



UNIVERSIDAD
DE MURCIA

DEPARTAMENTO DE FILOLOGÍA
INGLESA

Plaza de la Universidad • 30071 MURCIA (España)

Tel.: (34) (68) 36 31 91 • Fax: (34) (68) 36 31 85

7 MAR 2002



D. Pascual Cantos Gómez, Profesor Titular de Universidad del Área de Filología Inglesa y Director del Departamento de Filología Inglesa, INFORMA:

Que la Tesis Doctoral titulada "THE CATEGORISATION OF THE SOUNDS OF ENGLISH: EXPERIMENTAL EVIDENCE IN PHONOLOGY", "LA CATEGORIZACIÓN DE LOS SONIDOS DEL INGLÉS: EVIDENCIA EXPERIMENTAL EN FONOLOGÍA", ha sido realizada por D. Jose Antonio Mompeán González, bajo la inmediata dirección y supervisión de D. Antonio Barcelona Sánchez, y que el Departamento ha dado su conformidad para que sea presentada ante la Comisión de Doctorado.

Murcia, a 11 de Febrero de 2002





Don ANTONIO BARCELONA SÁNCHEZ, Profesor Titular de Universidad del Área de Filología Inglesa en el Departamento de Filología Inglesa AUTORIZA:

La Presentación de la Tesis Doctoral titulada “La Categorización de los Sonidos del Inglés: Evidencia Experimental en Fonología”, realizada por Don José Antonio Mompeán González, bajo mi inmediata dirección y supervisión, en el Departamento de Filología Inglesa y que presenta para la obtención del grado de Doctor por la Universidad de Murcia.

En Murcia, a 23 de enero de 2002

Fdo: Antonio Barcelona Sánchez.

UNIVERSIDAD DE MURCIA



1611270

CONTENTS

TABLE OF CONTENTS	i
ACKNOWLEDGEMENTS	iv
INTRODUCTION: OBJECTIVES AND JUSTIFICATION	vi
CHAPTER I: ON CATEGORISATION	1
1.1. Categorisation: a Basic Cognitive Process.....	1
1.1.1. Categorisation and Perception.....	2
1.1.2. Categorisation, Cognition, and Memory.....	3
1.1.3. Categorisation and Behaviour.....	5
1.1.4. Categorisation and Language.....	6
1.1.5. Types of Categories.....	7
1.2. Categorisation Research with Adult Humans.....	12
1.2.1. Classical Models of Categorisation.	12
1.2.1.1. Learning, Mental Representation, and Storage.	13
1.2.1.2. Classification.	14
1.2.1.3. Decadence of the Classical View: Questioning the Paradigm.....	15
1.2.1.3.1. Failure to Specify Defining Features.....	15
1.2.1.3.2. Use of Non-Necessary Properties in Categorisation.	17
1.2.1.3.3. Unclear Category Members.	17
1.2.1.3.4. Typicality and Typicality Effects.....	19
1.2.1.3.5. Taxonomic Phenomena.....	24
1.2.1.4. The Utility of the Classical View.....	30
1.2.1.5. The Emergence of Similarity-Based Categorisation Models.	33
1.2.2. Probabilistic Models of Categorisation.	34
1.2.2.1. Learning, Mental Representation, and Storage.	34
1.2.2.2. Classification.	35
1.2.2.3. Advantages and Disadvantages of Probabilistic Models of Categorisation.....	36
1.2.3. Exemplar Models of Categorisation.	38
1.2.3.1. Learning, Mental Representation, and Storage.	39
1.2.3.2. Classification.	39
1.2.3.3. Advantages and Disadvantages of Exemplar Models of Categorisation.	40
1.2.4. Mixed Models of Categorisation.	42
1.2.4.1. Learning, Mental Representation and Storage, and Classification.	42
1.2.4.2. Advantages and Disadvantages of Mixed Models of Categorisation.	43
1.2.5. Insufficiency of Similarity-Based Categorisation Models.	44
1.2.5.1. The Emergence of the Theory-Based Account of Categorisation.	45
1.2.6. Categorisation Models: Final Concluding Remarks.	48
1.3. Categorisation Research with Other Populations.	51
1.3.1. Categorisation in Children.	51
1.3.1.1. Categorisation by Children: Issues.	52
1.3.2. Categorisation in Pre-Verbal Infants.	55
1.3.2.1. Categorisation in Infants: Issues.	56
1.3.3. Categorisation in Non-Human Species.	59
1.3.3.1. Categorisation in Avians.	59
1.3.3.2. Categorisation in Primates.....	61
1.3.3.3. Interspecies Comparison.	62

CHAPTER II: ON THE CATEGORISATION OF SOUNDS.67

2.1. The Categorisation of Sounds: Experimental Evidence.	67
2.1.1. The Categorisation of Sounds by Adults.	68
2.1.2. The Categorisation of Sounds by Children.	72
2.1.3. The Categorisation of Sounds by Infants.	75
2.1.4. The Categorisation of Sounds by Non-Human Species.....	77
2.2. The Categorisation of the Sounds of English and its Relation to the Field of Phonology.....	80
2.2.1. Descriptive Phonology vs. Psycholinguistic (Psychological) Phonology.	82
2.2.2. Success of the Different Approaches to the Task of Phonology.	83
2.2.2.1. Criticisms about the Use of Experiments in Phonology.	90
2.2.2.2. Advantages of the Use of Experiments in Phonology.....	93
2.2.2.3. The Complementariness of Descriptive and Psycholinguistic Phonology.....	97
2.3. Concept Formation as an Experimental Paradigm.	98
2.3.1. Parts of a Concept Formation Experiment.	100
2.3.1.1. The Instructions.	100
2.3.1.2. The Training (or Learning) Session.....	101
2.3.1.3. The Test Session.....	104
2.3.1.4. The Post-Experimental Interview.....	105
2.3.2. Interpretation of the Data.	106
2.3.2.1. Number of Subject Who Form the Category.....	106
2.3.2.2. Trial-to-Criterion Scores and Time to Solution.....	107
2.3.2.3. Subjects' Correct vs. Incorrect Responses to Different Stimulus Types.....	108
2.3.2.4. Subjects' Reaction Times.....	109
2.3.2.5. Subjects' Verbal Report in the Post-Experimental Interview.....	109
2.4. Versions of the Concept Formation Experimental Paradigm.....	110

CHAPTER III: EXPERIMENTAL EVIDENCE FROM THE CATEGORISATION OF SOUNDS: A PSYCHOLINGUISTIC PHONOLOGICAL INVESTIGATION.....115

3. 1. Introduction.....	115
3.1.1. Overview of the Present Experiments.	115
3.1.1.1. Aims of the Experiments and Selection of the Issues under Investigation.....	115
3.1.1.2. Selection of Subjects.	116
3.1.1.3. Arrangement of the Experimental Tasks.	117
3.1.1.4. Stimuli, Materials, and Apparatus.....	118
3.1.1.5. Specific Procedural Aspects of the Different Parts of the CF Experiments.....	121
3.1.1.6. Data Collection and Interpretation.	124
3.2. Experiment 1. The Category "Consonantal Sounds" in English.	126
3.2.1. Design: Rationale, Purpose of the Experiment, Research Questions, and Hypotheses.....	126
3.2.2. Method.	134
3.2.2.1. Subjects.	134
3.2.2.2. Stimuli.	134
3.2.2.3. Procedure.	138
3.2.2.3.1. Instructions.....	138
3.2.2.3.2. Training Session, Test Session, and Post-Experimental Interview.....	139
3.2.3. Results.	139
3.2.3.1. Training and Test Sessions.....	139
3.2.3.2. Post-Experimental Interview.	149
3.2.4. Discussion.	152
3.3. Experiment 2. The Category "Oral Plosive" in English.	158
3.3.1. Design: Rationale, Purpose of the Experiment, Research Questions, and Hypotheses.....	158
3.3.2. Method.	163
3.3.2.1. Subjects.	163
3.3.2.2. Stimuli.	163
3.3.2.3. Procedure.	166
3.3.2.3.1. Instructions.	167
3.3.2.3.2. Training Session, Test Session, and Post-Experimental Interview.....	167
3.3.3. Results.	167
3.3.3.1. Training and Test Sessions.	167
3.3.3.2. Post-Experimental Interview.	176
3.3.4. Discussion.	180

3.4. Experiment 3. The Category “Phoneme / p /” in English.	185
3.4.1. Design: Rationale, Purpose of the Experiment, Research Questions, and Hypotheses.....	185
3.4.2. Method.	194
3.4.2.1. Subjects.	194
3.4.2.2. Stimuli.	195
3.4.2.3. Procedure.	198
3.4.2.3.1. Instructions.	199
3.4.2.3.2. Training Session, Test Session, and Post-Experimental Interview.....	199
3.4.3. Results.	200
3.4.3.1. Training and Test Sessions.	200
3.4.3.2. Post-Experimental Interview.	206
3.4.4. Discussion.	208
3.5. Experiment 4. The Category “Allophone [p ^h]” in English.	216
3.5.1. Design: Rationale, Purpose of the Experiment, Research Questions, and Hypotheses.....	216
3.5.2. Method.	220
3.5.2.1. Subjects.....	220
3.5.2.2. Stimuli.	220
3.5.2.3. Procedure.	220
3.5.2.3.1. Instructions.	223
3.5.2.3.2. Training Session, Test Session, and Post-Experimental Interview.....	223
3.5.3. Results.	224
3.5.3.1. Training and Test Sessions.	224
3.5.3.2. Post-Experimental Interview.	229
3.5.4. Discussion.	234
CHAPTER IV: CONCLUSIONS	239
4.1. Introduction.....	239
4.2. Categorisation, The Categorisation of Sounds, Phonology, and Experimentation.....	239
4.3. CF as an Experimental Paradigm.....	240
4.3.1. Evidence for Specific Categories: Within-Category Facts.	241
4.3.2. Evidence for Between-Category Similarities and Differences.	242
4.3.3. Experimental Evidence and the Solution to Phonological Problems.	247
4.3.4. The Utility of the Concept Formation Paradigm.....	250
4.4. Directions for Future Research.....	251
4.5. Conclusion.....	256
REFERENCES	259

ACKNOWLEDGEMENTS

I would like to thank a number of people who have contributed to a greater or lesser extent to the attainment of this dissertation.

My deepest appreciation is to my adviser, Antonio Barcelona Sánchez, from whom I have always received considerable help, encouragement and support in absolutely every respect.

I would further like to express my thanks to the following people for their help in obtaining the bibliographic materials needed to conduct this research. These include researchers from the University of Murcia (Alfredo Brito, Javier Valenzuela, Laurence Erussard, Teresa Moralejo, Ana Mansilla, M^a Mar Vilar) as well as from other universities in Spain (María Josep Cuenca, University of Valencia; Joseph Hilferty, University of Barcelona) and in other countries (Dahlia Sabry, University of Modern Sciences and Arts, Cairo, Egypt; Anne Slon, Swietokrzyska Academy, Department of Modern Languages, Kielce, Poland; Lawrence Barsalou, Emory University, Atlanta, USA; Francisco Serra, University of Hull, UK; Raymond Weitzman, California State University, Fresno, USA; Andrew Wedel, University of California in Santa Cruz, USA; David Eddington, University of New Mexico, USA). My thanks also go to the library staff at the universities of Murcia and Granada and to the Servicio de Préstamo Interbibliotecario at the University of Murcia for their priceless help.

A number of colleagues also contributed to selecting the subjects that took part in the experiments reported in this dissertation. These include Lourdes Cerezo, Ana Rojo, Laura Campoy, and Lorena Asensio. I would also like to thank the International Relations Office at the University of Murcia for their help in contacting the subjects to whom I am also indebted for their sincere cooperation and willingness to perform the tasks.

I am grateful to Andrew Wedel for useful discussions at all stages of this project, to Rafael Monroy Casas and Francisco Gutiérrez Díez for extremely valuable suggestions, and to Katherine Connolly and Cristina Soriano for helpful comments. I especially appreciate the efforts of David Eddington, who provided detailed corrections on the final draft and helped me clarify a number of major substantive issues in the whole presentation. I alone, however, am responsible for any errors of form, content and interpretation

I have also benefited from the technical assistance of Alvaro Martínez García (graphic design), Óscar de Jodar Bonilla (DartPro assistance), Elizabeth Murphy and Caroline Lips (stimulus recording), Francisco Luis García Abad and Juana María Vivo (statistical analyses).

In this respect, I am particularly indebted to my friend and colleague Lorenzo Fernández Maimó, who designed the computer programme with which the experiments were performed.

I am also indebted to many friends for their invaluable support and understanding while this research was being conducted. These include Suleimán Salmán Tudela, José Luis Jiménez Penalva, Juan Murcia Calatrava, and Jesús Cano Sánchez. My greatest debt in this respect is no doubt with Veridiana Martínez García, who has patiently gone through all the stages of this dissertation.

Finally, I would like to dedicate this dissertation to my parents, Agustina González Zaragoza and Pedro Mompeán Flores, my grandparents (Francisca Zaragoza Martínez, María Flores, Antonio González Almagro), and my brother Pedro Mompeán González.

José Antonio Mompeán González

Murcia, February 2002.

INTRODUCTION: OBJECTIVES AND JUSTIFICATION

Phonology can benefit by embracing the data and methods from fields as diverse as acoustics, psychology, and ethnology. Phonology does not lose its identity by this; what sets phonology off from other disciplines is its *questions*, its *end*, and its *methods or means*.

John J. Ohala (1983: 241).

One of the most essential cognitive abilities is that of **categorisation**. For the past thirty years an enormous body of experimental research has been devoted to categorisation. Different disciplines, like cognitive psychology, anthropology, philosophy, and linguistics have contributed to a better understanding of this basic cognitive ability. The findings of this literature have changed and are still changing much of our understanding of the mind and the way it works.

In general, *categorisation* refers to or is a synonym of *classification*. In fact, many researchers have defined categorisation with an explicit reference to classification. The following definitions of categorisation may suffice to illustrate the point: “the process by which people assign objects to categories” (Smith 1990: 33) and “a cognitive process in which people decide whether an instance is a member of a category by comparing the instance with their conceptual representations” (Lin & Murphy 1997: 1153), “determining that a specific instance is a member of a concept (for example, this particular creature is a *guppy*) or that one particular concept is a subset of another (for example, *guppies* are *fish*)” (Smith & Medin 1981: 7), “the means by which people decide whether or not something belongs to a simple class” (Medin & Smith 1984: 114), etc. Categorisation seems to be then, according to these definitions, the process by which people classify things. Classifying a given entity as a dog, or as a chair is, according to these definitions, an act of categorisation. Classifying a dog, a cat and a horse as “animals” would also be an act of categorisation.

However, although categorisation behaviour has been extensively investigated experimentally in different fields of research (most notably cognitive psychology, developmental psychology, social cognition, etc.), the experimental study of categorisation in

linguistics has been much more in the field of linguistics and perhaps even more blatantly neglected in the field of phonology, which represented, in many respects the origin of modern linguistics.

The present dissertation tries to contribute to the experimental study of categorisation from the field of experimental phonology (see Jaeger & Ohala 1986). However, to better understand the relevance of categorisation for phonology, the thesis provides an intellectual tour from the extensive body of research conducted on categorisation with visual stimuli, to the study of the auditory categorisation abilities of different subject populations, to the specific study of the categorisation of sounds in phonology with a psychological commitment. The relationships between categorisation, the categorisation of sounds, phonology, and experimentation, the present dissertation are strung together in three chapters, plus a final chapter that summarises the relationships. The outline of the present dissertation is as follows:

Chapter I, called “On Categorisation”, offers an overview of the categorisation research carried out during the past eighty years (in particular the last thirty years). This overview provides the background to the study of the categorisation of sounds in this dissertation. The intention of this chapter is twofold: to provide a sound cognitive theoretical basis to the study of phonological categorisation and to offer an state-of-the-art account of past and present research to those interested in categorisation, cognition, language, linguistics, and phonology, and the relationships between them.

Chapter II, called “On the Categorisation of Sounds” shows how a great deal of experimental research has already more or less explicitly concerned itself with the classification of sounds by speakers. The chapter also deals with the long-standing debate on whether phonology should be a mere taxonomic discipline with an autonomous methodology or whether phonological descriptions should reflect psychological reality and should draw on methods and findings from other related experimental sciences to validate psychological claims. The latter position is adopted. In this respect, since the present dissertation tries to provide experimental psycholinguistic evidence on certain phonological problems and endorses the view that phonological descriptions should be guided and checked by experimental evidence, the choice of an appropriate experimental technique to study the categorisation of sounds as a source of evidence is justified. One promising experimental paradigm, known as “concept formation” is discussed in great detail and adopted to conduct the studies presented in this dissertation.

Chapter III, called “Experimental Evidence from the Categorisation of Sounds: A Psycholinguistic Phonological Investigation”, begins with an overview of the four experiments presented in this chapter. Next, the four experiments carried out with native speakers of English. Four different experiments are reported which show how speakers of English are able to create or retrieve from long-term memory categories of sounds and classify sounds as members of four different categories.

These four experiments also provide evidence as to different traditional specific problems in phonological theory related to each of the phonological categories under investigation in each of the four experiments respectively. These problems have received little experimental consideration in phonology. The category studied are those of the consonantal sound, the oral stop, the phoneme / p / and [p^h] in sections 2.1. through 2.4. respectively.

Chapter 4, “Conclusions”, summarises the connecting lines the previous chapters and compares the results of the four experimental studies conducted in the previous chapter. The chapter also examines the usefulness of experimental evidence for the studies conducted and in general for any experimental task similar to the ones conducted and proposes directions for the future application of the experimental paradigm used, namely “concept formation” in phonology or related disciplines.

Finally, the section “References” gathers the extensive (but still highly selective) body of literature cited throughout the present dissertation.

CHAPTER I.

ON CATEGORISATION.

1.1. Categorisation: A Basic Cognitive Process.

Categorisation has frequently been claimed to be a basic, essential or fundamental mental process. A few quotations may suffice to illustrate the point: “categorisation may be considered one of the most basic functions of living creatures” (Mervis & Rosch 1981: 89), “categorisation... is a fundamental cognitive ability” (Lin & Murphy 1997: 1154), “one of the most basic human cognitive processes is the ability to categorise” (Corrigan 1989a: 2), “categorisation is fundamental to human experience” (Niedenthal *et al.* 1999: 337), “there is nothing more basic than categorisation to our thought, perception, action and speech” (Lakoff 1987: 5), “categorisation is essential” (Mervis & Pani 1980: 496), “categorisation is an unquestionably important type of cognitive process” (Busemeyer *et al.* 1997: 2), “categorisation is one of the most basic cognitive functions” (Corter & Gluck 1992: 291), “a basic cognitive function is to categorise” (Medin *et al.* 1999: 366), etc.

In what sense might categorisation be basic? Apparently, considering categorisation as a basic cognitive ability can at least be interpreted in two different ways: in a technical and in a more colloquial or informal sense.

In a technical sense, categorisation is a basic cognitive process in Flavell’s (1985) operationalisation of the term. Flavell proposed to divide memory phenomena into four distinct (but not mutually exclusive) types: basic processes, knowledge (semantic memory), strategies, and metamemory (see also Rosch & Mervis 1981: 94). Basic cognitive processes are characterised by two main features: they undergo no significant development with age (other than that due to maturation) and they are unconscious processes: people are not aware of their actual workings. In principle, categorisation seems to comply with these two characteristics. First, categorisation starts with life: the ability to categorise is innate and is demonstrated in infants as early as 3 months (see 1.3.2.). This ability is present almost from birth and, although it develops with age, it basically remains the same throughout one’s life. Second, categorisation is an unconscious process that people typically perform automatically and effortlessly (Barsalou

1992a: 16) and if people become aware of it at all, it is only in problematic cases (Lakoff 1987: 6).

However, most of the statements that categorisation is basic do not simply try to convey the idea that categorisation is a particularly important function. To show why this should be so, we may ask three questions the answers to which further clarify why categorisation may be so important. First, *who* categorises? The answer is that every human being possesses that ability: categorisation is a universal capacity in the sense that all human beings possess it. As Jaeger (1980: 7) puts it, “the perceptual and cognitive structures which allow man to categorise are innate and universal”.¹ Second, *when* does categorisation take place? Categorisation occurs whenever we perceive stimuli or wherever relevant knowledge might be brought to bear, which is most of the time (Lin & Murphy 1997: 1154; Medin *et al.* 2001: 366). Furthermore, categorisation occurs in all sensory modalities, not just vision (Barsalou 1992a: 15) and it may not be infrequent that several senses (e.g. sight and hearing) may be simultaneously involved in an act of categorisation. Categorisation seems to be a pervasive, ever-present process in mental life (Bornstein 1984: 313). Third, what is the *role* of categorisation in human cognitive functioning? The answer to this question most clearly reveals why categorisation should be considered as a basic cognitive ability. More specifically, categorisation seems to be intimately linked to perception, cognition, memory, behaviour, and language (Bornstein 1984: 313-315). This point is nicely summarised in Lingle and his co-worker’s statement that “from elementary perception to complex reasoning tasks, categories and categorisation processes pervade our mental functioning” (Lingle *et al.* 1984: 73-74). The following four sections illustrate this.

1.1.1. Categorisation and Perception.

Categorisation seems to be essential to perception. According to Bruner, “perception involves an act of categorisation” (1957: 123). To Hunt, “the act of categorising sensory input is basic to our perception of the world about us” (1962: 5). Categorisation appears to be basic to perception in the sense that it promotes structured perception by surmounting natural variation and making sense of the sheer diversity of the stimuli organisms encounter (Rosch 1975d; Smith & Medin 1981; Wilder 1981; Younger & Cohen 1985). In discussing categorisation, dozens of researchers have emphasised the idea that organisms are confronted with a world with an infinite variability of unique, distinguishable or distinct stimuli of many different types (e.g. Barsalou 1992a; Bourne 1966; Corrigan 1989a; Hulse *et al.* 1980; Hunt 1966; Jaeger 1980a; Markman 1983; Medin *et al.* 2001; Mervis 1980, 1985; Mervis & Pani 1980; Mervis & Rosch

¹ However this ability is not restricted to the human species (see section 1.3.3.).

1981; Rosch 1978).² The fact that the world provides us with such a tremendous diversity poses a perceptual problem. If organisms were to perceive each stimulus not previously encountered as exclusively unique, they would be rapidly overwhelmed by the extreme diversity of their environment and would be unable to make any sense of their experience (Smith & Medin 1981: 1).

However, organisms simplify this problem by means of the ability to categorise. In order to cope with this enormous diversity of varied and varying stimuli, categorisation allows people to order their experience by considering some of the stimuli as similar or equivalent rather than as unique (Hulse *et al.* 1980; Hunt 1966; Jaeger 1980a; Lingle *et al.* 1984; Markman 1983; Medin & Schwanenflugel 1981; Mervis & Pani 1980; Murphy & Wisniewski 1989b; Rosch 1975d, 1978; Rosch *et al.* 1976; Smith & Medin 1981; Wisniewski & Medin 1991: 237). Categorisation permits them to organise the input from their environment into cognitively manageable units by directing them to perceive groups of experiences as being somehow similar so that the complexity of the environment is reduced (Bornstein 1984: 314). In other words, categorisation makes it possible for an organism to reduce the limitless variation in the world to manageable proportions (Jaeger 1980a : 7) and simplify the environment to some degree. In this sense, organisms achieve one of their most basic needs: a sense of stability and order in their environment (e.g. Mervis 1985: 293). Categorisation allows organisms to perceive stimuli in the environment as being organised into a system of structures and relationships, and therefore allows them to achieve a sense of order and continuity.

However, categorisation affects perception in yet another way. Psychologists have long been intrigued by the possibility that the concepts that people learn influence their perceptual abilities. For example, as experts like orthopaedic surgeons learn to distinguish amongst the types of fractures that there are, they seem to acquire new ways of perceptually structuring the objects to be categorised. In other words, categorisation can produce perceptual learning that lasts beyond the categorisation act. In fact there is some experimental support for the notion that experience in acquiring new categories can alter perceptual sensitivity and the ability to make perceptual discriminations (Goldstone 1994a; MacKintosh 1995: 196-1998).

1.1.2. Categorisation, Cognition, and Memory.

Researchers also agree on the facilitative role of categorisation and categories for cognition and memory. First, through categorisation we create abstract mental categories or concepts from experience with stimuli, which is “one of the hallmarks of human intelligence”

² Although stimuli are never the same, organisms are sometimes not able to discriminate or distinguish them. In this case, categorisation becomes *categorical perception*, and discrimination is entirely or severely limited (e.g. Bomba 1984; Goldstone 1994).



(Busemeyer *et al.* 1997: 1).³ Categorisation allows concept learning, a fundamental aspect of intelligent thought (Wisniewski & Medin 1991: 237).

Second, cognitive categories help us recognise familiar information by providing “receptacles” (i.e. the conceptual categories) into which incoming information can be encoded in subsequent acts of categorisation. Organisms are constantly exposed to objects that they have never seen before and yet categorise them as members of familiar classes. Therefore, those receptacles make the unfamiliar familiar (Bornstein 1984: 314; Lin & Murphy 1997: 1154; Mervis 1980: 279; Mervis & Pani 1980: 497; Spalding & Murphy 1996: 525).

Third, not only do categories allow organisms to identify novel stimuli as members of a given familiar category or categories⁴ but also to apply or generalise more knowledge previously acquired about that category to new members through drawing inferences (e.g. Bornstein 1984: 314; Malt *et al.* 1995; Mervis 1980: 279; Murphy & Ross 1994: 148; Ross & Murphy 1996: 736).⁵ As Medin and his co-workers put it:

If each thing we encountered was unique and totally unlike anything else we had ever known, we would not know how to react to it or make any useful predictions about its properties. We would be literally lost in a sea of new experiences, helpless to employ any of our prior knowledge to navigate it.

Douglas L. Medin et al. (2001: 366).

Categories are useful because, once the category assignment of an object is known, a person can predict both the internal (invisible from the surface) attributes of the instance in question or its functions (Mervis 1985: 293; Wisniewski & Medin 1991: 237). For example, an organism searching for food needs to know whether a particular plant part is poisonous, nutritive, sweet, tough, and so on (Corter & Gluck 1992: 291). Having identified something as a specific type of mushroom, the organism may then understand whether it is dangerous or not (Medin *et al.* 2001).⁶

³ Some researchers distinguish between “concepts” (the mental representation of entities or their “intension”) and “categories” (the set of entities in the world which are grouped together on the basis of some criterion, that is, concepts’ “extension”). However, the term “category” can also be used to refer to a collection of memories of experienced stimuli that are stored in memory as well as to a set of real-world stimuli that categorisers believe belong together. The former use of the term category is made more explicit if the adjectives conceptual, cognitive or mental precede it.

⁴ This is so because categories allow cross-classifications or the assignment of one entity to more than one non-contrasting categories (e.g. Barsalou 1982; Murphy 1993a; Murphy & Ross 1999; Ross & Murphy 1999). A dog, for example, can be classified as a “dog”, a “pet”, a “German shepherd”, a “mammal”, a “vertebrate”, a “carnivore”, a “quadruped”, a “land-dweller”, a “guard dog”, etc. Cross-classification is a widespread phenomenon in social categorisation. Clearly, individual people fit a very large number of social categories simultaneously. A person may be classified as a democrat, a fiscal conservative, a feminist, and a golfer at the same time).

⁵ Although the initial goal in categorisation is to figure out what something is, classification is usually not an end in itself or the ultimate goal of categorising. Knowing that an animal is a dog is of little use unless that classification allows one to accomplish some goal, such as deciding whether the animal can be approached or if it is likely to be a pet (Ross & Murphy 1999: 528). For this reason drawing inferences about the already-classified entity which are relevant to interacting with it is essential. For example, after having classified some animal as a “dog”, we can infer that it is “likely to bark”, “might attempt to attack a stranger”, “is an affectionate pet”, etc. Inferences may concern the origin of the entity or its features, the behaviours of the entity likely to be observed, the relevance of the entity to the perceiver’s goals, or plans for interacting with the entity to achieve these goals, etc.

⁶ Classification and inference-drawing are two of the most widely discussed functions of concepts. However, other functions of concepts are also important to cognition (see Medin *et al.* 2001; Medin & Barsalou 1987; Medin & Smith 1984; Rey 1983; Rips 1995; Smith & Medin 1981 for discussions). The most widely studied function apart from classification and inference-drawing is conceptual combination or the construction of new concepts from combining existing ones like “orange dog”, “pet fish”, etc. (e.g. Hampton 1982, 1987, 1988a, 1988b, 1991; Jones 1982; Osherson & Smith 1981, 1982; Wisniewski 1996; Zadeh 1982).

Finally, another advantage of the categories created through categorisation is that they reduce the necessity of constant learning because most new stimuli can be immediately understood as an instance of some type of stimuli previously encountered. As a consequence, those stimuli can be acted upon appropriately on the basis of past learning (Jaeger 1980: 7; Mervis & Rosch 1981: 89). Therefore, fewer items need to be retained individually after encoding (Bornstein 1984: 314). In addition, grouping objects into categories allows for more efficient storage of information in memory about these groups of objects (Bornstein 1984: 314; Corter & Gluck 1992: 291). Also, it becomes easier to remember previously encountered stimuli. In fact, without categories organisms would probably be unable to remember more than a minute fraction of what they encounter. Concepts serve to enhance memory capacity (Bornstein 1984: 314; Smith & Medin 1981: 1).

Due to all the ways in which categorisation is related to cognition and memory, it is not surprising to find statements in the literature claiming that “categorisation is a central part of intelligent thought” (Murphy & Ross 1994: 148).

1.1.3. Categorisation and Behaviour.

Categorisation and cognitive categories are largely unobservable internal phenomena. However, the workings of categorisation and the existence of such categories are inferred from certain behavioural patterns. Thus, the first way in which categorisation is related to behaviour is that categorisation causes organisms to respond similarly to members of one class of stimuli and differently to members of other classes of stimuli. A category is believed to exist in the mind whenever two or more distinguishable stimuli are treated equivalently and differently from members of other categories (Bomba 1984; Bornstein 1984: 314; Bourne 1966: 2; Hunt 1966: 1; Medin *et al.* 1987; Mervis 1985: 293; Mervis & Pani 1980: 496-497; Mervis 1980: 279; Mervis & Rosch 1981: 89; Murphy & Brownell 1985: 70; Spalding & Murphy 1996: 525; Wasserman *et al.* 1992:374). In this respect, researchers have emphasised that this equivalent treatment or “sameness-of-response” can take many different forms, such as labelling distinct objects or events with the same name, performing the same action on different objects, etc.⁷ For example, a person might indicate the category “bird” by calling an eagle, a robin, a chicken and an ostrich each a “bird” or s/he might indicate the category “things that can be sat on” by sitting on a chair, a couch, a stool, and a tree stump or by performing both motor and verbal acts at the same time.

⁷ However, Epstein (1982) warns that behavioural variation is overlooked or insufficiently emphasised in most studies of categorisation. Epstein claims that behavioural acts vary as much as stimuli. Organisms may respond to a series of different stimuli with similar actions or linguistic labels, but these will never be the same. Different dogs, for instance, may be called “dog”, but different instances of the word “dog” will vary in pitch, volume, duration, etc. Furthermore, even in cases in which subjects press buttons to categorise stimuli (as in laboratory experiments) the button pushes vary in topography, latency, magnitude, and duration. Consequently, categorisation implies treating different stimuli as the same performing behavioural acts that can themselves be considered as (but are not) the same.

Categorisation influences behaviour in still another way. Classifying something as a member of a particular category or another determines which type of behaviour organisms perform in the future towards that stimulus once it has been classified. This has to do with the inferences that categorisation promotes. However, a very similar, if not equal, treatment towards the members of a category, although useful in general, may sometimes cause problems. To treat the members from a common category equally on every situation may be problematic, sometimes even unfair. Specifically, categorising a set of objects can lead us to treat the members of a category as more similar to each other than they really are simply because of their common category membership (Medin *et al.* 2001: 369-371). Categorisation can both exaggerate (between category) differences and inappropriately minimise (within-category) differences. This risk is particularly evident in social categorisation, where stereotypes may sometimes have dangerous consequences derived from people's social behaviour.

1.1.4. Categorisation and Language.

Categorisation has also been suggested to be basic to an understanding of the nature of language acquisition/learning, development and language use (e.g. Anglin 1977; Bruner *et al.* 1956; Clark 1973; Nelson 1973; Schlesinger 1982; etc.). According to Nelson, "categorization of sound patterns and of objects and events in the real world is basic to learning a language" (Nelson 1973: 21). Categorisation has also been claimed to be essential to language comprehension (van Gelder 1993), speech perception or recognition (Hunt 1962; 5; Lotto 2000), early lexical development (Gopnik & Meltzoff 1987, 1992; Poulin-Dubois *et al.* 1995), or word use (Bowerman 1978, 1980; Roberts & Horowitz 1986), etc. In many respects then, language is "a categorising activity" (Labov 1973: 342).

As far as lexical acquisition is concerned, for example, a strong relation has been reported between an active and high level of categorisation behaviour and the so-called "naming explosion", that is, the rapid increase in the acquisition of object nouns that occurs in the second year of life. The general consensus is that children with more names in their vocabulary have better categorisation skills (Gopnik & Meltzoff 1987; Poulin-Dubois *et al.* 1995).

Word use has also been claimed to be directly related to categorisation. More specifically, it is often conceived of as a routine act of categorisation. The act of naming or reference which associates a linguistic sign with an element of the extra-linguistic world can then be considered as an act of categorisation. As Bowerman puts it (1980: 278), the ability to categorise "permits symbols such as words to stand not only for unique objects and events but also for whole arrays of discriminably different stimuli". This, in turn, makes communication easier because if each individual entity needed a distinct name, our language would be staggeringly complex. For Tversky (1986: 63), "naming an object categorises it: when individuals call an object a table,

they are saying that, for the purpose at hand, it is equivalent to other tables (Tversky 1986: 63). In short, naming an object (e.g. “this is a table”) is a way of categorising it because it implies that that particular unique entity is similar to other members of the category. In Bourne’s words: “language assists in our efforts to deal and contend with a potentially overpowering real world of varying and unique objects” (1966: 107), a function that words serve since at least the second year of life (Nelson *et al.* 1978: 960).

However, it is paradoxical that if words serve to categorise environmental stimuli, this function of language is itself possible because “our entire system of naming objects and events... presupposes the ability to group experiences into appropriate categories” (Medin & Schwanenflugel 1981: 355).

In short, thanks to categorisation and cognitive categories, each new stimulus in the world is perceived, processed, remembered, acted upon or talked about not as unique, but rather as an instantiation of a category already in the mind. Clearly, categorisation implies a basic, economical, and sophisticated cognitive approach to the world (Bornstein 1984: 315). Categorisation simultaneously embraces unity and diversity, constancy and variation, and it is central to a host of significant mental functions critical to an organisms’ survival (Wisniewski & Medin 1991: 237). It is not surprising then that man has also been defined as “a categorising animal” (Labov 1973: 342) or as “the classifying animal” (Berlin *et al.* 1973: 214). In light of the above discussion then, Lakoff’s assertion that “without the ability to categorise, we could not function at all, either in the physical world or in our social and intellectual lives” (1982: 3, 1987a: 6) does not seem exaggerated.

However, basic as categorisation is to cognition, extreme proposals reducing all cognition to a kind of categorisation are not appropriate. The view that all cognitive operations are just mental categorising activities is too extreme. For example, important as categorisation is for language understanding or sensorimotor control, it is not the most essential component of those processes (van Gelder 1993). Categorisation should therefore be placed *within* cognition, but cognition is not a kind of categorisation. Similarly, it is useless to reduce all speech events to a categorising activity as language serves many other different functions just as there are thousands of behavioural acts that do not have a categorising end.

1.1.5. Types of Categories.

The different types of categories studied in the categorisation literature reflect the heterogeneity of stimuli with which our cognitive systems deal. Apparently, any kind of stimulus can be the object of categorisation since individuals seem to have categories for everything we can think about (Lakoff 1987a: 9).

It is obvious that there are different kinds of categories, at least in the everyday sense of *different* and *kinds* (Medin *et al.* 2001: 393). However, it is very difficult to establish taxonomies of types of categories and distinctions between different kinds of categories are not always clear-cut. Categories might differ in structure, processing assumptions, content, etc. Taxonomies of categories have been proposed, for example, in terms of the degree of physical variation of category members (Bornstein 1984: 315-326). Ideally, specific criteria for what should count as a different category type should be very explicit. The review presented below covers research conducted experimentally with adult human subjects and merely refers to the types of categories that researchers have distinguished. Most research has dealt with common, concrete categories, usually divided into two main types: *natural* categories (natural kinds such as “birds”, “cats”, “trees”, “leaves”, “vegetables”, “fish”, etc.) and *artefactual* categories (human artefacts such as “tools”, “toys”, “cars”, “chairs”, “furniture”, “clothing”, “vehicles”, etc.).⁸

Natural and artefactual categories have been extensively studied in categorisation research due to their suitability for testing different categorisation models (see 1.2.). These categories apparently produce no biases for or against any particular model of categorisation (Smith & Medin 1981: 5). These categories are also advantageous in that their members are usually common, familiar and depictable and relatively well-understood, both by researchers and their subjects (Tversky 1990: 335). Therefore, their relative simplicity makes them attractive for laboratory experiments (Busemeyer *et al.* 1997: 2) although this does not mean that other types of categories (see below) are not.

In addition to work on natural and artefactual categories, “artificial categories” constructed for a particular experiment have also been widely used in many laboratory investigations. Artificial categories provide a powerful scientific tool for exploring phenomena observed initially in natural categories. This is so because they are easier to manipulate systematically as the features that specify membership in those categories can be stated precisely and their contribution in determining category membership can be assessed in a straightforward way. In this way, researchers can have a more precise control over the instances that subjects are exposed to in acquiring a given category and the variables of interest, something that does not occur with natural or artefactual categories (Rosch & Mervis 1975: 591).

Artificial categories are of two types. The first type includes categories constructed carefully to mimic natural categories. These include drawings of novel artefacts (Lin & Murphy

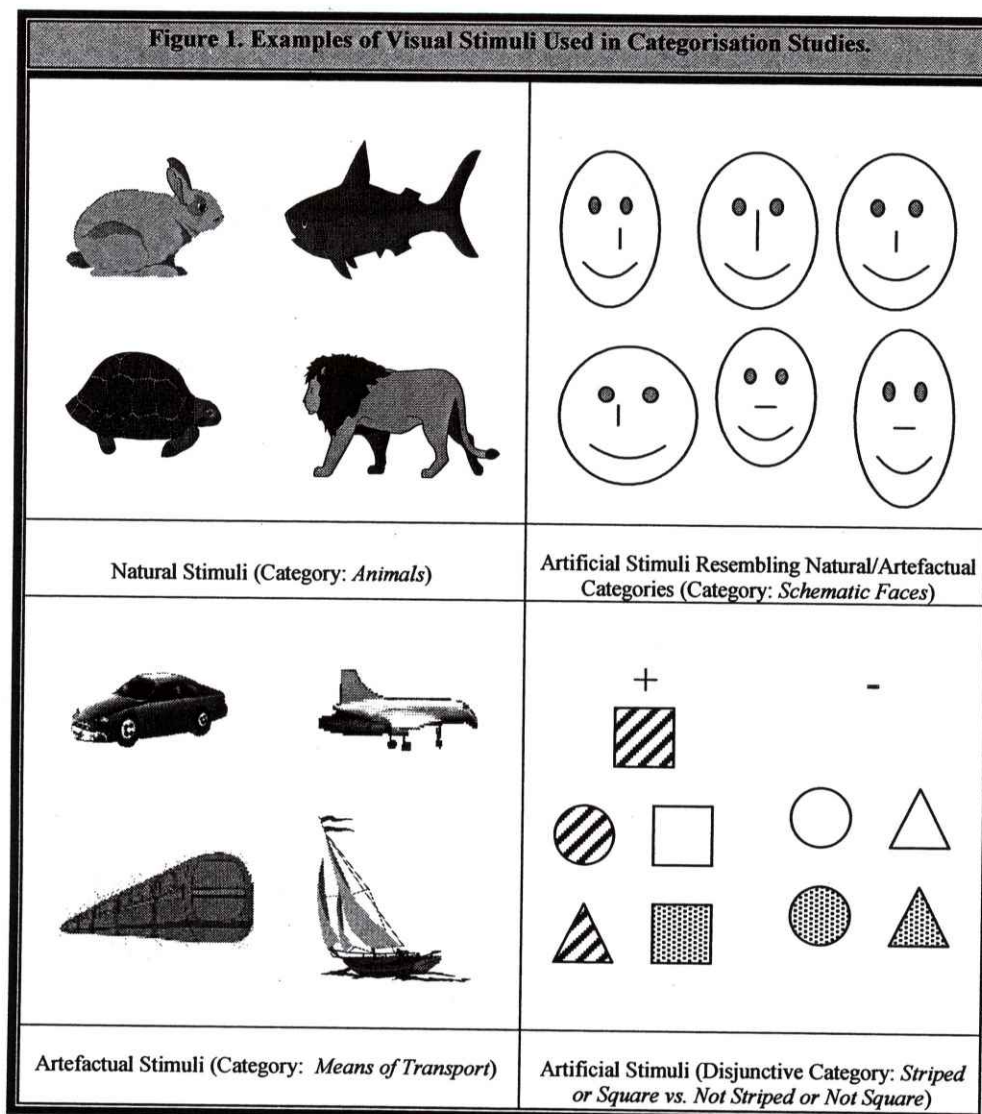
⁸ The focus on these two types of categories derives from research in the field of cognitive psychology, with the influence of the early work of William Battig and William Montague (1969), Eleanor Rosch (e.g. Rosch 1973b, 1975b; Rosch & Mervis 1975; Rosch *et al.* 1976), or Edward E. Smith (e.g. Smith *et al.* 1974), who sparked interest in such categories amongst psychologists (see e.g. Armstrong *et al.* 1983; Hampton 1979; McCloskey & Glucksberg 1978, 1979; Medin & Schwanenflugel 1981; Murphy & Wisniewski 1989a; Ross & Murphy 1999; Tversky & Hemenway 1984, for half a dozen influential papers) and the fields of

1977; Murphy & Smith 1982), schematic drawings of faces (Goldman & Homa 1977; Kellogg 1981; Kemler Nelson 1984; Lamberts 1994, 1995; Martin & Caramazza 1980; Medin & Schaffer 1978; Murphy & Ross 1994; Neumann 1977; Nosofsky 1991a; Reed 1972, 1978), drawings of starfish-like things (Ahn & Medin 1992), drawings of imaginary animals (Malt 1989), multidimensional objects (Lamberts & Freeman 1999), caricature irises (Anderson & Fincham 1996), etc.

The second type includes completely arbitrary categories. This subtype of categories was very often used in the typical concept formation experiments of the behaviourist and information processing eras. The instances of those categories are usually constructed from a small number of separable orthogonal perceptual dimensions (e.g. form, colour, size) that vary according to a small number of features (e.g. red, blue, green, for the colour dimension). In other words, any instance combines features out of any logically possible combination in an arbitrary manner. Studies of this second type include dot patterns (Homa & Chambliss 1975; Homa & Cultice 1984; Homa *et al.* 1973; Homa & Vosburgh 1976; Homa *et al.* 1987; Knapp & Anderson 1984; Omohundro 1981; Posner 1969; Posner & Keele 1968, 1970; Posner *et al.* 1967; Robbins *et al.* 1978; Rosch *et al.* 1976; Strange *et al.* 1970), geometric figures, designs or patterns (Bourne 1982; Dennis *et al.* 1973; Goldstone 1994a; Lamberts & Brockdorff 1997; Medin & Schwanenflugel 1981; Medin *et al.* 1984; Murphy & Ross 1994; Neumann 1974; Nosofsky 1991d), letter strings (Reber & Allen 1978; Rosch & Mervis 1975; Rosch *et al.* 1976; Whittlesea 1987), etc.

Despite the fact that natural, artefactual and artificial categories have been extensively used by categorisation researchers (see figure 1 for visual examples of such type of stimuli), these have also turned their attention to other types of categories. In this respect, it is useful to remember the proposal of Bruner and his co-workers (Bruner *et al.* 1956) that there are three main types of categories: formal, functional, and affective. Natural, artefactual and artificial categories are usually defined in terms of their perceptual physical characteristics so they are “formal” because they are defined in terms of their physical “form” and appearance. However, other types of categories largely composed of physical instances are based on functional relevance. For example, “ad-hoc” categories are categories constructed on a specific occasion to achieve a current, immediate goal (e.g. Barsalou 1981, 1983, 1984, 1985; Barsalou *et al.* 1986; Barsalou *et al.* 1987; Ross & Murphy 1999). For example, if someone is going camping for the first time, he or she may construct the ad hoc categories “places to go camping” or “things to take camping”, etc. These categories may be instantiated by objects as physically dissimilar as a tent, a lamp, a sleeping bag, etc. However, these objects are members of the category because they serve to achieve the goal of going camping.

cognitive anthropology (e.g. Brown *et al.* 1976; Kay & McDaniel 1978; Kempton 1978, 1981; Labov 1973) and ethnobiology (e.g. Berlin 1978; Berlin *et al.* 1973; Brown *et al.* 1976; Stross 1973), with their extensive work on folk taxonomies.



Other types of categories created for their social functionality are the so-called “social” categories studied in the field of social perception. These are categories that refer to particular age groups or occupations, races, religions, nationalities, personality traits, etc. In general we can distinguish three main subtypes of social categories: groups of people like “us”, “them”, etc. (Allen & Wilder 1975; Billig & Tajfel 1973; Brewer *et al.* 1981; Howard & Rothbart 1980; Nelson & Miller 1995; Smith & Zárate 1990; Tajfel & Billig 1974; Tajfel *et al.* 1971; Taylor 1981; Wilder 1981), traits of people, like “quarrelsome”, “extrovert”, “talkative”, “helpful”, etc. (Andersen & Klatzky 1987; Borkenau 1990; Buss & Craik 1983; Cantor & Mischel 1977; Chaplin *et al.* 1988; Hampson 1982; Hoffman *et al.* 1981; Isen *et al.* 1992; John *et al.* 1991; Read *et al.* 1990a, 1990b; Wojciszke & Pienkowski 1991), and “profession” stereotypes like “librarian”, “worker”, “politician”, “comedian”, etc. (Andersen & Klatzky 1987; Brewer *et al.*

1981; Cantor & Mischel 1979a, 1979b; Cohen 1981; Dahlgren 1985; Hastie *et al.* 1990; Trentin & Salmaso 1990).

Categories are also created out of some emotional or affective reason. The so-called “emotional-response” categories are those in which category members belong in the same category because they evoke a similar emotional response (e.g. Niedenthal & Halberstadt 1995; Niedenthal *et al.* 1999). For example, a hermit crab could be grouped together in a formal category together with snow crabs and lobsters because of their perceptual similarity. However, a child who cherishes it as his childhood pet could group it together with other things that evoke happiness like his skateboard and his Star Wars action figures in a category like “things that make me happy”. Everyone’s own category of “friends” is certainly an emotional-response category.

Goal-derived and emotional response categories are still largely categories of objects. However, a large proportion of human categories are not categories of things but categories of abstract entities (Lakoff 1987: 6). These include categories as heterogeneous as works of art, beliefs, crimes, just decisions, instincts, rules, sciences and kinds of work (Hampton 1981), intelligence (Neisser 1979), the self (Roberts *et al.* 1977), emotions (Fehr 1988; Fehr & Russell 1984; Shaver *et al.* 1987), musical motifs and themes (Welker 1982), painting styles (Hartley & Homa 1981), furniture art styles (Whitfield & Slatter 1979), psychiatric diagnoses (Cantor *et al.* 1980; Horowitz *et al.* 1981; Murphy & Wright 1984), physical illnesses (Brooks *et al.* 1991; Medin *et al.* 1982; Shanks 1991), worker ability, self-esteem, or thirst (Busemeyer *et al.* 1997), computer programming concepts like “sorting” or “searching” (Adelson 1985), physics problems (Chi *et al.* 1981), mathematical concepts (Armstrong *et al.* 1983), mental states like “hesitation” and “bewilderment” (Chaplin *et al.* 1988), commercial products (Loken & Ward 1990), environmental scenes like “indoors” or “outdoors” (e.g. Tversky & Hemenway 1983; Tversky 1986, 1990), event categories like “doing the dishes” or “going to a restaurant” (e.g. Abelson 1981; Barsalou & Sewell 1985; Bower *et al.* 1979; Fivush 1987; Graesser *et al.* 1979; Morris & Murphy 1990; Rifkin 1985; Ross & Murphy 1999), psychological situations, social situations, cultural situations and ideological situations (Cantor *et al.* 1982), locative categories like “in”, “above”, “below” or “beside” (Erreich & Valian 1979), “tallness” (Dirven & Taylor 1988), “highness” (Brownell & Caramazza 1978), metalinguistic categories like “simple declarative sentence” (Corrigan 1986), the categories “sentence”, “declarative sentence”, “exclamative sentence”, “imperative sentence” and “intransitive” sentence (Corrigan 1991), etc.

The specialised literature has explicitly dealt with the similarities and differences between categories quite often. The characteristics of trait categories have been compared with those of categories of states and objectual categories (Chaplin *et al.* 1988) and with those of goal-derived categories (Hoffman *et al.* 1981). In addition, the features of visual categories have been compared with those of auditory categories (Bornstein 1987). However, most studies have

compared the characteristics of natural and/or artefactual categories with some other type of categories like abstract concepts (Hampton 1981), social categories (Borkenau 1990; Cantor & Mischel 1979; Dahlgren 1985; Medin & Smith 1984; Trentin & Salmaso 1990; Wattenmaker 1995), ad-hoc categories (Barsalou 1983, 1985, 1987; Barsalou *et al* 1986; Barsalou *et al.* 1987; Mervis & Smith 1984), environmental scenes (Tversky 1990), psychiatric diagnostic categories (Cantor *et al.* 1980), psychological, ideological, social and cultural situation categories (Cantor *et al.* 1982), event categories (Barsalou & Sewell 1985; Morris & Murphy 1990), sentences (Corrigan 1991) or even computer programming concepts (Adelson 1985).

1.2. Categorisation Research with Adult Humans.

Reviewing the categorisation research conducted with human adults almost inevitably requires a revision of the different categorisation models proposed to account for the human categorisation ability. For a large number of years different categorisation models have tried to provide an account of the ability to categorise. Almost every model of categorisation shares the assumption that most categories are neither God-given or neuro-genetically wired but that individuals learn them through experience with category members.⁹ However, beyond this common assumption, few similarities exist amongst the different views. Different categorisation models disagree on the nature of learning, mental representation and the processes used in classification.

The purpose of the rest of this section is to review this extensive literature to provide the reader with an overall view of the categorisation research conducted with adult human subjects (but see Barsalou 1992a; Barsalou & Hale 1993; Homa 1984; Komatsu 1992; Lingle *et al.* 1984; Medin 1989, 1999; Medin & Barsalou 1987; Medin & Smith 1984; Mervis 1980; Mervis & Rosch 1981; Smith & Medin 1981; etc., for complementary discussions). Although there are usually differences between the specific submodels within each type, in general, we can distinguish four main types of models: classical, probabilistic, exemplar, and mixed. Probabilistic, exemplar, and mixed models represent alternatives that emerged in opposition to the problems that plagued the classical view, to which we now turn.

1.2.1. Classical Models of Categorisation.

⁹ Although categorisation researchers have traditionally assumed that people acquire knowledge primarily through experience, many theorists also assume that some knowledge has a biological basis in the cognitive system, requiring only modest experience to become activated and tuned. However, although such categories are often largely innate, no category in any organism is completely innate (Barsalou 1992a: 22). Instead, experience with the environment shapes all innate categories. The emphasis on innate biases must be tempered by the obvious fact that experience plays a significant role in the acquisition of categories (Medin & Barsalou 1987: 466-468).

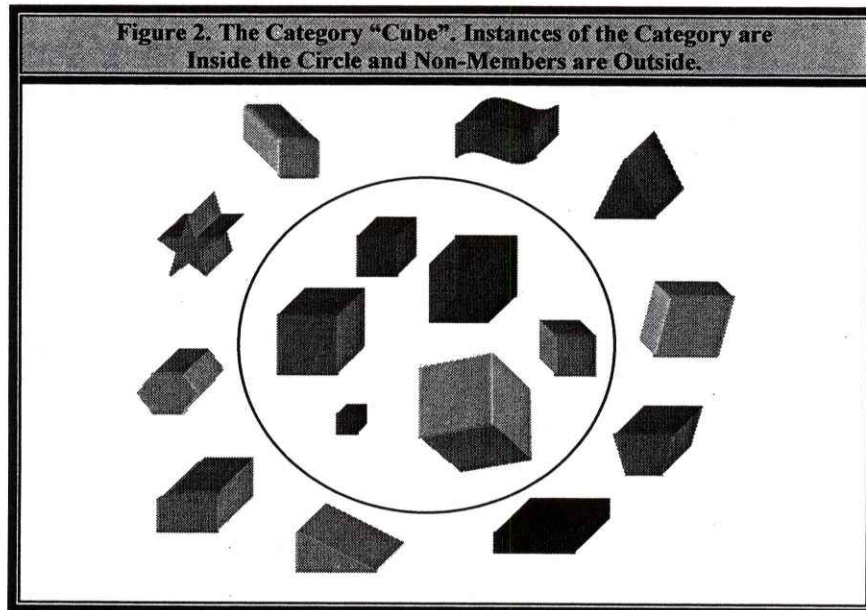
The classical (or definitional) view of concepts goes back to Aristotle. This view was, until recently, the dominant position in philosophy (see Margolis 1994; Rey 1983, 1985; Sutcliffe 1993 for recent advocacy). It was also the dominant view in psychology with the traditional concept formation and learning experiments of the behaviourists initially (e.g. Hull 1920) and the information processors later (e.g. Bourne 1966, 1970; Bruner *et al.* 1956; Hunt 1962; Neisser & Weene 1962), and amongst educational or developmental psychologists (e.g. Bruner *et al.* 1966; Clark 1973) and linguists (e.g. Chomsky & Halle 1968; Katz & Postal 1964; etc.) as well as in many other disciplines.

1.2.1.1. Learning, Mental Representation, and Storage.

Category learning consists of two aspects according to the classical view: first, the abstraction of relevant attributes that define the category from category members and second, the creation of higher order information containing the logical relationships or rules amongst those attributes. This process has been conceived of in two different ways. The extensive behaviourist psychology literature of the first half of the 20th century assumed that subjects learned defining features in a passive and gradual way. The subject was considered as a passive recipient of incoming information who did not operate on it in any essential way (e.g. Hull 1920). However, since Bruner (e.g. Bruner *et al.* 1956; Hunt 1962) researchers began to consider subjects' conscious and active hypothesis testing in the learning of relevant features and the logical rules combining them. The subject was then conceived of as an active participant that always entertained some hypothesis about the unknown concept in the process of forming concepts. Each example and non-example that the subject encountered was believed to provide a test of a current hypothesis (see also Bolton 1977: 1-5; Bourne 1966: 25-44; Hulse *et al.* 1980: 277-283; Hunt 1962: 47.102; Picas 1966: 29-155 for reviews of both orientations).

Irrespective of the position adopted, the representation of a concept or category is, according to the classical view, an abstraction consisting of a small set of features shared by all entities classified as members of the category. In addition, this abstraction is a summary description that specifies features that are singly necessary and jointly sufficient for category membership. For a single feature to be necessary and sufficient, *all* instances of the concept must share it. For example, the geometric concept "cube" is definable by the following defining features: 1) "closed" figure, 2) "six sides, 3) "with equal sides in length", and 5) "with equal interior angles". Each of these features is by itself necessary and together they are sufficient if something is to be classified as a cube. In other words, having all of these attributes is sufficient for something to be a cube, and to be a cube, it is absolutely necessary that an object have six sides, that these sides be equal in length, etc. Figure 2 shows the clear-cut boundaries separating the examples of the category "cube" (inside the circle), which comply with all the

characteristics mentioned above from other geometric figures that lack at least one of the criterial attributes and that cannot be classified as instances of the category “cube”.



Although classical models of categorisation are usually considered to require a small storage capacity, actually they require a large one, even more than exemplar models (see 1.2.3.1). This is so because a classical concept does not have necessary and sufficient conditions if there are any exceptions to them. If an exemplar does not satisfy the set of necessary and sufficient features (usually called “rule”) for a category but it is nevertheless assigned to the category by an external source, a new set of features must be computed. To ensure that there are no exceptions, memories of exemplars must then be stored to allow recomputation of rules that are not confirmed by an entity that is classified as a member of a given category. It is usually necessary to examine every previous exemplar memories when revising a rule. Consequently, classical models do not only store defining rules but must also maintain a record of all exemplars. As a consequence, the discovery of classical rules often requires extensive computation and storage capacity.

1.2.1.2. Classification.

According to the classical view, once categorisers have learned and know the category, they should be able to determine whether or not any particular entity is a member of the category simply by determining whether it possesses all the defining features of the conceptual category (or concept) or not. Categorisers simply have to test likely candidates for category membership against the category abstraction (or rule), which is the only information they

retrieve to perform classification, never any of the exemplars. This form of classification requires that each new instance meet the defining criteria associated with its category. A novel entity is assigned to a category if and only if it possesses all of the defining features of the abstract representation. Anything that does not have all of those properties is not in the class and will not be classified as a member of that class.

1.2.1.3. Decadence of the Classical View: Questioning the Paradigm.

Classical models of categorisation offer intuitively logical and appealing accounts of categorisation. However, although the classical view has always had its critics (e.g. Smoke 1932; Ryle 1949; Wittgenstein 1953), after the many centuries during which it held sway, the classical view became seriously questioned due to the many experimental results obtained since the early 1970s that appear to contradict all of its assumptions. In retrospect, the resulting subsequent decline of the classical view is not surprising because this view of categorisation was mainly a non-empirical, philosophical position arrived at on the basis of a priori speculation (Lakoff 1987a: 6). Over the centuries, it simply became part of the background assumptions taken for granted in most scholarly disciplines and taught not as an empirical hypothesis but as an unquestionable, definitional truth. However, since categorisation started to be investigated experimentally, a wealth of experimental data appeared to contradict the traditional view of categories, which ran into numerous problems to explain the new findings. In fact, the major conclusion to emerge from one of the most important books on categorisation (Smith & Medin 1981) is that most facts about categorisation obtained empirically fit alternative views of categories and categorisation like the probabilistic and exemplar views (see 1.2.2. and 1.2.3. respectively) better than classical models.

The seeming incompatibility of the classical view with different experimental phenomena can be addressed by examining the different criticisms for which empirical evidence has been gathered over the years. These include the failure to obtain defining features, the use of non-necessary features in categorisation, the existence of unclear category members, the phenomenon of typicality and typicality ratings, and the inconsistency of many taxonomic phenomena with the classical view of taxonomic organisation.

1.2.1.3.1. Failure to Specify Defining Features.

The heart of the classical view is its assumption that every concept has a set of necessary and sufficient features. However, decades of analyses by linguists, philosophers, psychologists, and others have apparently failed to turn up the defining properties of most natural and artefactual categories. Many researchers have since claimed that there is probably no set of

defining features that will determine category membership for all category members in most categories. Smoke (1932) was one of the first to argue that “common elements” in Hull’s (1920) sense were probably a fiction. In this respect, Ryle (1951) coined the expression “polymorphous concept” to refer to many philosophically important concepts (such as “thinking”) and most everyday terms (such as “solicitor”) for which there appeared to be no defining features. In a similar vein, Wittgenstein suggested that such types of concepts were the rule rather than the exception. In his well-known analysis of the category “game” (1953), he argued that finding definitional characteristics for “game” on whose basis games can be clearly distinguished from non-games is impossible as no property appears true of all games. Instead, members of the category are connected by a network of overlapping similarities or “family resemblances” in the sense that various members of the category share properties with some other games but not with all. Therefore, it is possible for two games to be members of the same class without overlapping in features. Game A may resemble (“share properties with”) game B and game C. In addition, game D may resemble games B and C. However, game D may not share any property with game A.

This state of affairs is by no means peculiar to the category “game”. Following Wittgenstein, a host of researchers in psychology, linguistics, anthropology, etc., have since endorsed the view that most categories are seldom defined in terms of a few criterial (i.e. “sufficient and necessary”) properties and provided experimental evidence for this (e.g. Dennis *et al.* 1973; Hampton 1979, 1981, 1995; Herrnstein 1985; Kempton 1981; McCloskey & Glucksberg 1979; Rosch & Mervis 1975; etc.). Mervis and Rosch (1975), for example, had experimental subjects list features for various instances of the category “furniture” and found that not a single feature was listed for every member.

In this respect, it has been claimed that the mental illusion that most categories have defining properties seems to derive from the fact that there are certain properties in concepts that provide a moderate degree of “stability” in category representations because they have been frequently processed in conjunction with a category on so many occasions that they become automatised. For example, the category “bird” automatically activates the property “flies” across contexts. However, these features, although produced by many (but rarely by all) subjects within the same context, are far from definitional in the sense of the “classical” view (e.g. Barsalou 1982; Barsalou & Medin 1986; Barsalou *et al.* 1987). However, such frequently processed features do not form “definitional cores” but “experientially-based cores” (Barsalou & Medin 1986). The latter differ from definitional cores in two respects. First, experientially-based cores are not definitional as there is, in principle, no reason why any core property should be either necessary or sufficient for category membership. Instead it is simply information that occurs frequently enough for a category in someone’s experience to become automatically activated every time the representation of a category is accessed, regardless of the current

context. Second, experientially-based cores are not necessarily static but can vary between and within individuals because differential experiences cause different properties to become automatically activated. This, however, contradicts the assumption of the classical view that what objectively defines a category never changes; according to the classical view, a category should always be represented by the same static and well-bounded packet of information across different individuals and across different contexts.

1.2.1.3.2. Use of Non-Necessary Properties in Categorisation.

According to the classical view, the possession of features which are nondefining by members of a category is irrelevant to category membership and these features should not be used in classifying an entity as a category member. However, most of the properties people list for the exemplars of most categories are not true of all exemplars. In addition, the fact that the distribution of these properties correlates with classification times strongly suggests that people use non-necessary properties to determine category membership (e.g. Hampton 1979; Malt 1993, 1994; Malt & Johnson 1992; Rips *et al.* 1973). For example, although “flies” is a feature that does not apply to all members of the category “bird” (e.g. “ostrich”, “penguin”, “duck”), it is usually used to classify an entity as a bird.

1.2.1.3.3. Unclear Category Members.

Since the classical view assumes that judgements about category membership are based on defining properties, this type of category membership judgements should lead to no errors or variations in the classification of an entity as a member of a given category. Categorisers simply have to compare the defining features of the to-be-categorised entity against the summary description of a category. If the entity meets the necessary and sufficient features of the concept, it is a category member, if it does not, it is not a member. There are no reasons to expect inconsistencies in classification.

However, this prediction of the classical view often fails. Categories do have ambiguous category members, sometimes referred to as “unclear cases”. These are exemplars that are not classified the same way 100% of the time (Medin & Barsalou 1987: 471). When deciding whether a list of instance concepts belong in a category or when asked to judge membership in questions like “is a rug a piece of furniture?”, “is a wheelchair a vehicle?”, “is a tomato a fruit?”, “is a tomato a vegetable?”, etc. people are sometimes inconsistent and consider an entity as a member of a particular category and at some other times as a non-member (Barsalou 1983; Bellezza 1984; Brownell & Caramazza 1978; Hampton 1979; Kempton 1978, 1981; Labov 1973; McCloskey & Glucksberg 1978; Smith *et al.* 1974; etc.). These inconsistencies exist



within individuals and across individuals on the same or even on different occasions. McCloskey and Glucksberg (1978), for example, studied categories like “disease”, “vehicle”, or “animal” and found out not only that people differed frequently in the entities they considered as category members, but also that a given individual frequently changed his or her mind over a one month period.

The number of such cases is, however, generally small in comparison with the number of exemplars that are clearly classified. However, the classical view precludes the existence of *any* unclear cases. Category boundaries are, according to the classical view, clear-cut and no fuzzy areas are permitted: only two degrees of membership (i.e. member and non-member) are allowed for one and the same entity on all occasions. As a consequence, the existence of even a few unclear cases is a problem that classical models do not satisfactorily account for

There may probably be a host of reasons why certain items do not receive the same category assignment on every occasion. The context in which an entity is classified appears to be an important variable. In a series of experiments involving cups and cup-like containers, Labov (1973) showed subjects line drawings of cups and other vessels (mugs, bowls, etc.) one by one and asked them to name every container. One experiment that involved what Labov called a “neutral context” had subjects classify different containers as members of the categories “cup”, “bowl”, “mug”, or “vase”. The results showed that subjects did not agree amongst one another. In subsequent experiments, however, when the context involved one of three different scenes (a coffee-drinking situation, a dinner table situation with the object filled with mashed potatoes and a scene where the objects were standing on a shelf with cut flowers in them) the context also affected subjects’ assignment of entities to different categories.

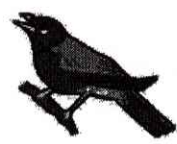
Another variable responsible for the existence of unclear cases has been shown to be the induction of positive affect. Isen and Daubman’s (1984) study suggests that subjects in whom positive affect is induced tend to include more entities in categories than subjects in whom such a positive mood has not been induced. Age also affects fuzzy category boundaries. In a study with subjects of 3, 4, 5, and 25 years of age, category boundaries, although found to be fuzzy at all ages, became reliably less fuzzy with age (Alexander & Enns 1988). It has also been found that a change in the instructions given to the subjects affects which members of a category are judged as members of that category or not (Homa *et al.* 1987). For example, when subjects employ conservative instructions (e.g. “assign members to categories only if one is reasonably certain of the choice”), they include fewer items in a category than when they employ more liberal instructions (e.g. “make efforts to assign members to categories even if one is not reasonably certain”).

1.2.1.3.4. Typicality and Typicality Effects.

Of all the experimental findings used as evidence against the classical view, perhaps the best known is the phenomenon of instance typicality, which has been referred to as “typicality”, “prototypicality”, “representativeness”, “exemplar goodness”, etc. The typicality of an instance refers to how representative of its category it is perceived to be. For example, if people are asked to judge to what extent members of a category like “bird” can be regarded as good, representative or typical examples of the category, they tend to consider “robin” or “sparrow” as more typical birds than “duck”, and “duck” as more typical than “penguin” or “ostrich”, etc. (see table 1 and figure 3 below). In other words, typicality refers to a continuum of category representativeness, ranging from the most typical members of a category and continuing through less typical members to the most atypical ones. Typicality even extends beyond the category to non-members. “Butterfly”, for example, is a better non-member of the category “bird” than “helicopter” is. “Helicopter” is, in turn, a better non-member of the category “bird” than “chair”. Consequently, non-members of a category also vary in how good non-members they are of the category they do not belong to (Barsalou 1983, 1985, 1987; Barsalou & Medin 1986).

Table 1. Typicality Ratings for Some of the Members of the Category “Bird”. 1= very typical, 7 = very atypical (Rosch (1975b)).			
Category member	Mean typicality ratings	Category member	Mean typicality ratings
<i>Robin</i>	1.02	<i>Owl</i>	2.96
<i>Sparrow</i>	1.18	<i>Swan</i>	3.16
<i>Dove</i>	1.46	<i>Flamingo</i>	3.17
<i>Parakeet</i>	1.99	<i>Duck</i>	3.24
<i>Eagle</i>	1.75	<i>Peacock</i>	3.31
<i>Hummingbird</i>	1.76	<i>Egret</i>	3.39
<i>Hawk</i>	1.99	<i>Turkey</i>	4.09
<i>Parrot</i>	2.07	<i>Ostrich</i>	4.12
<i>Pheasant</i>	2.69	<i>Penguin</i>	4.53
<i>Toucan</i>	2.95	<i>Bat</i>	6.15

Figure 3. Instances of the Category "Bird" from Most Prototypical to Least Prototypical (According to Rosch 1975b).



Robin



Sparrow



Bluebird



Dove



Parakeet



Eagle



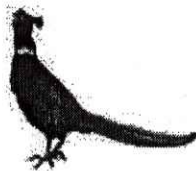
Hummingbird



Hawk



Parrot



Pheasant



Toucan



Owl



Swan



Flamingo



Duck



Peacock



Egret



Turkey



Ostrich



Penguin



Bat

Traditionally, the standard procedure for assessing subjects' ratings of the typicality of items as members of categories is a 7-point rating scale technique (e.g. Rosch 1973b, 1975b; Rosch & Mervis 1975) ranging from 1 (=very good example), through 4 (=moderately good/bad example), to 7 (=very bad example). What subjects are instructed to do is to write a number next to different members of a specific category listed on a sheet. This number represents the extent to which they feel each member is typical of its category. In this respect, a large body of research on typicality judgements using this particular scale or modified versions of it shows that, under identical experimental conditions, subjects' typicality ratings are highly reliable statistically. In other words, people agree on which instances of categories are more or less typical.¹⁰

Typicality ratings have been obtained in all types of categories like natural or artefactual categories (e.g. Hampton & Gardiner 1983; Malt 1994; McCloskey & Glucksberg 1978; Rips *et al.* 1973; Smith *et al.* 1974), perceptual categories like *colours* (Nosofsky 1988b; Rosch 1973a, 1975c), product categories (Loken & Ward 1990), goal-derived categories (Barsalou 1981, 1983, 1984, 1985; Barsalou *et al.* 1986), abstract concepts like *work* or *crime* (Hampton 1981), linguistic categories (Corrigan 1986), personality trait categories (Borkenau 1990; Buss & Craik 1983; Cantor & Mischel 1977; Chaplin *et al.* 1988; Hampson 1982; Isen *et al.* 1992; Read *et al.* 1990a, 1990b; Wojciszke & Pienkowski 1991), "person" categories (Cantor & Mischel 1979a, 1979b; Dahlgren 1985), art styles (Whitfield & Slatter 1979), psychiatric categories (Cantor *et al.* 1980), locative categories (Erreich & Valian 1979), computer programming categories (Adelson 1985), "emotion" categories (Cantor *et al.* 1979b; Fehr *et al.* 1982; Shaver *et al.* 1987), sentences exemplifying proverbs (Honeck *et al.* 1985), phonetic categories (e.g. Hodgson & Miller 1996; Volaitis & Miller 1989, 1992; Wayland *et al.* 1994). In fact, every human category studied has been shown to be capable of eliciting typicality ratings from subjects.

¹⁰ Typicality ratings reflect the current processing of a category, not the supposedly invariant cognitive or conceptual structure of a category in long-term memory. Typicality ratings are simply a current rank ordering of different category instances from most to least typical. Typicality does not carry any representational assumptions so it does not provide any specific theory of mental representation (Rosch 1978; Barsalou 1987) despite early and nowadays widely abandoned views to the contrary (e.g. Rosch 1975b). One of the clearest indications that typicality ratings do not represent an invariable conceptual structure is that typicality ratings vary as a consequence of a host of factors like the linguistic context in which category instances are processed (Roth & Shoben 1983), the phonetic context (Miller & Volaitis 1989, Miller *et al.* 1997; Volaitis & Miller 1992; Wayland *et al.* 1994), the point of view subjects adopt (Barsalou & Sewell 1984; Barsalou *et al.* 1986), the mood in which they are (Isen & Daubman 1984; Isen *et al.* 1992; Niedenthal *et al.* 1999), the time interval between two different rating sessions (Barsalou *et al.* 1986; Hampton & Gardiner 1983), the level of abstraction in a taxonomy at which the category is processed (Mervis & Roth 1981; Roth & Mervis 1983; Smith *et al.* 1974), the processing of an entity as a member of an isolated category or as a member of a category that is the product of conceptual combination (Hampton 1982, 1987; Medin & Shoben 1988; Osherson & Smith 1981, 1982; Smith & Osherson 1984, 1987; Smith *et al.* 1988; etc.), the age of subjects (Bjorklund *et al.* 1983; Duncan & Kellas 1978; Keller 1982; Rosch 1973a) or the cultural-linguistic background of subjects (Atran 1998; Hampton & Gardiner 1983; Lin & Schwanenflugel 1990; Schwanenflugel & Rey 1986; Segalowitz & Poulin-Dubois 1990). None of these studies shows that typicality is an arbitrary phenomenon. They merely show that when subjects make judgements of typicality, they draw upon many different sources of knowledge, depending on the circumstances and that the determination of typicality is a highly flexible dynamic and context-dependent process (Barsalou 1985, 1987, 1992a: 176-177; Segalowitz & Poulin-Dubois 1990).



One of the reasons why typicality has been the focus of so much interest and research lies in its strong influence on performance in a wide range of experimental tasks or naturally-occurring phenomena of roughly three main kinds: cognitive processing and memory, language use and communication, and finally, category learning and conceptual development. Typicality has been shown to be related to virtually all of the major dependent variables used as measures in psychological research. The effects of typicality on those variables are usually called “(proto)typicality effects”.

Typicality effects have been obtained with different tasks related to cognitive processing and memory. The typicality of a category member seems to predict how long it takes someone to fixate on the names of category instances while reading sentences (Carroll & Slowiaczek 1986), how long it takes to verify propositions in sentence verification tasks including category members (Keller 1982) and how long it takes someone to classify an item as a member of a category in speeded category verification tasks (e.g. Armstrong *et al.* 1983; Duncan & Kellas 1978; Glass & Meany 1978; Hampton 1979; McCloskey & Glucksberg 1979; McFarland *et al.* 1978; Mervis *et al.* 1976; Rips *et al.* 1973; Rosch 1973b, 1975b, 1975c; Rosch *et al.* 1976; Smith *et al.* 1974; Smith *et al.* 1978).¹¹ Typicality also predicts the direction in similarity judgements between category members varying in typicality (e.g. Tversky & Gati 1978) and in equivalence judgements (Rosch 1975a); it predicts strength of inductive (e.g. Osherson *et al.* 1990; Rips 1975) and deductive (Cherniak 1984) inferences about category members, perceived probability that instances belong to categories (e.g. Shafir *et al.* 1990), rated degree of truth value of category membership propositions (e.g. Oden 1977) and ease of recall (e.g. Bjorklund *et al.* 1982; Bjorklund *et al.* 1983; Cantor & Mischel 1979b; Greenberg & Bjorklund 1981; Heider 1972; Keller & Kellas 1978).

Typicality has also been shown to be related to several phenomena related to language use and communication. Typicality predicts, for example, acceptance of linguistic “hedges” (Kempton 1978, 1981; Lakoff 1973)¹², the extent to which the names of category members can be substituted for their related category name in a sentence (Rosch 1977), the order and probability of production of category members in a free listing task (e.g. Battig & Montague 1969; Hampton 1979; Hampton & Gardiner 1983; Mervis *et al.* 1976; Rosch & Mervis 1975; Rosch *et al.* 1976; Wilkins 1971), in more naturalistic situations (Kelly *et al.* 1986) or the order in which a short list of exemplars to denote a superordinate category term in American Sign Language is used (Newport & Bellugi 1978). Typicality also predicts order of selection of members of categories (e.g. Heider 1972; Whitfield & Slatter 1979; etc.) and which category members will be named with a higher superordinate category name in parents’ (or caretakers’) input to children (White 1982).

¹¹ For other variables determining reaction time in such tasks see Chumbley (1986).

¹² Hedges are qualifying terms like “true”, “technically”, “virtually”, “sort of”, etc.

The third main group of variables for which typicality has been shown to be a good predictor of performance is that related to developmental and/or category learning phenomena. First, a wide variety of experimental tasks like non-verbal sorting, non-verbal selection, picture-naming, name-recognition, etc. have shown that typicality predicts the order in which children learn category members (e.g. Anglin 1977; Bauer *et al.* 1995; Bjorklund *et al.* 1983; Blewitt & Durkin 1982; Carson & Abrahamson 1976; Heider 1971; Lin *et al.* 1990; Mervis 1980; Mervis & Pani 1980; Mulford 1979; Rosch 1973b; Rosch *et al.* 1976; White 1982) and adults learn a new (artificial) category (e.g. Heider 1972; Mervis *et al.* 1975; Rosch & Mervis 1975; Rosch *et al.* 1976).¹³

This extensive body of literature on typicality ratings and typicality effects is one of the most powerful experimental findings against the classical theory of categorisation. On the strictly classical view, it simply makes no sense to ask “to what extent” a member of a category is a good member of its category. In principle, since each category is defined by necessary and sufficient conditions, no member that satisfies such conditions should be a better example of the category than any other. Any one stimulus which fits the definition of the concept (possesses the defining attributes in the correct combination) should be as good an example of the concept as any other, that is, should be automatically included as a full and equal member as any other in every respect, including its representativeness as a member of the category. Thus, no category member exemplifies its category better than any other member (e.g. Mervis & Rosch 1981: 95; Rosch 1973a: 328-329, 1973b: 112-113, 1977: 1980). For example, if the concept is “red square”, with size as an irrelevant attribute, and all squares are homogeneous with respect to number of sides and angles, it makes no logical sense to ask the subject whether a small or a large red square is a better example of the concept ‘red square’. However, the large body of evidence mentioned demonstrates that these predictions of the classical view of categorisation are wrong. Paradoxically, even degrees of representativeness are obtained with the so-called “logically-defined” artificial categories like geometric designs or figures (Bourne 1982; Nosofsky 1991d) or mathematical concepts like “even number” (Armstrong *et al.* 1983) that

¹³ Although investigators agree on the importance of typicality, they do not concur on its explanation. Two main types of determinants of typicality are usually distinguished: materialistic or non-materialistic factors. The former refer to either the material structure of the human perceptual apparatus, or the material characteristics (structural, statistical or referential) of the category instances while the latter refer to socio-cultural, linguistic or conceptual factors (Geeraerts 1988). Materialistic explanations refer to purely structural factors (e.g. Boster 1988; Hampton 1979, 1981; Homa *et al.* 1981; Malt & Smith 1984; McCloskey & Glucksberg 1979; Mervis & Rosch 1981; Rips *et al.* 1973; Rosch 1975b, 1978; Rosch & Mervis 1975; Rosch *et al.* 1976; Roth & Mervis 1983; Smith *et al.* 1974; Tversky & Hemenway 1983), purely physiological factors (e.g. Heider 1971, 1972; Nathan 1986; Rosch 1973a, 1973b, 1974, 1975c), or statistical factors (Dahlgren 1985; Malt & Smith 1982; McCloskey 1980; Mervis *et al.* 1976; Nosofsky 1988b; Rosch *et al.* 1976). Non-materialistic factors include reference to perceived correlated attributes (e.g. Malt & Smith 1984), perceived frequency of occurrence of category members (Barsalou 1985; Loken & Ward 1990), perceived word frequency (Segalowitz & Poulin-Dubois 1990), perceived familiarity (Ashcraft 1978; Barsalou 1985; Glass & Meany 1978; Hampton & Gardiner 1983; Lin & Schwanenflugel 1990; Loken & Ward 1990; Malt & Smith 1982; McCloskey 1980; Schwanenflugel & Rey 1986), ideals (Barsalou 1981, 1983, 1985, 1987, 1991; Borkenau 1990; Chaplin *et al.* 1988; Read *et al.* 1990a, 1990b), social salience (Whitfield & Slatter 1979), idealised cognitive models (Lakoff 1987a), etc. Researches have also found that different factors may determine the typicality of a particular category at the same time (e.g. Barsalou 1981, 1985; Barsalou & Sewell 1985; Loken & Ward 1990; Nosofsky 1988a), that different factors may determine typicality in different categories (see e.g. Barsalou 1985 and Hampton 1981) or at different taxonomic levels (e.g. Loken & Ward 1990) and that the determinants of the typicality of the members of a particular category may vary depending on the context in which the category is processed (Barsalou 1985, 1987).



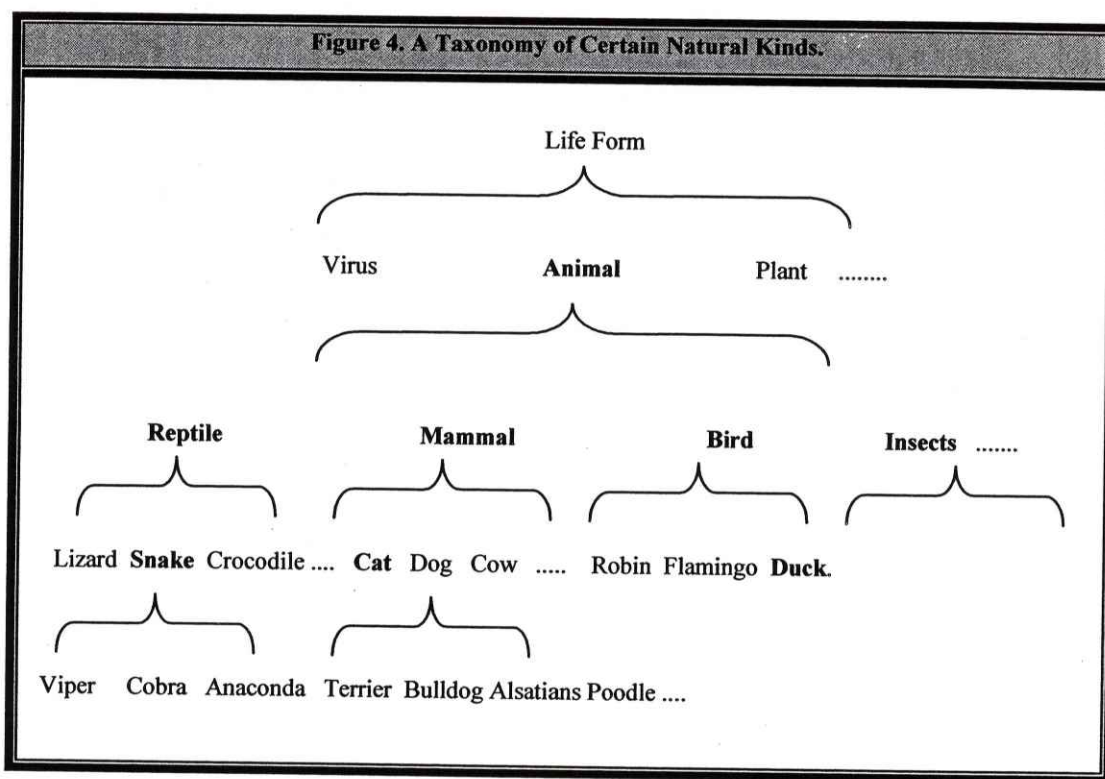
have traditionally been the banner of the classical view and the most likely candidates for having defining attributes.

1.2.1.3.5. Taxonomic Phenomena.

One striking characteristic of human categories is that they do not usually exist independently of one another in memory. Categories tend to be organised into systems where they are related to each other in various ways. In this respect, it should be noted that hierarchical organisation is a particularly efficient way to store information and much of our cognitive activity seems to be hierarchically organised (Neisser & Weene 1962: 640). However, two essential but different hierarchical systems can be distinguished: “partonomies” and “taxonomies” (Barsalou 1992a; Berlin 1978, 1992; Callanan 1985; Markman 1989; Murphy & Lassaline 1997; Tversky 1986, 1989, 1990; Tversky & Hemenway 1984). A partonomy reflects a subdivision of knowledge into parts. A taxonomy reflects a subdivision of knowledge into kinds. The primary difference between them is to be found in their different organising relations.¹⁴ In taxonomies concepts are organised by the “type” or “is-a-kind-of” relation, which specifies that one concept is a type, instance, or subset of another. Thus, a robin is a type of bird, which is a type of animal. Another way to refer to a type relation is “class inclusion”. Taxonomic organisation is usually conceived of as a vertical axis in which there are different levels of abstraction which are more inclusive as we move upwards and less inclusive as we move downwards. Each level of abstraction is, in turn, occupied by different contrasting categories. Each category within a taxonomy includes others (unless it is the lowest level category). For example, the category “animal” includes not only “mammals”, but “birds” and “reptiles” as well. On the next level, the class “mammal” comprises not only “dogs”, but “cats”, “cows”, “lions”, “elephants” and “mice”. Still further down, the class “dog” includes contrasting categories like “terriers”, “bulldogs”, “Alsations”, “poodles”, and other breeds of dogs (see figure 4). On the contrary, partonomies do not reflect class inclusion relations. Partonomic concepts are organised by the “part relation”, which specifies that one concept represents a part

¹⁴ Other important differences exist between partonomies and taxonomies. First, a partonomy is a consequence of an “analytic” attitude, of a top-down investigation, in which a whole is decomposed into parts. In contrast, a taxonomy seems more commonly to be a consequence of a “synthetic” attitude, more likely to be a bottom-up enterprise in which exemplars are grouped on the basis of common and distinctive features (Tversky 1990). Second, categories in taxonomies allow cross-classification. A given entity may be categorised in many alternative ways (e.g. as a vehicle, as a car, as a Volkswagen). In a partonomy cross-classification is not possible as the principle of “mutual exclusivity” holds. If an object is given a name, it cannot be given another (Markman 1984, 1987). Third, in taxonomic hierarchies, property inferences are usually permissible (Corrigan 1991; Markman & Hutchinson 1984; Murphy & Lassaline 1997; Tversky 1986, 1989; Tversky & Hemenway 1984). If apples are spherical, for example, then so are subcategories of apple. Property inferences are not usually appropriate in partonomies: cars may be beautiful but their pistons probably are not. However, taxonomies and partonomies also share some characteristics. First, the number of levels into which they decompose is similar (maximally six) (Brown *et al.* 1976: 81). Second, the phenomenon of typicality exists in both types of knowledge structures. Whereas in taxonomies some members of categories are better examples of the category than others, in partonomies some parts are also better parts than others (Hemenway 1981; Tversky 1986, 1989, 1990; Tversky & Hemenway 1984). For example, “arms” and “legs” may be basic parts of the “body” because they are perceptually and functionally distinct from other parts of the body. Third, both types of organisation are asymmetric relations (“dog” is a kind of “animal”, but “animal” is not a kind of “dog”; similarly, a “finger” is a part of a “hand” but a “hand” is not a part of a “finger”). Finally, the vertical phenomenon of a basic level is present in both (Rifkin 1985; Tversky 1989, 1990).

of another. Thus, in the well-known body part partonomy, a finger is a part of a hand, which is a part of the body. Fingers are not included in the class of hands, which are not included in the class of bodies.



The study of folk taxonomies of speakers in non-western pre-scientific and industrialised Western societies has yielded a view of taxonomic organisation which differs from the classical view of taxonomic organisation in important ways, showing a discrepancy between the classical theory of categories and a cognitively adequate theory of categories. In this respect, certain factors related to the taxonomic organisation of categories have also been mentioned as evidence against the assumptions of the classical theory.

First, classical taxonomies (particularly traditional scientific taxonomies) appear to have an excessive number of levels (Ungerer & Schmid 1996). An obvious reduction in taxonomic complexity (of 13 major levels in the scientific classification to six (e.g. Berlin 1978, 1992; Berlin *et al.* 1973; Brown *et al.* 1976) or three (Rosch 1978; Rosch *et al.* 1976a) has been shown to reflect the actual level of abstraction in people's folk taxonomies. According to Berlin and his co-workers, from the top to the bottom of a taxonomy, the levels are called "unique beginner", "life form", "generic", "specific", and "varietal". Rosch and her co-workers call their three levels "superordinate" (which includes Berlin's first two levels), "basic", and "subordinate"

(which includes Berlin's last two ones).¹⁵ In short, scientific classifications may be fascinating in their complexity and rigidity, but they are not suitable for human categorisation. In a sense, scientific taxonomies are neither mind-sized nor mind-oriented (Ungerer & Schmid 1996: 60-62).

Second, classical taxonomies are characterised by their elegance and exhaustiveness: the elaboration of taxonomies is complete with no missing levels in any taxonomy. This is so because the aim of scientific classifications is to provide a complete system of taxonomic levels. However, anthropological work has shown that complete elaboration of folk taxonomies seldom occurs. The experiential folk taxonomies emerging from everyday categorisation are neither complete nor fully consistent and may very well be prepared to skip some levels. Folk taxonomies, unlike scientific taxonomies, are characterised by gaps, inconsistencies and alternative paths. Higher or lower levels seem to have a subsidiary status, because they are not fully developed where there is no need for additional categorisation (Ungerer & Schmid 1996: 66). In Tzeltal, for example, the categories "corn" and "bean", in the middle of a taxonomic hierarchy, are not included in any superordinate category (Berlin 1978, 1992; Berlin *et al.* 1973; see also Brown *et al.* 1976).

Third, the classical view assumes that the summary description of a concept like "robin" always includes or consists of the defining properties of its superordinate (e.g. "bird"), plus those that distinguish it from other members at the same level of abstraction like "eagle", "sparrow", etc. In addition, the category "bird" includes all the properties of its immediate superordinate (i.e. "animal") plus those that also distinguish it from other members of "animal" at its own level like "mammal", "reptile", etc. To put it differently, as we move down any hierarchy, each category is supposed to possess exactly the features of the immediately higher category, plus one or more additional distinguishing features. Category members at the same level of abstraction all share the features of the immediately higher category, but each is distinguished from the other categories at the same level by the presence of certain features. So the defining features of a concept are nested, at least theoretically, within those of its subordinates. This means that a concept like "robin" has more common properties and fewer distinctive ones with its immediate superordinate (i.e. "bird") than with its distant one (i.e. "animal").

However, this does not always seem to be the case. Evidence suggests that a substantial degree of nesting occurs, there is by no means complete nesting, that is, not all of the properties of a superordinate concept are included in subordinate ones. For example, not all features of a general category are listed for a subset if that subset is an atypical member of the category; "chicken", for instance, is an instance of the category "bird" (though an atypical one) but not all

¹⁵ Though researchers may differ as to the number of abstraction levels, any taxonomy is best considered as a continuum of differentiation with no strict layers on which any category falls (e.g. Murphy & Brownell 1985; Rosch *et al.* 1976; Tanaka & Taylor

the features of “bird” (e.g. “flies”, “small”) are listed for “chicken” (e.g. Rosch *et al.* 1976). In short, one category at a particular level of categorisation does not always share all of the features that define the immediately superordinate category (Rosch & Mervis 1975; Sloman 1997). In addition, although the classical view implies and predicts that a particular concept like “robin” should always be judged more similar to an immediate superordinate one like “bird” than to a distant one like “animal”, this prediction works on some occasions but not all (e.g. Collins & Loftus 1975; McCloskey 1980; McCloskey & Glucksberg 1979; Rosch *et al.* 1976; Roth & Mervis 1983; Smith *et al.* 1974). Thus, for example, “chicken” and “duck” are consistently rated as more similar to “animal” than to “bird” (Rosch & Mervis 1975; Rosch *et al.* 1976).

Fourth, the perfect nesting of category inclusion relations implied by the classical view does not always occur. Although category organisation usually appears to contain some form of higher-order clustering, it is not always transitive (e.g. Collins & Loftus 1975; Hampton 1982, 1988a; Sloman 1997; Smith *et al.* 1974). Hampton (1982, 1988a) has also shown intransitivity in the categorisation of everyday objects. He found, for example, that people affirmed that “a car headlight is a kind of a lamp” and that “a lamp is a kind of furniture” but not that “a car headlight is a kind of furniture”.

Finally, the strongest and most widely investigated taxonomic phenomenon presented as evidence against a classical view of categorisation is the finding that a particular level of specificity, which is in the middle of a general-to-specific taxonomic hierarchy, enjoys psychological salience or primacy. This is the “generic” level (in ethnobiological terms) or “basic” (in cognitive psychological terms) level.¹⁶ For example, the basic level of abstraction in the hierarchy “animal-mammal-dog-poodle-European poodle” is “dog”, and this level of abstraction apparently has psychological primacy.

For a large number of years, investigators have showed that the basic level has a special status in a variety of tasks and this has been shown not only with natural or artefactual categories (e.g. Rosch *et al.* 1976; Rosch 1978) but also with artificial categories (Mervis & Crisafi 1982; Murphy & Smith 1982), environmental scenes (Tversky 1986, 1990, 1991; Tversky & Hemenway 1983, 1984), event categories (Morris & Murphy 1990; Rifkin 1985),

1991; Ungerer & Schmid 1996).

¹⁶ Researchers have been interested in discovering what determines that a given category is “basic”. In general, there are two main types of determinants: structural/perceptual-cognitive and cultural-epistemic. Different explanations of the basic level emphasise that the basic level is determined by the structure of the world as it is perceived and processed by cognitive systems (see e.g. Corter & Gluck 1992; Jolicoeur *et al.* 1984; Jones 1983; Lin *et al.* 1997; Medin 1983; Mervis & Crisafi 1982; Mervis & Mervis 1982; Mervis & Rosch 1975, 1981; Murphy 1991; Murphy & Brownell 1985; Murphy & Smith 1982; Rosch 1976, 1977, 1978; Rosch & Mervis 1975; Rosch *et al.* 1976). However, it has been argued that the basic level is not solely determined by features in the environment and our cognitive and perceptual evolutionary capacities. On the contrary, the basic level has been shown to vary as a function of general cultural significance (Berlin *et al.* 1973; Berlin 1992; Dougherty 1978; Rosch *et al.* 1976; Stross 1973) as well as by individual familiarity, expertise or knowledge (Honeck *et al.* 1987; Medin *et al.* 1997; Rosch *et al.* 1976; Tanaka & Taylor 1991). In short, the basic-level is determined by both the structure of the world and by the contributions of the human perceiver or categoriser like his/her goals, culture, expertise, knowledge, etc. (Dougherty 1978; Mervis 1980: 291-292). Structural factors as well as non-structural (cultural and knowledge) factors interplay to define for a given subject, or a population of subjects the basic level in a given taxonomy.

social, ideological, cultural and psychological situations (Cantor *et al.* 1982), psychiatric diagnoses (Cantor *et al.* 1980), social categories like person categories and trait categories (Brewer *et al.* 1981; Cantor & Mischel 1979; Dahlgren 1985; John *et al.* 1991), emotions (Fehr & Russel 1984; Shaver *et al.* 1987), computer programming concepts (Adelson 1985), sentences (Corrigan 1991), etc.

The special centrality of the basic level is revealed in experimental tasks of different sorts. Some tasks reflect the contents of category knowledge and inference drawing. The basic level is the level at which subjects list more attributes for category members (Horton & Markman 1980; Mervis & Greco 1984; Mervis & Rosch 1981; Murphy & Brownell 1985; Rosch 1978; Rosch *et al.* 1976).¹⁷ The richer attribute structure or feature information that basic level categories possess may be the reason why the basic level is the level at which more inferences can be drawn, in particular, in comparison with superordinate categories (Gelman & Markman 1986; Gelman & O'Reilly 1988).

The basic level is also the highest level at which it is easiest to form a relatively concrete mental image of an average member of the category (in the absence of that object) which is isomorphic to an average category member, an ability known as "imaging capacity" (Bolton 1977: 56). People have mental images of basic-level categories like "chairs" but they do not have abstract mental images of superordinate categories like "furniture" that are not images of basic-level objects like chairs, tables, beds, etc. (Rosch 1976, 1978; Rosch *et al.* 1976).

Some of those tasks reflect our perception of objects. These are tasks that depend on the (shape) appearance of objects. The basic level is the highest level in which category members have similar overall perceived shapes and, as a consequence, the average shape of a number of, for example, chairs is still recognisable or identifiable as a chair (Rosch 1976, 1978; Rosch *et al.* 1976).¹⁸ Other tasks reflect people's behaviour towards members of categories. In this respect, the basic level is the highest level in a taxonomy at which a person uses similar motor actions or movements for interacting with category members (Rosch *et al.* 1978).¹⁹ Still, other tasks reflect communication about objects. Basic-level categories (and basic-level category names) are primarily used when identifying objects in controlled free-naming tasks (e.g.

¹⁷ On the contrary, fewer attributes are listed of category members at the superordinate level ("furniture", "vehicle", "mammal") and there is virtually no increase for subordinate categories (e.g. "rocking chair", "sports car", "retriever") over the basic level unless expert knowledge is developed in relation to subordinate categories (e.g. Tanaka & Taylor 1991). In addition, it has been found out that at least for natural and artefactual categories, most of the attributes listed for both basic-level and subordinate categories refer to physical "parts" like "arms", "legs", "eyes", etc. (e.g. Hemenway 1981; Mervis & Greco 1984; Tversky 1986, 1990, 1991; Tversky & Hemenway 1983, 1984, 1991). However, parts are neither necessary nor sufficient for establishing a basic-level structure (Murphy 1991a, 1991b). The few attributes listed for superordinate categories are abstract attributes that refer to the functions of objects.

¹⁸ On the contrary, the average shape of a number of superordinate objects is not readily recognisable as a piece of furniture. Members of superordinate categories do not share a common shape. Also, some gain in similarity of shapes occurs for subordinate category members (e.g. two kitchen chairs) but it is so small when going from the basic to the subordinate level that the basic level is again preferred (Rosch *et al.* 1976).

¹⁹ Although it is easy to describe the motor movements we perform when interacting with all members of a basic level category like "table", it is very difficult to describe the ways in which we habitually use or interact with superordinate categories as a whole (e.g. "furniture" in general). In addition, subjects behave in a very similar way to members of subordinate categories, but no more movements are made in common to subordinate than to basic level categories (Rosch 1978; Rosch *et al.* 1976).



Jolicoeur *et al.* 1984; Murphy & Brownell 1985; Murphy & Wisniewski 1989a; Rosch *et al.* 1976; Smith *et al.* 1978; Tanaka & Taylor 1991). When given an unlabelled picture of, say a Ford sedan, most people identify it spontaneously as a “car” rather than as a “vehicle” or as a “Ford”) when no need for specificity or generality is required.²⁰ In addition, such identifications are usually faster at the basic level than at any other abstraction level. In more naturalistic situations like normal everyday conversation, basic-level category names are also more frequently used (Anglin 1977; Berlin *et al.* 1973; Brown 1958, 1976; Cruse 1977; Downing 1977).

Research on folk taxonomic hierarchies has revealed developmental differences in the way in which these two types of concepts might be acquired. Some developmental psychologists have hypothesised that basic-level categories are learned first and that they are easier to learn (Anglin 1977; Blewitt 1983; Blewitt & Durkin 1982; Callanan 1985; Daehler *et al.* 1979; Horton & Markman 1980; Markman *et al.* 1980; Mervis 1980, 1987; Mervis & Crisafi 1982; Rosch *et al.* 1976; Tversky 1985, 1989; Waxman *et al.* 1991).²¹

In a word, it appears that one of the levels of organisation is most basic, fundamental or functional in classifying entities into an optimal number of differentiable categories than either higher levels called superordinate or lower levels called subordinate. These findings pose many problems to the classical view.²² According to this view, there should be no level of categorisation in taxonomic hierarchies that is more “basic” than any other. There is no reason for assigning special status to any particular level of categorisation. As Mervis points out:

Because, according to the traditional theory, category divisions are arbitrary, the theory does not discuss the different levels of generality (e.g. rocking chair vs. chair vs. furniture) at which objects may be categorised. All these levels are implicitly treated as the same; no one level is more reasonable than another.

Carolyn Mervis (1980: 285).

²⁰ This is so unless the to-be-identified object (e.g. a chair) has to be categorised as part of a scene or context, such as living room with a sofa, tables, and lamps in which case categorising the object at the superordinate level (e.g. “furniture”) is just as fast as categorising the object at the basic level (i.e. “chair”) (Murphy & Wisniewski 1989a) or when subjects possess a high degree of expertise, in which case spontaneous naming of entities occurs at the subordinate level (Tanaka & Taylor 1991). In general, the need for specificity or generality in the information conveyed may require the use of subordinate or superordinate level category names (Barsalou & Billman 1989; Cruse 1977; Rosch *et al.* 1976).

²¹ Several studies (e.g. Lassaline *et al.* 1992; Mervis & Crisafi 1981; Murphy & Smith 1982; Murphy & Brownell 1985; etc.) have ruled out the possibility that the basic level is due to linguistic factors like word length and frequency, which reflect properties of the category names rather than properties of conceptual representations. However, basic-level names have special characteristics that differentiate them from the category names of superordinate and subordinate categories, like being shorter, underived, morphosyntactically regular, etc. (see Berlin 1978, 1992; Berlin *et al.* 1973; Brown 1958, 1976; Brown *et al.* 1976; Lakoff 1987: 33, 46; Mervis & Rosch 1981 for more lengthy discussions), or the first to be learned developmentally (Anglin 1977; Blewitt & Durkin 1982; Dougherty 1978; Mervis 1980, 1984; Mervis & Mervis 1982; Poulin-Dubois *et al.* 1995; Rescolda 1980; Rosch *et al.* 1976; Shipley *et al.* 1983; Stross 1973; White 1982) or primarily used by parents or caretakers in their speech to children (Anglin 1977; Blewitt 1983; Brown 1956, 1976; Callanan 1983, 1985; Poulin-Dubois *et al.* 1995; Shipley *et al.* 1993; White 1982).

²² Although most research on taxonomic hierarchies has focused on the special status of basic-level categories, some researchers have also examined both the superordinate and subordinate levels of abstraction. This research has dealt with the reasons why superordinate categories (e.g. Murphy & Wisniewski 1989a; Ungerer 1994; Ungerer & Schmid 1996: 77-79) and subordinate categories (Ungerer 1994; Ungerer & Schmid 1996: 78-79) are created and the different types of superordinates (Markman 1985; Wisniewski *et al.* 1996) and subordinates (Biederman *et al.* 1999; Jolicoeur *et al.* 1984; Murphy & Brownell 1985) that have conceptual significance.



1.2.1.4. The Utility of the Classical View.

The criticisms against the classical view are very strong, but none of them, some researchers believe, undermines it in a way that the classical view has to be abandoned completely (Medin & Smith 1984; Smith & Medin 1981). In fact, replies to such criticisms are occasionally made by advocates of the classical view. Due to its long tradition, many researchers have tried to defend the classical view in different disciplines like cognitive psychology (e.g. Armstrong *et al.* 1983; Gleitman *et al.* 1983; Martin & Caramazza 1980), philosophy (e.g. Margolis 1994; Rey 1983, 1985; Sutcliffe 1993), linguistics (e.g. Wierzbicka 1990), etc. These authors have generally argued that the classical view has been rejected prematurely.

The most promising modern attempt to salvage the classical view of categories, which developed as a reaction to the experimental data adduced against it, appeared to be the well-known “core-identification procedure” classical model of categorisation (e.g. Armstrong *et al.* 1983; Gleitman *et al.* 1983; McNamara & Sternberg 1983; Medin & Smith 1984; Rey 1983, 1985; Smith & Medin 1981; Smith *et al.* 1984; Smith *et al.* 1974; Osherson & Smith 1981, 1982). This modern version of the classical view, also known as the “binary model” of concept structure (Hampton 1988, 1991, 1993), proposes that there are two kinds of features in a concept representation, the “core” defining features which work according to the classical model, augmented by a set of non-defining but characteristic features. For example, the defining features of concepts like (human) “male” and “female” might have to do with genetics (e.g. “producing gametes (spermatozoa) that can fertilise female gametes (ovata)” in the case of male). Core properties comprise a static rule or definition that ultimately determines category membership. However, to quickly classify someone as a male or female we might rely on non-defining characteristics such as hair length, height, facial expression, and clothing that represent a mixture of secondary sexual characteristics and cultural conventions. In addition, reliance on non-defining, “identification” features would explain most of the experimental findings adduced against the classical view like typicality ratings and effects, the existence of unclear cases, etc.

This modern version of the classical view is not devoid of problems. It is, for instance, paradoxical that the core-identification procedure model appeared to be promising because it is dependent on nonnecessary (“identification”) features for classification, the backbone of probabilistic models of categorisation, which are a set of models proposed as alternatives to the classical view (see 1.2.2.). In other words, the core-identification model appears to be promising because it moves toward the probabilistic view. In addition, one of the most fatal flaws of the core-identification procedure is that it is almost impossible to distinguish experimentally between core (or defining) and identification (merely characteristic, non-defining) features (e.g.

Barsalou & Medin 1986; Barsalou *et al.* 1987; Malt 1994; Malt & Johnson 1992; Murphy 1995; see also Smith *et al.* 1984). As Malt points out:

The hybrid proposals present essentially the same paradox that the original defining features view did: while the notion of core features provides an attractive answer to certain questions about category membership, there is no direct evidence for such features for most concepts.

Barbara C. Malt (1990: 290).

Given that the assumptions of classical models of categorisation appear to be so problematic when confronted with the substantial experimental evidence against them, theorists have frequently wondered why the classical view still looks so intuitively compelling and why most of us have the initial intuition that our concepts conform to the classical view and have defining features (Medin *et al.* 2001: 372). Perhaps, a possible reason may be that the classical view is both our everyday folk theory of what a category is and the principal technical theory (Lakoff 1987: 5). Medin and Ortony (1989) tried to explain the intuition that the classical view is our folk theory of categories and categorisation with the notion of “psychological essentialism”. This view does not imply that things have defining features or “essences”, but rather that people’s conceptual representations of entities reflect such a belief (see also Malt 1990; McNamara & Sternberg 1983; Medin & Ortony 1989; Medin & Wattenmaker 1987; Rips 1989; Smith & Samuelson 1997; Wattenmaker *et al.* 1988). The existence of such beliefs may derive, as discussed earlier (see 1.2.1.3.1.), from the existence of experientially-based cores. However, psychological essentialism should not be equated with the classical view that concepts are represented with criterial, definitional attributes that guide categorisation. As demonstrated by Malt (1990, 1993, 1994; Malt & Johnson 1992), although people may believe that concepts have essential features and even state them, they do not generally use them in categorisation. For example, although people believe that being “H₂O” is a necessary and sufficient feature to consider something as “water”, they do not use this information as the primary determinant of what liquids they call “water” (Malt 1994). In short, there is usually a gap between conscious beliefs about concepts like “water”, which may be based entirely on the idea of an essence, and the knowledge on which classification is actually based.

The question remains unanswered as to whether the classical view is of any use. It is often argued that the domain of the classical view might be restricted to the so-called “well-defined” or “logically-defined” categories, that is, those (usually artificial) categories for which there is a clearly specified rule that all category members fully comply with. In fact, most of the experimental literature supporting the classical view of categorisation has been dominated by the use of that type of categories (e.g. Bourne 1970; Neisser 1987a). However, some recent work (Nosofsky 1984, 1986, 1991d; Nosofsky *et al.* 1989) has evinced that an exemplar model of categorisation (see section 1.2.3.) is also capable of predicting classification performance for logically-defined categories and that people’s mental representations of well-defined concepts

may often be characterised better in terms of exemplar storage solely. The claim in Nosofsky (1991d) is not that people cannot apply and use simple defining rules when instructed to do so. What is claimed is that in learning logically-defined concepts through induction from training exemplars, memories for the individual exemplars may form a dominant component of the category representation and account for classification.

Perhaps a case in which (at least) the classification assumptions of the classical view might be useful is that of “differential diagnosis” (Schank *et al.* 1986). In differential diagnosis, necessary and sufficient features are useful when it is essential that instances be correctly assigned to a given category. This happens when accuracy is the top priority in categorisation. This hypothesis ties in with frequent references by cognitive psychologists (e.g. Estes 1986a; Kemler Nelson 1984; Medin & Smith 1984; Rosch 1977; Smith & Medin 1981) and linguists (e.g. Lakoff 1987a; Taylor 1989; Ungerer & Schmid 1996) that, although insufficient as a general theory of human categorisation, the classical view of categorisation may come into its own whenever there is a need for technical, precise and rigid definitions and classifications in educational, scientific or legal contexts. This may occur in the domains of mathematics, geometry, biology, physics or even in the legal system where the accurate classification of an entity or event as an “even number”, a “square”, a “molecule”, or as “perjury” is absolutely necessary. Similarly, one of the uses of distinctive feature models, as Kempton claims (1981: 201) in discussing dictionary definitions, is that they may provide simple utilitarian definitions. Nevertheless, these must be thought of as convenient fictions, not representing everything that there is about word meanings or concepts. As a consequence, dictionary definitions say little about the conceptual validity of the classical view.

However, there is also some experimental evidence that appears to support the classical view of categorisation at least to some extent. First, it has been found that at least when the features of concepts are not easily integrated into meaningful coherent categories, subjects often construct categories defined by single dimensions (or unidimensional categories). These categories are similar to classically-defined categories (Ahn & Medin 1992; Medin *et al.* 1987; Spalding & Murphy 1996). Second, people happen to have a bias towards developing rules for classifying stimuli into categories, although if any are obtained that summarise all exemplars of a category, they may not be used in classification (Martin & Caramazza 1980). Third, defining features are usually obtained, for example, for proper nouns (Mcnamara & Sternberg 1983) although the defining attributes are not usually defining for all subjects, but only for individual subjects. Finally, in explicit rule-instruction conditions, Nosofsky and his co-workers (1989) observed patterns of classification performance conforming to the predictions of classical models of categorisation.

For these reasons, some researchers have emphasised that the classical theory should not be disregarded at all and that it can be reconciled with other different models of categorisation

that have been originally set up as alternatives against the classical view (Hawkins 1988; Rosch 1983).

1.2.1.5. The Emergence of Similarity-Based Categorisation Models.

Due to the limited experimental evidence supporting the classical view of categorisation and the fact that most of the experimental evidence gathered on categorisation contradicts most of the assumptions of the classical view, an increasing shift from the classical to alternative views of categorisation like the probabilistic, exemplar and mixed probabilistic-exemplar views in categorisation research has been made since the early 70s. This work suggested that, although the classical view of categorisation is not altogether wrong, it had failed as a general theory of human everyday categorisation.

The alternative models of categorisation devised in the 1970s and 1980s appeared to be more promising because they were both “experimental” and “experiential”. They are experimental in the sense that these models have derived from empirical research, unlike the classical view, which was mainly a non-empirical, philosophical position (see 1.2.1.3.). In addition, the new models were also “experiential” in the sense proposed by Lakoff (1982, 1987a) while classical models are “objectivist”.

The classical view assumed that the defining features of conceptual categories mirrored the supposedly necessary and sufficient features of the members of supposedly objective real-world categories. Because the “objective” world contains various kinds of things that share defining properties, human subjects’ task involves finding out those defining properties and representing them in memory. In other words the categories of mind were assumed to fit the categories of the world. In addition, since the defining features of real-world categories were presumably stable, different people should have the same concept, assuming that everyone has reasonably good access to the objective world or to information about it.

On the contrary, alternative models of categorisation appeared to be essentially experientialist in that they assumed that conceptual categories are determined by the structure of the world *as well as* by how individuals experience and interact with it given their human bodies, cognitive organisation, culture, etc. Conceptual categories were not considered to necessarily mirror any purportedly preexisting logically-organised world and their structure need not be accounted for by an association with something in objective reality. Instead, concepts or categories are formed and learned through the reciprocal interaction of cognitive structures and environmental events. Alternative models of categorisation (probabilistic, exemplars, mixed) are experientialist in that they reflect people’s experience and changing knowledge about categories. These models assume that people’s representations of categories vary dynamically with their experience (e.g. Barsalou & Medin 1986).

Apart from their “experientialist” standpoint, the main assumption of all the alternative models of categorisation proposed in the 70s and the 80s was that categorisation was driven by similarity. Similarity was considered as an extremely powerful principle that could account for most of the problems that plagued the classical view. It should be stressed that similarity had no place in classical models. Whereas similarity underlies categorisation in probabilistic, exemplar, and mixed models, it has nothing to do with categorisation in classical models. In classical models, an entity belongs to a category only if it strictly satisfies the category’s rule. An entity cannot be a category member if it is similar to a rule but must match the rule perfectly, partial satisfaction being insufficient.

1.2.2. Probabilistic Models of Categorisation.

The first alternative to the classical view of categorisation to be discussed is the probabilistic view of categorisation which still shares many assumptions with the classical view but differs from it in important respects.

1.2.2.1. Learning, Mental Representation, and Storage.

As in the case with classical models of categorisation, probabilistic models assume that the categoriser calculates some sort of summary statistics that describe the category as a whole as he/she is exposed to different category instances. The categoriser then forms an abstract representation of each category derived from a series of learning experiences. This process is usually referred to as “prototype” or “schema” abstraction (e.g. Anderson *et al.* 1979; Elio & Anderson 1984).²³

Like the classical view of categorisation, probabilistic models also assume that the representation of a concept is essentially an abstract summary description of an entire class that applies to all the instances of the category. However, unlike classical models, the representation does not need to contain singly necessary and jointly sufficient features. Instead, the representation includes features that are representative or characteristic of the exemplars of a category, namely, attributes that are highly probable across category members. In other words, the features that happen to enter the summary representation of a concept are those which have a substantial probability of occurring in instances of the concept.

Although probabilistic models agree on this representational assumption, they typically differ in the type of category-level information or statistics that the learner is supposed to

²³ These summary representations have sometimes been called prototypes or schemas. However, the label “prototype” has also been used to refer to other things other than “summary representations”, like, for example, very typical members in typicality rating tasks. One of the reasons why Smith and Medin (1981) decided to call probabilistic models “probabilistic” and not “prototype models” is because of the polysemy and ambiguity in the cognitive literature of the label “prototype”.

calculate and store. Most models claim that this information is based on some sort of central tendency of the instances' properties or patterns. This central tendency may include means -or the average of the observed feature values in category members- (e.g. Estes 1986b; Homa & Chambliss 1975; Posner 1969; Posner & Keele 1968, 1970; Posner *et al.* 1967; Reed 1972, 1978; Reed & Friedman 1973; Rips *et al.* 1973; Rosch *et al.* 1976), modes -or the most frequent values of the features- (e.g. Collins & Loftus 1975; Estes 1986b; Goldman & Homa 1977; Hampton 1979; Hayes-Roth & Hayes-Roth 1977; Kellogg 1981; McCloskey & Glucksberg 1979; Neumann 1974; Rosch *et al.* 1976; Smith *et al.* 1974), or both means and modes (Smith & Medin's 1981 "component" model). Other models propose that the summary description includes not only central tendency but also the shape of the distribution of feature values (Flannagan *et al.* 1986; Fried & Holyoak 1984) or the ideal values of attributes (Barsalou 1985; Kay & McDaniel 1978, Oden & Massaro 1978).

Although probabilistic models maintain that the abstract representation is what is essentially learned and stored, they do not exclude, in principle, learning (and subsequent storage) of individual exemplars. However, probabilistic models assume that, as a number of category instances are encountered, the generalisations gain more strength in memory than any of the specific items that generated them. Thus, category-level information is more robustly encoded than specific item memories. Such memories are, however, necessarily stored in memory for the summary statistics to be calculated and updated as new category instances are encountered. Probabilistic models require, like classical models, both extensive computation to perform learning and storage capacity.

1.2.2.2. Classification.

Probabilistic models perform classification quickly. This is so because they do not need to examine large numbers of exemplars: only the summary description calculated at the time of learning has to be retrieved in order to decide whether a novel instance is a member of a given category. Subjects first retrieve this information and then they determine whether or not a candidate instance for category membership is a member of a category on the basis of how similar that instance is to the category's summary description. In order to do so, the categoriser's cognitive system compares the novel entity's structural description to that of the category representation retrieved. Instances above some threshold of similarity to the abstract representation are classified as category members while all others that do not reach that threshold are not. In addition, if a new entity is similar to the summary descriptions of more than one category, it is classified as belonging to the category whose abstract summary it most resembles. The processing assumption is then that classification is merely a matter of assessing similarity rather than applying a definition.

1.2.2.3. Advantages and Disadvantages of Probabilistic Models of Categorisation.

The different probabilistic models account for most of the basic phenomena that troubled the classical view of categorisation. First, probabilistic models do not require sets of “defining” properties for which it is so difficult to find empirical evidence. The only requirement on properties to be part of a summary representation is that they be characteristic or probable of the members of the category. Second, since summary representations do not include necessary and sufficient features, classification of novel entities as category members can be driven by features that are merely characteristic (i.e. probabilistic). The use of nonnecessary features in categorisation is thus explained. For example, two entities can be classified as members of a certain category on the basis of certain features that each of them shares with the summary description but although the summary representation possesses all of them these features may not be exactly the same in both entities. Third, probabilistic models claim that unclear cases appear when a to-be-classified entity is not sufficiently close to the threshold level of similarity for membership in a particular concept (e.g. a “sea horse” is not sufficiently similar to the summary representation of “fish”), when an object is close but hardly crosses the threshold for more than one concept, (e.g. “tomato” is equally similar to both “fruit” and “vegetable”) or both. Fourth, the probabilistic view explains typicality by assuming that the typicality of an instance as a category member reflects how similar its properties are to those of the category representation. The more similar an entity is to the summary representation, the more typical it is (e.g. Hampton 1979, McCloskey & Glucksberg 1979, Rips *et al.* 1973, Smith *et al.* 1974). Fifth, probabilistic models explain why a perfect nesting of features and concepts does not exist. “Chicken”, for example, may be rated as more similar to “animal” than to “bird” because it shares more attributes with “animal” than with “bird”. In addition, resuming Hampton’s (1982) example, “car headlight” is sufficiently similar to “lamp” and “lamp” to “furniture” but “car headlight” is not sufficiently similar to the abstract summary representation of “furniture” to be considered as a kind of furniture. In addition, the existence of a basic level in hierarchical taxonomies is explained in the following way. The instances categorised as members of a basic level are more similar amongst one another and dissimilar from members of contrasting categories at the same level of abstraction than instances of members of categories at higher or lower levels of abstraction (Jolicoeur *et al.* 1984; Mervis & Crisafi 1982; Mervis & Mervis 1982; Mervis & Rosch 1981; Murphy & Brownell 1985; Murphy & Smith 1982; Rosch 1976, 1977, 1978; Rosch & Mervis 1975; Rosch *et al.* 1976). Apparently, there exists a greater perceptual similarity among the exemplar memories of a category at the basic level than amongst categories at any other level of abstraction and a greater dissimilarity from members of other categories at the same level of abstraction.

Despite the successful solutions that probabilistic models provide to the problems that plagued the classical view of categorisation, they have not been free from further theoretical as well as empirical inconveniences.

First, although probabilistic models do not preclude the mental storage of exemplar memories, they assume that the summary description calculated at the time of learning is more robustly encoded than those memories, which are not used to classify novel entities. However, much recent work has demonstrated that people use exemplars in classification (see 1.2.3.). Unfortunately, probabilistic models have little to say about the use of exemplars in classification, a shortcoming that even advocates of probabilistic models acknowledge (e.g. Posner 1968, 1986).

Second, it has often been argued that probabilistic models may not adequately capture all of people's knowledge about concepts (e.g. Anderson & Finchman 1996; Medin *et al.* 1982; Medin & Smith 1984; Michalski 1993; Smith & Medin 198). For example, probabilistic models are unable to account for people's extensive knowledge of correlated attributes. People know, for instance, that birds are typically small and typically sing, and also that these properties (i.e. "small" and "sings") are correlated: if one occurs, the other must occur as well. In fact, an extensive literature is available showing that people know a great deal about correlated attributes and use them in classification (e.g. Anderson & Finchman 1996; Malt & Smith 1984; Medin & Schwanenflugel 1981; Medin & Shoben 1988; Medin *et al.* 1982; Medin *et al.* 1987; Murphy & Wisniewski 1989b; Wattenmaker 1991; Wattenmaker *et al.* 1986). In short, probabilistic representations seem to discard too much information that can be shown to be relevant to human categorisation like correlated attributes and other more extensive and complex knowledge structures.

Third, probabilistic models do not usually handle context effects (Michalski 1993; Smith & Medin 1981), despite the fact that there is ample evidence that context can greatly affect concept usage and that how well an item belongs to a concept depends on the context in which it is presented (Medin *et al.* 2001). For example, the context in which to-be-categorised entities are found facilitates the categorisation of instances from expected categories but interferes with the categorisation of instances of unexpected categories. An appropriate context generally aids identification while an inappropriate context hinders it (Palmer 1975). Context effects are not limited to classification. Phenomena like instance typicality are also context-dependent. For example, Roth and Shoben (1983) showed that typicality judgements vary as a function of particular contexts. For example, "tea" is judged to be a more typical beverage than "milk" in the context of "secretaries taking a break" but this ordering reverses for the context of "truck drivers taking a break". Similarly, a "harmonica" is a typical musical instrument in the context of a campfire but atypical in the context of a concert hall. Unfortunately, probabilistic models have little to say about these context effects.



1.2.3. Exemplar-Based Models of Categorisation.

A basic distinction amongst categorisation models is whether the explanation for the categorising behaviour entails primarily a representation in memory of category-level information or memories of the separate or particular instances of the category studied. The former are known as ‘abstraction’ or ‘generalisation’ models and the latter as ‘exemplar’ or ‘instance’ models.

Most models of categorisation are abstraction models. These models posit that the cognitive system acts on input during the learning of categories in two ways: first, by determining which elements of a situation are relevant and which irrelevant; and second, by creating, in an unconscious and automatic way, new higher order information not given previously in any particular exemplar. The nature of this higher order information generated varies, however, for different models. It consists of defining features in classical models and of characteristic ones in probabilistic models. In addition, abstraction models assume that the cognitive system uses the resultant categorical information to classify novel items.

In general, psychology has shown considerable faith in the existence of an automatic abstraction mechanism. However, many studies suggest that the abstraction of regularities may be far from automatic when categorisers are exposed to novel material and that, even when regularities are known, they are not sufficient to explain categorisation behaviour. This is precisely the underlying assumption in the second alternative to classical models of categorisation discussed below and known as the exemplar view. Exemplar models do not require the abstraction of new summary information not given previously in any particular exemplar. Instead, this alternative assumes that the category instances encountered at the time of learning are stored in memory and that the abstraction of higher-order, not previously encountered summary descriptions, if it ever occurs, is secondary. In fact, a substantial number of exemplar models is available (Brooks 1978, 1987; Brooks *et al.* 1991; Estes 1986b; Heit 1992, 1994; Hintzman 1986; Hintzman & Ludlam 1980; Jacoby & Brooks 1984; Kossan 1981; Kruschke 1992; Lamberts 1994, 1995, 1998; Medin & Schaffer 1978; Nosofsky 1984, 1986, 1987, 1988a, 1988b, 1991a, 1991b, 1991c, 1991d; Nosofsky *et al.* 1989; Reber & Allen 1978; Smith & Medin 1981; Walker 1975; Wattenmaker 1991; Whittlesea 1987, 1997). This enormous body of literature catalogues a great deal of experimental support for the exemplar view.

1.2.3.1. Learning, Mental Representation, and Storage.

According to a generic version of exemplar models of categorisation, the categoriser does not calculate any kind of abstract summary generalisation at the time of learning a category. Learning in exemplar models only requires that, as subjects are exposed to category members, they memorise descriptions of individual examples of the category. Therefore, exemplar models require minimal computation to perform learning as there is no need to update a summary representation as each new category member is encountered. The cognitive system simply stores each new exemplar upon encountering it.

In general, exemplar models maintain that categories stored in memory do not include summary generalisations abstracted from exemplar memories but some models do (e.g. Medin & Schaffer 1978). Such abstractions though, are less robustly encoded in memory than exemplar memories and are not used in classification. Exemplar models claim that a category or concept is primarily represented by separate descriptions of some of its specific exemplars. In assuming that the learner stores mental representations of exemplars, grouped by category, exemplar models generally require a large memory capacity to store exemplars.

1.2.3.2. Classification.

According to exemplar models of categorisation, people classify new entities or instances on the basis of their similarity to memories of previously experienced and stored category members retrieved at the time of classification. Instances are assigned to the category with the most similar exemplar memories on average, with the highest number of similar exemplars or with the most similar exemplar to the to-be-classified entity.

Exemplar models of categorisation generally require more processing computations than probabilistic ones. While in abstraction models the same representation is retrieved for classifying different instances, in exemplar models a large proportion of exemplars must be examined on average before a positive classification can be reached. However, this mental operation is unproblematic for most exemplar models. The reason for this is that exemplar models assume that, in order to find the most similar exemplar(s) in memory, the cognitive system scans the whole memory array automatically and unconsciously on every trial by means of a parallel process. Specifically, the cognitive system compares the structural description of the unknown entity to all exemplar memories across all categories simultaneously, not serially. As a consequence, exemplars are supposed to be examined in a relatively short time.

From what has been said so far concerning learning, storage and classification, exemplar models can be considered as characterised by “late computation” while abstraction models are characterised by “early computation” (Estes 1986b). The term *early computation* implies that

the bulk of the mental operations needed in the categorisation process are accomplished during the course of learning with only the results of these being consulted at the time of the decision. For example, in abstraction models, summary representations of categories are developed during the course of learning and updated as information comes in on each trial, but when a test item is to be categorised, it is compared only to the current representation of the categories under consideration. In contrast, *late computation* characterises exemplar models in which the learner simply memorises exemplars during learning but is conceived to consult all exemplars in memory at the time of classification, which requires expensive computations.

1.2.3.3. Advantages and Disadvantages of Exemplar Models of Categorisation.

Exemplar models of categorisation account for the problems that plagued the classical view in a generally satisfactory way. First, since there is no reason why different exemplars need have the same properties, there is no reason to expect defining properties. The only requirement on properties is that they characterise at least one exemplar, a very different assumption from the one in probabilistic models, in which a property must have a substantial probability of occurring among the concept's instances for it to be included in the category representation. Second, since the features of category members need not be true of any other category member, any feature or conjunction of features can be used to classify an entity as a category member. The use of nonnecessary properties in classification is therefore easily explained. Third, according to exemplar models, unclear cases can arise when an object is similar to exemplar memories of more than one category (e.g. a "tomato" is similar to the exemplars of both "fruit" and "vegetable"), or when an object is not very similar to the exemplars of any category (e.g. "a sea horse" is not very similar to the exemplars of the categories "horses" or "fish"). Fourth, exemplar models account for the existence of typicality ratings and effects in the following way: the more similar an instance is to the other members of its category and the less similar it is to members of contrasting categories, the higher its typicality will be (Nosofsky 1988b). Finally, the exemplar view explains most phenomena associated with taxonomies. The lack of perfect nesting of features and concepts is easily explained by the exemplar view. "Chicken", for example, might be rated more similar to "animal" than to "bird" because the particular exemplars associated with "chicken" may share more properties with the exemplars of "animal" than with the exemplars of "bird". Also, a "car headlight" is sufficiently similar to exemplars of the category "lamp" and the exemplars of the latter to those of the category "furniture", but a car headlight may not be sufficiently similar to the exemplars of the category "furniture". In addition, the explanation for the existence of a basic level is very similar to the explanation provided by probabilistic models.

Despite providing a more or less successful account of the problems that tarnished the classical view of categorisation, exemplar models are not free from some of the flaws attested in probabilistic models.

First, a serious problem with exemplar models is that, although people clearly know and use numerous abstractions about categories that they induce from exemplars, these models have little to say about the formation of these abstractions, their content, or their use (Barsalou 1992a; McClelland & Rumelhart 1985; Michalski 1993).²⁴ Second, although some exemplar models can account for correlated attributes (Medin 1989; Medin & Schaffer 1978; Nosofsky 1988b), exemplar models are criticised because they are unable to capture all the knowledge people have about categories. Third, exemplar models provide an insufficient account of context effects. The basic idea is that prior context raises the probability of retrieving certain exemplars in a representation and not others. For example in a sentence like “the man lifted the piano”, the context preceding “piano” may increase the availability of exemplars of heavy pianos (that is, exemplars whose representations emphasise the property of weight), thereby making it likely that one of them will actually be retrieved (Smith & Medin 1981: 159), but exemplar models do not account for this.

Apart from these three problems, exemplar models also have specific weaknesses that probabilistic models did not. The first problem is people’s apparent capacity to construct new categories without experiencing exemplars of that category, a theoretical flaw that even proponents of exemplar models have come to acknowledge (e.g. Medin & Shoben 1988). This contradicts the exemplar theory’s assumption that classification is always based on previously experienced and remembered exemplars. Without further assumptions, exemplar models cannot account for how people are able to decide that something belongs to a category if they have never encountered an example of that category (Rips 1995). Second, a serious problem of exemplar models is their lack of constraints on what properties enter into concepts or even on what constitutes a concept. According to exemplar models, it seems that the only relation between the exemplars in a given representation is that they all point to the same concept. This, unfortunately, would unduly justify the totally unnatural grouping into a category of any array of entities. It is true, however, that this is possible with emotional-response categories or ad-hoc categories (see 1.1.5.). Third, the exemplar view assumes that the cognitive system stores a tremendous, almost unlimited amount of idiosyncratic exemplar information for categories. This presupposes a large memory (or storage) capacity. However, although no limit on people’s ability to acquire new information has been reported so far (Barsalou 1992a: 27), it seems implausible that each instance categorisers ever encounter becomes a separate part of the representation of any category (Smith & Medin 1981: 147). If that is the case, there must be

²⁴ This problem is the reverse of the one in probabilistic models (i.e. their failure to capture the idiosyncratic information of exemplars).

some means of restricting the (theoretically infinite) number of exemplars that might enter the category representation for an exemplar model to be actually plausible in cognitive terms.

Although exemplar models have problems of which there was no trace in probabilistic models, exemplar models also have advantages over probabilistic ones. Exemplar models have been praised for their ability to account for the learning of categories whose exemplars can vary widely and have little in common (Barsalou 1992a: 27). In addition, Medin (1989, 1999) claims that exemplar models predict sensitivity to category size, instance variability (or the variability of exemplars) and account for some context effects.

1.2.4. Mixed Models of Categorisation.

Probabilistic and exemplar models of categorisation have often been contrasted as though they were mutually exclusive. More recently, however, investigators have argued that both learning, storage, and classification in categorisation often involve both abstracted category generalisations and stored exemplar information. Some categorisation models posit that category representations include both summary information and exemplar memories. They raise the possibility that the representation for a single category is mixed, that is, partly probabilistic, partly exemplar (e.g. Barsalou 1990b; Busemeyer *et al.* 1984; Elio & Anderson 1981; Homa 1984; Homa *et al.* 1981; Lingle *et al.* 1984; Malt 1989; Medin 1986; Medin & Smith 1981; Medin *et al.* 1984; Medin *et al.* 1983; Smith & Medin 1981; Smith & Zárate 1990).

1.2.4.1. Learning, Mental Representation and Storage, and Classification.

Mixed models assume that, at the time of exposure to novel instances of a category, both memories of specific exemplars and a summary abstraction describing some sort of central tendency are learned and stored in long-term memory. Most importantly, these models assume that each type of information is encoded as robustly as the other, in equal terms. For example, in learning the category “dog”, a person might store specific exemplars like, for example, a dog next door named Fido, a friend’s dog called Rover, abstractions over a specific subset of exemplars like “Alsations”, “Pekinese”, etc. as well as a summary abstraction of “dog” which does not correspond to any particular dog or breed. Therefore, mixed models assume that the representation of a concept or category like “dog”, contains different members varying in their degree of abstraction. The exemplars themselves can be represented in different ways, partly depending on whether they are subsets of dogs that correspond to specific breeds, like “terrier”, “Alsatian”, “Pekinese”, etc., or real-world or imagined specific instances like “the cartoon Pluto” or “my dog Lucky”. The former are, of course, abstractions. While “Pekinese” is

something of an abstraction in that it abstracts over different specific dogs, it seems qualitatively less abstract than “Rover”, which abstracts over a specific entity.²⁵

Mixed models maintain that, when making category membership decisions, the summary representation is more accessible in some cases while at some other times it is any of the exemplar memories. Consequently, the to-be-categorised instances are compared to the summary representation or to different exemplar memories to perform classification decisions.

1.2.4.2. Advantages and Disadvantages of Mixed Models of Categorisation.

Mixed models account for the problematic findings of the classical view integrating the explanations for such findings of both probabilistic and exemplar models of categorisation. However, although mixed models constitute an effort to reconcile apparently competing views, the same problems as those described for strictly probabilistic and exemplar models reappear in mixed models. These include an insufficient account of context effects, a failure to represent all the knowledge people have about categories, the lack of constraints on possible attributes and categories, the assumption of an unrestricted storage capacity, etc.

1.2.5. Insufficiency of Similarity-Based Categorisation Models.

Part of the reason why the alternative models of categorisation to the classical view discussed above have not been completely satisfactory is because of their strong and almost exclusive reliance on the principle of “similarity”. Apparently, similarity is a powerful mechanism but not sufficient to account for all the facts known about categorisation. The attempt to account for the phenomena that similarity-based categorisation models were unable to sufficiently explain motivated, at least partially, the second of two major shifts in categorisation research (Busemeyer *et al.* 1997; Medin 1989; Murphy 1993b). The first one is the shift from the assumption that conceptual representations have defining properties (the classical view) to the idea that they contain properties that are only characteristic of category members (the probabilistic, exemplar or mixed views). The second shift consists in considering conceptual representations as being organised around more complex knowledge structures (conveniently called “theories”) instead of being exclusively structured by similarity.

It should be mentioned that, as far as the former shift is concerned, practically all of the various models proposed as alternatives to classical models of categorisation share the following three assumptions. First, conceptual coherence is based on similarity: the association of different instances under a given category seems coherent because such instances are similar

²⁵ Viewed in this way, it may be sensible to conclude that rather than permitting abstraction or not, the different models of categorisation differ on the type of information that they abstract (e.g. Medin *et al.* 1983: 612ff.). Even exemplar models permit

to one another and dissimilar to members of other categories. Second, classification is driven by similarity. Classification is believed to involve a feature matching process and computation of feature similarity carried out by the cognitive system between a to-be-classified entity and the conceptual representation (summary abstraction, exemplar memories or both) with which the entity is compared. The likelihood of assigning some entity to a category depends on the entity's reaching some "threshold of similarity" to the category representation.²⁶ Third, summary representations or exemplar memories are made up of collections of relatively independent features. Fourth, similarity-based categorisation models assume that long-term memory contains rather invariant, static representations of categories and that they are retrieved from long-term memory into working memory and are used on all occasions in which a category is processed. Finally, concepts are relatively independent of one another and the few relationships established amongst them (class inclusion, category contrasts at the same level of abstraction) are based on similarity.

However, categorisation researchers soon began to note problems with these assumptions. First, the notion of similarity seems too unconstrained to give a satisfactory account of conceptual coherence. It is not at all clear whether some members belong to the same category because they are similar or whether they are similar because they are in the same category. Similarity could be a by-product of conceptual coherence rather than its determinant. Second, even if classification is driven by similarity, models do not explain why certain features of a concept are relevant to conceptual representation and are used in such processes. Unless constraints specify some properties as relevant and others as irrelevant, similarity becomes a meaningless construct (Wisniewski & Medin 1991: 238). However, even if a successful account were available, similarity would not always account for the assignment of entities to either of two (or more) contrasting categories (Medin 1989; Medin & Wattenmaker 1987; Murphy & Medin 1985). Third, conceptual representations do not seem to be satisfactorily described in terms of collections of unrelated features or "feature lists". This representational language is partly responsible for the failure of similarity-based models to capture all the conceptual knowledge people have about connections between features and between features and concepts. Fourth, the assumption that all the contents of conceptual representations of categories are invariantly retrieved from long-term memory into working-memory and used on all occasions in which a category is processed prevents models from accounting for context effects in a satisfactory way. Finally, the notion of similarity does not always account for the class inclusion relations or more complex relationships between conceptual and more complex knowledge

abstractions as exemplar memories are really abstractions over specific instances.

²⁶ For a very detailed description of this process see Barsalou (1990: 268-269, 1992a: 25). However, although this process is usually conceived of as a bottom-up feature-matching process in which the feature description of the entity is compared with the feature description of the category representation (or category representations) stored in memory, an increasing body of evidence has shown that the similarity relationships underlying classification may also be based on holistic, overall or global similarity rather

structures. Although similarity-based categorisation models have helped researchers understand the complex knowledge structures like taxonomies (but by no means explain them completely), they have typically failed to capture the extensive knowledge people have about concepts and the interconnections existing amongst them.

1.2.5.1. The Emergence of the Theory-Based Account of Categorisation.

The dissatisfaction with similarity-based models of categorisation and their assumptions led to the emergence of the “theory-based” view of conceptual structure and categorisation, the second major shift in categorisation research. As mentioned above, the shift consists in viewing concepts as being organised around “theories” instead of being exclusively based on similarity. “Theories” is the name used to refer to the set of beliefs that people have about the interrelations and causal connections between the features of a concept, between features and concepts, and between concepts themselves. The beliefs in question may be “sketchy, naive, stereotyped or incorrect” (Rips 1995: 76) and so they are better conceived of as “any of a host of mental explanations rather than a complete, organized, scientific account” (Murphy & Medin 1985: 290). Theories usually include people’s ideas about what makes an instance a member of the category (i.e. explanations), some specification of the normal or default properties that such an instance possesses and an account of the relation between both (Rips 1995: 76).²⁷

The emergence of the “theory” view, an experientialist one like similarity-based categorisation models (see 1.2.1.5.) and accompanied by an extensive body of experimental research²⁸, improves on probabilistic, exemplar and mixed models of categorisation in the following ways. First, it notes the insufficiency of similarity as a general explanation of conceptual coherence (Medin *et al.* 2001). Coherence can be achieved, according to the theory-view, in the absence of any obvious source of similarity. Goal-derived ad-hoc categories (e.g. Barsalou 1983) or emotional-response categories (Niedenthal *et al.* 1995) are clear examples of

than the more analytic featural similarity (Kemler Nelson 1984, 1988; Smith, J. D. 1989; Smith & Kemler Nelson 1984; Smith & Shapiro 1989; Ward & Scott 1987; Ward *et al.* 1990).

²⁷ “Theories” have been referred to using a wide variety of names like “background knowledge” (Lin & Murphy 1997), “complex knowledge structures” (Murphy 1993b), “naive theories” (Rips 1989; Smith & Samuelson 1997), “folk theories” (Malt 1990), “intuitive theories” (Barsalou 1992a; Barsalou & Medin 1986; Murphy & Wisniewski 1989b), “micro-theories” (Honeck *et al.* 1985), or even “mini-theories” (Rips 1995). To be sure, the various names have sometimes been used to avoid possible confusions with other senses of the word “theory” in the now extensive literature on the issue. In addition, other kinds of similar knowledge structures have been proposed by cognitive scientists such as “frames” (e.g. Barsalou 1992b; Fillmore 1985), “scripts” (e.g. Schank & Abelson 1977), “schemata” or “mental models” (e.g. Barsalou 1992a), “idealized cognitive models” (Lakoff 1987a, 1987b), etc.

²⁸ Since the emergence of the theory-based view of concepts (see also Heit 1997; McCauley 1987; Murphy 1993a; Neisser 1987a, 1987b), various studies have empirically examined the relationships between background knowledge and concepts. These studies have examined the effects of knowledge on typicality ratings (e.g. Honeck *et al.* 1985; Rips 1991; Rips & Collins 1993), concept (or category) learning (e.g. Heit 1994; Kaplan & Murphy 1999; Medin *et al.* 1987; Murphy & Allopenna 1994; Nakamura 1985; Pazzani 1991; Spalding & Murphy 1996; Wattenmaker *et al.* 1986; Wisniewski 1995; Wisniewski & Medin 1994, 1991), conceptual combination (e.g. Cohen & Murphy 1984; Hampton 1988b; Medin & Shoben 1988; Murphy 1988, 1990; Rips 1995), item classification after the category is well learned (e.g. Keil 1987; Lin & Murphy 1997; Lin *et al.* 1997; Malt 1990; Pazzani 1991; Rips 1989; Smith & Sloman 1994; Wattenmaker 1995), conceptual development (e.g. Gelman & Markman 1986; Hayes-Taplin 1992; Markman 1989; Nelson 1983), expert categorisation phenomena (Honeck *et al.* 1987), cross-classification (Ross & Murphy 1999), etc.



this. In fact, some classifications blatantly contradict perceptual similarity (like classifying a “dolphin” and a “bat” as “mammals” and not as “fish” or “bird” respectively). For example, conceptual coherence is provided by the theories people have. Concepts are coherent to the extent they fit people’s background knowledge or naive theories about the world. People’s theories and knowledge of the real world, and not similarity, play a major (but not exclusive) role in conceptual coherence and serve to structure concepts.²⁹

Second, theories easily justify why certain features are chosen and represented in concepts while others are not. They can indicate which attributes of objects should be considered in computing similarity (Medin *et al.* 1993, Murphy & Medin 1985), and how these attributes should be weighted. In addition, the theory view acknowledges that, once the features of a domain are known and the relevant aspects of similarity defined, feature matching processes may provide an adequate account of classification (Murphy 1993b). According to Rips (1995), deciding whether a particular instance belongs to one or other category comes down to determining which theory corresponding to the different categories best explains the properties manifested by the instance. People’s background knowledge serves to provide explanations for why particular instances are members of a category (Wattenmaker *et al.* 1988). Third, the theory view rejects the view of concepts as lists of independent features, which derives from the representational language that most similarity-based categorisation models use. That representational language, known as “feature lists”, characterises people’s knowledge of a concept precisely as a list of relatively independent features. The theory view claims that although useful, feature lists can at best be considered as a rough approximation to human knowledge but by no means a complete picture of it. As Barsalou and Hale put it:

Feature lists are like a few pieces of a dinosaur’s skeleton, discovered by a palaeontologist. They provide hints of the dinosaur’s overall structure but are far from constituting it completely. Analogously, feature lists may pick out pieces of a structured representation, but these representational fragments do not describe the organisation of the underlying structure. Because of their expressive limitations in representing relations, feature lists serve primarily to represent knowledge fragments rather than knowledge structure.

Lawrence Barsalou & Christopher Hale (1993: 135)

The theory-based view explains why and how features are tied together to form a coherent concept, that is, the interproperty relationships in which features participate, a pervasive problem in categorisation models acknowledged as early as the behavioural era (Smoke 1932). The theory view suggests that knowledge of the domain provides links between features, that is, people have knowledge about the causal connections between features (Medin *et al.* 1987; Murphy & Medin 1985; Murphy & Wisniewski 1989b) and these correlations are

²⁹ However, the theory-view does not claim that similarity is a useless determinant of conceptual coherence. Recent reassessments of similarity testify to its importance in conceptual coherence (e.g. Goldstone 1994b; Hahn & Chater 1997; Hayes-Taplin 1992; Lamberts 1994; Medin *et al.* 1993). What the theory-view maintains is that both knowledge-based and similarity-based determinants of conceptual structure are not mutually exclusive but equally necessary for a complete account of category structure and conceptual coherence (Lamberts 1994; Malt 1995; Medin & Ortony 1989; Murphy 1993b; Niedenthal *et al.* 1996; Wattenmaker *et al.* 1988; Wisniewski & Medin 1991: 237).

embedded in one's knowledge and theories about the domain. Domain knowledge interprets and integrates the properties of category members. The correlation between "flies" and "wings" in the concept "bird" may exemplify the point. According to the theory view, these features are connected by virtue of their being part of people's naive theory of flying. People know that flying has something to do with supporting a body on air currents and that fat round things cannot support bodies on air currents, but that wide, thin things can. Therefore, there is a causal connection between having wings (which are wide, thin, and project from the body) and flying, rather than a simple association.³⁰ The assumption that features are interrelated provides an explanation for more complex feature relations than those of correlated attributes. To illustrate this, consider the concept of "car". According to the view of conceptual structure held by similarity-based categorisation models a mental representation (a summary representation or an exemplar memory) can be defined by a set of independent features. These include, amongst others, the following: "has wheels", "has doors", "has windows", "has a metal body", "has an engine", "transports people or goods", etc. On the contrary, the theory view assumes that the features are interconnected by the background knowledge that people have about cars, which refers to the 'causal, underlying beliefs about how various components of a car fit or work together to give rise to its function as a vehicle' (Lin & Murphy 1997: 1153). Thus, a belief such as "the engine turns the wheels, enabling the car to move about, and being able to move about in turn is a critical function of a car" in which the features are related to one another in different ways which would be considered as part of the knowledge associated with the concept and thus, as part of its conceptual representation.

Fourth, most views of cognition assume that knowledge consists of relatively static and well-bounded packets of information (e.g. definitions, probabilistic abstractions, exemplars, etc.) and that the same packet of information is used to represent all instances of a category across a wide range of contexts. However, some proponents of the theory view suggest that conceptualisations are, on the contrary, the conceptions of a category at a particular point in time, that is, how people construct or view a category on a specific occasion. The conceptualisation of a category on a particular occasion includes a very small subset of the total knowledge for a category in long-term memory, with the active subset varying widely from occasion to occasion. For example, the category "knife" at the dinner table activates a different set of characteristics from the ones activated by the category at a murder scene. The concepts that people use are constructed in working memory from knowledge in long-term memory by a

³⁰ This account of feature correlations avoids some of the problems of similarity-based models that account for correlated attributes with the independent feature list assumption. First, the features that are represented as being correlated out of the many possible ones are those that can be connected in people's knowledge base rather than any arbitrary connections in the feature list. Second, the correlation representation is not adding special links to the concept, but rather is capitalising on knowledge that we know people have. This knowledge outside the simple feature list can help structure the list and even influence what is in it. Finally, the theory could explain how people detect correct correlations among all the possible correlations in the world. The feature correlations are those that subjects expect to be correlated, then subjects do attend to their relationship -whether or not the features actually are correlated.

process sensitive to context and recent experience. Concepts appear to vary widely as a function of goals, current context, and recent experience (Barsalou 1982, 1987, 1989, 1992; Barsalou & Medin 1986; Barsalou *et al.* 1986, 1987; Medin & Barsalou 1987; Michalski 1993; Smith & Jones 1993; Smith & Samuelson 1997).

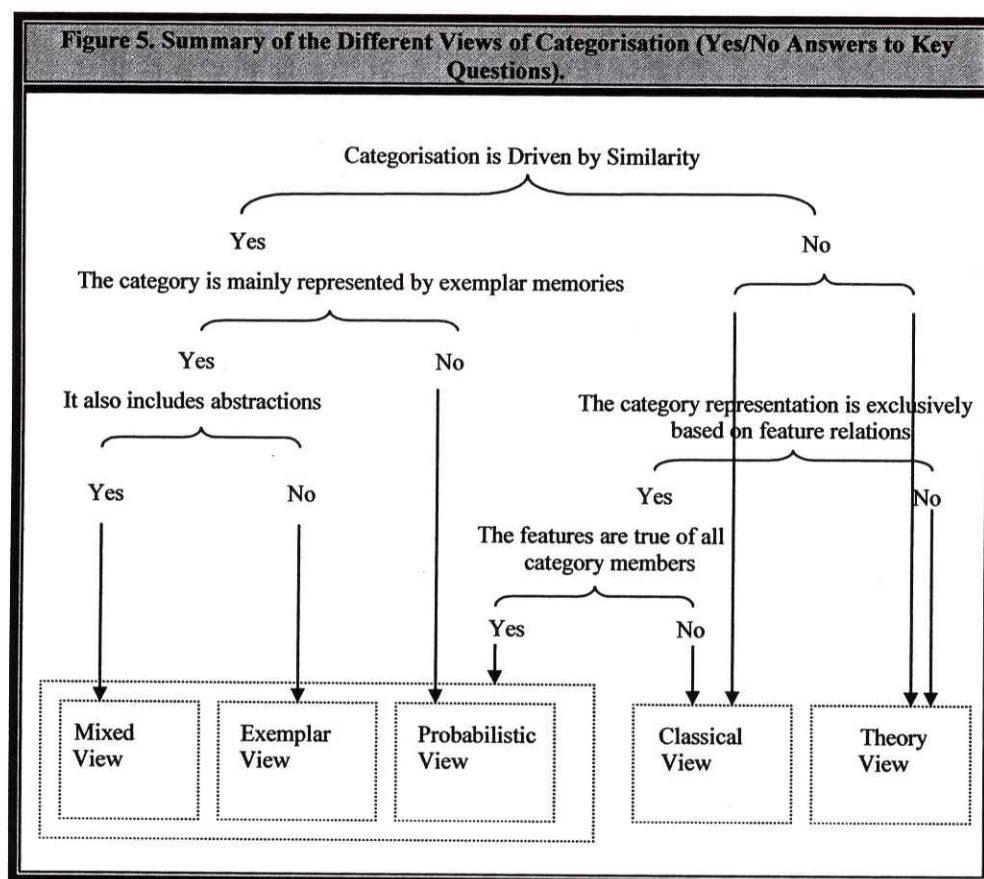
Finally, the theory-view also rejects the view of concepts as independent units of knowledge. Concepts are interconnected with one another in rich knowledge structures (apart from taxonomies) in much the same way as the features of the concept itself. In addition, a theory does not exist independently of its concepts and a concept is partly defined by the theories that it enters into (Murphy 1993b). This step forward is again to a great extent due to the increasing use of other representational formats like frames or mental models which overcome the limited potential of feature lists to represent critical conceptual relations between features and concepts (Barsalou 1992a, 1992b; Barsalou & Hale 1993).³¹ The theory-based view regards concepts as theoretical knowledge or, at least, as embedded in “theories”. To exemplify, the concept of “car” is part of a wider background knowledge structure or “theory” in which the concept of “car” is embedded: the “journey” theory. This “journey” theory, as Lakoff claims (1987a: 78), involves a structured scenario that specifies the steps to be followed in a journey (e.g. you have access to the vehicle, you get into it and start it up, you drive, row, fly, etc. to your destination, you park and get out, you are at your destination). In addition, the “journey” theory provides appropriate interrelationships amongst concepts and so, the concept of “car”, for instance, is interconnected with other concepts like “driver”, “goods”, etc. In addition, it may explain why certain category inclusions exist, like the fact that whales are mammals and not fish.

1.2.6. Categorisation Models: Final Concluding Remarks.

Although there is general agreement that categorisation is a fundamental human ability, there is surprisingly no general consensus about the nature of that ability (Medin 1989) as is shown in figure 5, which summarises the state of the art about the different categorisation views by answering key questions. This may be due in part to the fact that tests have seldom compared explicitly formulated models (see Ashby & Maddox 1993; Elio & Anderson 1981; Estes 1986b;

³¹ The distinction between a view of categorisation and a representational format is not usually well explained. A categorisation view seems to be a set of assumptions about learning, conceptual representation and classification while a representational formalism is a language in which those aspects are expressed (e.g. Michalski 1993). An example of the confusion between both notions can be exemplified referring to the so-called connectionist models of categorisation. These are simply categorisation models that assume that representations do reside in unique locations or form distinct traces in memory but that representations share many or all of the same storage elements and thus representations are distributed so one cannot properly point to a single memory trace. They differ from symbolic representations, which assume that symbols are localised representations occupying a single element of the computing resources. However, leaving aside the different representational language, these models are similar to prototype models (Gluck & Bower 1988; Shanks 1991), to exemplar models (Kruschke 1992) or mixed models (e.g. Knapp & Anderson 1984; McClelland 1981; McClelland & Rumelhart 1985), etc. It is just their representational language that changes. Any categorisation model (classical, exemplar, probabilistic or mixed) can use any representational language

Hayes-Roth & Hayes-Roth 1977; Martin & Caramazza 1980; Nosofsky 1991c; Reed 1972 for some exceptions). In addition, when such comparison has been made, the attempts to achieve differential empirical tests of these models have not always proved instructive. In fact, it is often difficult, if not impossible, to distinguish empirically amongst different categorisation models. As mentioned above (1.2.1.4.), findings traditionally believed to provide evidence for the classical view of categorisation were easily accounted for by an exemplar model (e.g. Nosofsky 1984, 1986, 1991d; Nosofsky *et al.* 1989), and data traditionally taken as strong support for the abstraction of probabilistic central tendencies of categories (e.g. Hartley & Homa 1981; Homa & Vosburgh 1976; Homa *et al.* 1973; Posner & Keele 1968, 1970; Strange *et al.* 1970) can also be predicted by exemplar models (e.g. Hintzman 1986; Hintzman & Ludlam 1980).



Nowadays, almost any model of categorisation (in particular probabilistic, exemplar, or mixed) can be constructed in such a way that it will account for many of the observed categorisation data equally well (see the papers mentioned in the preceding paragraph). Since many different processes and knowledge structures must be reflected in any model it is not surprising that very different models can explain the same behaviour (Barsalou 1990b; Malt

(feature lists, frames, mental models, etc.). The fact that similarity-based categorisation models have typically used feature lists as their representational format is not a consequence of their inability to do so but simply that frames have not been used in that way.

1989). Clearly new methods or experimental tasks to study categorisation are needed to answer questions about the nature of categorisation (Lamberts 1994).

In general, however, it seems reasonable to conclude that all learning, modes of mental storage and classification assumptions embodied in the different models of categorisation are, to some extent, available to human beings and that they are all needed to provide a satisfactory account of human categorisation. As Medin and Smith claim:

Categorisation theories are essentially alternative procedures, each of which can be evidenced when their eliciting factors are operative. In this sense, all models are correct (and all incorrect) at least some of the time.
Douglas L. Medin & Edward E. Smith (1981: 242).

An interesting direction for future categorisation research might then be to discover which “eliciting” factors determine the way a concept (or category) is learned, what exactly is stored in memory or which strategy is used to classify specific instances. Some work has already been conducted in this direction (Estes 1986b; Homa 1984; Smith & Medin 1981: 170-173). Some of the factors for which there is some experimental evidence at present are the age of subjects (Keil & Kelly 1987; Kossan 1981), the strategies that categorisers adopt (Medin & Smith 1981), whether the learner is analytic or non-analytic (Elio & Anderson 1984), whether learning is incidental or intentional (Homa *et al.* 1987; Lamberts 1994; Wattenmaker 1991), the type and availability of feedback (Homa & Cultice 1984), the kind of instructions that subjects follow (Homa *et al.* 1987), the moment after learning a category in which classification is performed (Robbins *et al.* 1978), the degree of intraclass similarity of category members (Goldman & Homa 1977; Homa & Cultice 1984; Homa & Vosburgh 1976; Homa *et al.* 1987), the presence or absence of unique or idiosyncratic features (Busemeyer *et al.* 1984; Kossan 1981; Malt 1989; Medin & Smith 1984; Medin *et al.* 1983; Wattenmaker 1991), the number of category members (Hartley & Homa 1981; Homa & Chambliss 1975; Homa & Cultice 1984; Homa & Vosburgh 1976; Homa *et al.* 1987; Homa *et al.* 1973; Homa *et al.* 1981), the number of categories involved (Homa 1984), the order of presentation of stimuli (Elio & Anderson 1984), etc.

As a concluding remark, we must say that several challenges to the study of categorisation lie before researchers in the next years. Although cognitive scientists have yet to develop a powerful theory of categorisation, they have come to understand why the issue is so formidable and recognise the fundamental problems they must solve to produce a satisfactory theory of categorisation. As MacWhinney puts it:

A full picture of the development of categorisation must not only show us how categorisation arises from biology, from cognition, from social interaction, and from language, but a full picture must also show how these very different impacts on categorisation compete and coexist. It is precisely when we consider these interworkings that we realise that we are in the deepest ignorance and it is here that the issues seem the most fascinating.

Brian MacWhinney (1984: 414-415).

1.3. Categorisation Research with other Subject Populations.

Most of the categorisation research has been concerned with adult human subjects. Nevertheless, an extensive body of research has also been conducted with younger human populations as well as with other non-human species. This research is important because it provides interesting insights into the capabilities of those populations as well as providing an external ontogenetic and phylogenetic perspective on adult human categorisation. A general overview of categorisation would be incomplete if the most interesting findings from this research were overlooked. In the rest of this introductory chapter the categorisation research conducted with children, infants, and non-human species will be reviewed with a special emphasis, as was the case with human adults, on visual categories.

1.3.1 Categorisation in Children.

The period of time between the beginning of language acquisition and puberty represents one of the most complex and intriguing ones for different reasons. This age group is particularly interesting because it is marked by the fact that children are acquiring language and developing cognitive, social and motor abilities at the same time.

One of the most active areas of research in cognitive development for some time has addressed the issue of how children acquire categories and categorise stimuli. Since the early '70s, many experimental studies have concentrated on children's ability to categorise the members of many natural, artefactual or specifically constructed artificial categories. Different experimental techniques have been used to study the categorisation abilities of children. Categorisation tasks usually take place in a laboratory set up as a playroom in which the experimenter and the child, usually accompanied by a parent, interact in different ways. Categorisation tasks are usually presented to children as games that they can play. Children may be asked to list the members of familiar categories in instance generation tasks (Nelson 1974b; Rosner & Hayes 1977), to name objects shown to them in "naming tasks" (Alexander & Enns 1988), to verify the category membership of a given object in a category in "category verification" tasks (Blewitt 1994; Callanan & Markman 1982; Kossan 1981), to sort objects into distinct groups in "free-sorting" tasks (e.g. Alexander & Enns 1988; Gopnik & Meltzoff 1987, 1992; Margand 1977), to express all they know about the objects presented (Anglin 1985), etc. Another very common technique is the "selection" task in which children are presented a haphazard array of objects and are asked to select (e.g. point to, grab, touch, place elsewhere, etc.) the object that exemplifies the category under investigation (e.g. Alexander & Enns 1988; Callanan & Markman 1982; Daehler *et al.* 1979; Déak & Bauer 1995; Erreich & Valian 1979; Kalish & Gelman 1992; Mervis & Crisafi 1982; Saltz *et al.* 1972; Salt *et al.* 1977; etc.).

Most studies have focused on natural categories, in particular different types of animals, food, fruits, vegetables, plants, leaves, flowers, colours, etc. (e.g. Blewitt 1994; Blewitt & Durkin 1982; Callanan 1983; Callanan & Markman 1982; Daehler *et al.* 1979; Déak & Bauer 1995, 1996; Gelman 1984, 1988; Gelman & Coley 1990; Gelman & Markman 1986, 1987; Gelman & O'Reilly 1988; Gelman *et al.* 1986; Heider 1971; Melkman *et al.* 1981; Mervis 1984, 1987; Mervis & Mervis 1982; Mervis *et al.* 1975; Nelson 1973, 1974b; Rosch *et al.* 1976; Rosner & Hayes 1977; Saltz *et al.* 1977; Saltz *et al.* 1972; Tversky 1989). In addition, artefactual categories have also been studied, like different types of vehicles, furniture, musical instruments, clothes, cutlery, toys, etc. (Blewitt 1994; Blewitt & Durkin 1982; Callanan 1983; Callanan & Markman 1982; Daehler *et al.* 1979; Déak & Bauer 1995, 1996; Gelman 1984, 1988; Gelman & O'Reilly 1988; Gopnik & Meltzoff 1987; Hupp & Mervis 1982; Melkman *et al.* 1981; Mervis 1984; Mervis & Mervis 1982; Nelson 1973; Rosner & Hayes 1977; Rosch *et al.* 1976; Saltz *et al.* 1977; Saltz *et al.* 1972; Tversky 1989).

Many studies on child categorisation have also been conducted with artificial categories. These include artificial categories resembling natural or artefactual ones like artificially constructed animals (Alexander & Enns 1988; Horton & Markman 1980; Kossan 1981; Markman *et al.* 1980; Merriman *et al.* 1991), toy-like objects (Merriman *et al.* 1991; Mervis & Crisafi 1982; Mervis & Pani 1980) or cartoon-type faces (Kemler Nelson 1984; Ward *et al.* 1990) and completely arbitrary categories like eight-sided polygons (Aiken & Williams 1973; Williams *et al.* 1977), rectangles (Gopnik & Meltzoff 1987), etc.

Finally, some types of abstract categories like some metalinguistic categories such as “noun” or “verb” (e.g. Anglin 1985; Maratsos 1982; Maratsos & Chalkley 1980) or concepts like “animacy” (Margard 1977), “front” and “back” (Levine & Carey 1982), “in”, “above”, “below”, “beside” (Erreich & Valian 1979) have also been studied.

1.3.1.1. Categorisation by Children: Issues.

Research on child categorisation has tried to discover the characteristics of young children's conceptual categories but also the similarities and differences between children and adults. Several issues have attracted the attention of researchers.

First, there has been growing interest in the complex roles and relationships of pre-linguistic object categories and linguistic input in learning early object words. Obviously, children have already acquired the categories behind many words on their own before they learn the word (Mervis 1987). Considerable research on categorisation in preverbal infants shows this (see section 1.3.2.). Many researchers claim that early words map previously achieved non-linguistic categories, which serve to stimulate and guide a search for a corresponding word in early word learning (e.g. Clark 1973, 1983; Levine & Carey 1982; Merriman *et al.* 1991;

Mervis 1984, 1987; Nelson 1973, 1974a, 1974b; Roberts 1988; Roberts & Horowitz 1986). However, some interactionist accounts also maintain that linguistic input helps in discovering and learning new categories (Bowerman 1980; Deák & Bauer 1997; Markman 1984, 1987; Markman & Hutchinson 1984; Mervis & Pani 1980; Mervis 1984, 1987; Nelson 1983; Nelson *et al.* 1978; Schlesinger 1982; Waxman & Hall 1993; Waxman *et al.* 1991). As experimental evidence has accrued in favour of both views, it seems plausible that there is some truth in both. In some cases, words may stimulate and shape the development of a non-linguistic category, while in others a pre-linguistic concept may be formed and function as a facilitator in word learning.

Second, it has been suggested that the concepts of young children are more concrete, less general, and more tied to perceptual attributes than those of older children (Nelson 1974b). In fact, traditional theories of cognitive developmental progress posit a striking age-related change in the criteria used to categorise objects. According to those theories, pre-schoolers are “perceptually-bound” and they group or sort objects on the basis of similarity of salient physical features (e.g. shape, size, colour, etc.) whereas older children and adults categorise on the basis of more abstract, symbolic, or conceptually important attributes. However, many empirical and theoretical studies have demonstrated that pre-schoolers (even as young as almost 2 years old) are not limited to categorising according to overall physical similarity (Deák & Bauer 1995, 1996; Gelman & Coley 1990; Gelman & Markman 1986, 1987; Gelman & O’Reilly 1988; Keil 1987; Markman 1987; Mervis 1987; Miller 1973; Nelson 1973, 1974a, 1974b; etc.). The traditional view (e.g. Bruner 1957; Bruner *et al.* 1966; Clark 1973; Daehler *et al.* 1979; Flavell 1985; Keil & Keller 1987; Melkman *et al.* 1981; Saltz *et al.* 1972; Tversky 1985; etc.) has been increasingly supplanted by the view that even very young children can categorise according to more conceptual non-obvious similarities as well as perceptual cues.³²

Third, it has also been suggested that the referential range of children’s categories is different from that of the categories of adults. An important aspect of children’s categories is that they are often broader or narrower than adults’ but seldom completely overlapping. Many category production tasks, language diary studies, sorting tasks, etc., have provided evidence for both overextension and underextension (e.g. Alexander & Enns 1988; Anglin 1977; Bowerman 1978, 1980; Nelson 1974b; Rescolda 1980; Rosner & Hayes 1977; Saltz *et al.* 1972; Saltz *et al.* 1977). These studies typically focus on the range of instances to which a particular category applies and the gradual extension of the category to encompass the full range of adult uses. Reasons like ignorance of adult culturally significant attributes and functions of objects in the adult world and/or different salience of particular attributes to children and adults, etc. usually explain this dissimilarity. In addition, it has been claimed that even if the same items are

included in a category (i.e. same extensional range), the cognitive representation of those categories usually differs in children and adults and that the internal representation of children's categories eventually changes to adapt to adult standards as a result of reassessing the importance of some features.

Fourth, although the ability to understand taxonomic hierarchies is present as early as two or three years of age (Blewitt 1994), the work of several researchers (e.g. Callanan 1983; Callanan & Markman 1982; Horton 1982; Markman 1978, 1979, 1983, 1984, 1987, 1989; Markman & Hutchinson 1984; Markman & Seibert 1976; Markman *et al.* 1980; Scholnick 1983; Tversky 1985) suggests that children find it easier to learn and sort according to "thematic" relations rather than "taxonomic" or "class-inclusion" relations. Thematic relations are characterised by a spatial, temporal or causal connection. For example children prefer to group "spider" and "web" together rather than "spider" and "ladybird" because 'the spider lives in a web' or they prefer to group together "dog" and "doghouse" rather than "dog" and "horse" because 'the dog lives in the doghouse'.³² However, as some more recent research has found evidence against that assumption (e.g. Bauer & Mandler 1989; Waxman and Hall 1993), it is presumably safer to conclude that the preference for one or another type of relation may depend on different factors or conditions like, for example, the presence vs. absence of novel nouns in the task (Markman & Hutchinson 1984; Waxman 1990; Waxman *et al.* 1991), the presence or absence of shared perceptual features amongst the members of taxonomically related categories (Fenson *et al.* 1989), etc.

Fifth, as far as taxonomic organisation is concerned, research on children's taxonomies has generally endorsed the standard Roschian view that pre-school children have little difficulty categorising objects at the basic level, but do not generally group different basic level objects into adult-like superordinate categories (e.g. Rosch *et al.* 1976; see also section 1.2.1.3.5.). Different studies have provided support for the developmental primacy of the basic level over the superordinate or subordinate levels, usually acquired in this order after basic-level categories have been learned (Anglin 1977; Blewitt & Durkin 1982; Callanan 1983; Daehler *et al.* 1979; Horton & Markman 1980; Mervis & Crisafi 1982).

Finally, evidence for the abstraction of criterial features and therefore, learning classical concepts has been difficult to obtain in children (e.g. Bruner *et al.* 1966). On the contrary, some studies have shown that children seem to learn central tendencies of categories as predicted by probabilistic categorisation models (Posnansky & Neumann 1976) or even learn specific exemplars (Kossan 1981). Irrespective of whether children learn and represent categories in

³² Much work on children's inductive inferences after classification in categorisation testifies to children's use of non-perceptual similarities in categorisation (Gelman 1984, 1988; Gelman & Coley 1990; Gelman & Markman 1986, 1987, 1989; Gelman *et al.* 1986; Kalish & Gelman 1992; Markman & Hutchinson 1984).

³³ Such studies reveal that there is a conceptual continuum of thematic relatedness that goes from the most cohesive relations found in partonomies of the kind "finger-hand", through those that are the referents of collective nouns like "family" (i.e. group of people) or "forest" (group of trees) to thematic relations of the "boy-and-his-bike", "dog-and-its-doghouse" type.

either probabilistic or exemplar modes, a common finding of categorisation studies on children is that the latter are sensitive to within-category typicality differences as revealed by many different categorisation tasks (Aiken & Williams 1973; Anglin 1977; Bjorklund *et al.* 1982; Bjorklund *et al.* 1983; Blewitt & Durkin 1982; Bowerman 1978, 1980; Carson & Abrahamson 1976; Duncan & Kellas 1978, 1982; Erreich & Valian 1979; Hupp & Mervis 1982; Keller 1982; Mervis & Pani 1980; Mervis *et al.* 1975; Mulford 1979; Rescolda 1980; Rosch 1973b; White 1982; Williams *et al.* 1977).

1.3.2. Categorisation in Pre-Verbal Infants.

Developmental psychologists have also turned their attention to the study of the categorisation abilities of pre-verbal (or pre-linguistic) infants. This research is deemed important, for example, for a better understanding of the initial stages of cognitive development (Eimas 1994) or for the understanding of infants' earliest acquisition of vocabulary (e.g. Gopnik & Meltzoff 1987, 1992; Roberts & Horowitz 1988; etc.).

Although the study of categorisation in infants may appear to be methodologically difficult due to the very early age of the subjects, exciting methodological advances have actually facilitated the exploration of categorisation abilities of infants prior to acquiring language. A common method to study infant categorisation is through the observation of infants' sequential touching and/or grouping of objects in object-manipulation tasks (e.g. Bauer *et al.* 1995; Gopnik & Meltzoff 1987; Mandler & Bauer 1988; Mandler & McDonough 1993; Mandler *et al.* 1991; Nelson 1973; Poulin-Dubois *et al.* 1995; Starkey 1981). In these tasks, infants are presented with collections of objects that can be sorted into distinct groups or touched sequentially. An alternative method is to observe infants' spontaneous manipulation of objects in real-world situations and observe functional differential responses to different objects. Mervis (1985), for example, inferred the learning of the category "horn" by observing an infant systematically blowing different toy-like horns on different occasions.

However, very young infants find difficulties in actively manipulating groups of small objects.³⁴ Consequently, an alternative and very frequently used technique is the visual habituation and related novelty preference paradigm (e.g. Cohen & Strauss 1979; Cohen & Younger 1983; Colombo *et al.* 1987; Quinn 1987; Roberts & Horowitz 1986; Ross 1980; Strauss 1979). In this paradigm, infants are initially given several familiarisation trials with different members of a certain category. Then they are shown either a novel member of the category, a novel member of a different category or both at once. If the infants have formed the

³⁴ This is due to infants' motor skill difficulties to grasp more than one object at a time, a talent not common until 9 months, which is very useful for object grouping, and to infants' difficulty in focusing their attention on objects long enough to grab them (Starkey 1981).

category, then they usually spend significantly more time looking at a stimulus from a novel category than from the familiar one because it is somehow more appealing.

Using any of these procedures, numerous studies have so far provided abundant evidence that infants of even as early an age as 3 months are able to learn a large number of natural, artefactual and artificial categories when exposed either to physical objects or to slides representing those objects. Amongst the natural categories infants learn are categories of animals like “cats”, “dogs”, “birds”, “horses”, etc. (Behl Chadha 1994, 1996; Eimas & Quinn 1994; Eimas *et al.* 1994; Mandler & Bauer 1988; Quinn & Eimas 1996a; Quinn *et al.* 1993; Younger 1990a), plants (Mandler *et al.* 1991), colours (Bornstein 1987; Bornstein *et al.* 1976), human faces (Cohen & Strauss 1979; Fagan 1976), food (Ross 1980), etc. Artefactual categories include different types of vehicles (Mandler & Bauer 1988; Mandler *et al.* 1991; Nelson 1973; van de Walle 1997), furniture (Behl Chadha 1994, 1996; Mandler *et al.* 1991; Rakison & Butterworth 1998; Ross 1980), etc. In addition, some studies have been conducted using artificial categories. These include both completely arbitrary categories like dot-patterns of geometric forms (Bomba & Siqueland 1983; Quinn 1987; Younger & Gotlieb 1988), geometric figures (Lécuyer 1991; Lécuyer & Poirier 1994; Starkey 1981) and artificial categories that mimic natural ones like leaf-like forms (Colombo *et al.* 1990), drawings of schematic human faces (Sherman 1985; Strauss 1979; Younger 1992), drawings of schematic animals (Colombo *et al.* 1987; Roberts 1988; Roberts & Cuff 1989; Roberts & Horowitz 1986; Younger 1985, 1990b; Younger & Cohen 1983, 1986), etc.

Finally, a few abstract categories like the spatial orientations “above vs. below”, “left vs. right”, “inside vs. outside” (Behl-Chadha & Eimas 1995; Quinn 1994) or line orientations like “oblique” and “vertical” (Bomba 1984; Quinn & Bomba 1986) have also been studied. The findings of these studies suggest that infants also learn these categories easily.

1.3.2.1. Categorisation in Infants: Issues.

In light of the demonstration that humans under one year of age are able to categorise many different types of stimuli, infancy researchers have begun to devote themselves to general issues in categorisation that have also attracted the attention of researchers studying categorisation in adults and children.

First, there has been some interest in whether pre-linguistic infants rely exclusively on perceptual features to perform categorisation or whether they have recourse to, at least, some kind of primitive conceptual features like “living thing”, etc. In this respect, although it has been argued that infants rely on non-perceptual features to perform categorisation decisions (e.g. Mandler & Bauer 1988; Mandler & McDonough 1993; Mandler *et al.* 1991), it is very likely that both the age of subjects and the nature of the stimuli presented to them (i.e., static pictorial

instances of categories) make it improbable that infants categorise on any basis other than perceptual cues (e.g. Quinn & Eimas 1996; Rakison & Butterworth 1988). In other words, although infants engage in some types of categorisation involving perceptual similarity prior to having extensive experience, which indicates that this ability has biogenetic bases (Bornstein 1984: 327), we have to wait for older subjects in order to find evidence of more functional or conceptual basis of categorisation (see 1.3.1.1.).

Second, the studies so far conducted with infants seem to favour the claim that infants usually abstract a probabilistic category representation rather than store all the individual instances of a category to which they are exposed. These abstract representations apparently consist of the average or modal values of the features of category members. For example, infants seem to abstract a mean prototype of schematic human faces (Strauss 1979), dot patterns of geometric forms (Bomba & Siqueland 1983; Quinn 1987; Younger & Gotlieb 1988), or drawings of schematic animals (Younger 1985, 1990b). They can also abstract a modal prototype of drawings of schematic male faces (Sherman 1985). Apparently, abstracting a prototype is useful to infants because their memory abilities are not fully developed and their memory load can be significantly reduced by representing information in summary form (Quinn 1987; Younger & Gotlieb 1988). However, although researchers like Sherman (1985) have argued against an exemplar encoding in infants, Bomba and Siqueland's (1983) findings show that it may be as difficult to distinguish experimentally in infants between models of categorisation as it is in the adult literature. Thus, exemplar encoding cannot be discarded in infants.

Third, it has been found that whether abstract summary representations or exemplar memories come to represent infants' prelinguistic categories, infants are sensitive to within-category differences. For example, when infants learn the category "bird" (Roberts & Horowitz 1986), prototypical birds (e.g. "sparrow", "robin", "bluejay") facilitate the learning of the category and are selected as the "prototypes". Apparently, all members are not treated equivalently but instead, there are some of them that appear to be perceptually more salient and therefore, more "prototypical" (see also Bauer *et al.* 1995). This parallels findings obtained with older children (e.g. Mervis & Pani 1980; Hupp & Mervis 1982; etc.). In addition, several studies have shown that infants, like adults, are sensible to correlated attributes in category members and that they classify entities into categories on the basis of clusters of correlated attribute values (Younger 1984, 1985, 1990a, 1992; Younger & Cohen 1983, 1985, 1986).

Finally, recent research has also been concerned with the level of abstraction of infants' categories. Research has clearly shown that infants can form categories that seem to fall within the adult basic level of categorisation. These categories are natural categories such as "cats" or "dogs" (Eimas & Quinn 1994; Eimas *et al.* 1994; Quinn & Eimas 1996; Quinn *et al.* 1993), or artefactual categories such as "chairs" and "couches" (Behl Chadha 1994, 1996). However,

infants can also form categories that closely correspond to the adult superordinate level of categorisation. These include natural categories such as “plants” (Mandler *et al.* 1991), “mammals” (Behl-Chadha 1994, 1996) or “animals” (Bauer *et al.* 1995; Mandler & Bauer 1988; Mandler & McDonough 1993; Mandler *et al.* 1991; Rakison & Butterworth 1998; Ross 1980; van de Walle 1997) and artefactual categories such as “furniture” (Behl-Chadha 1994, 1996; Mandler *et al.* 1991; Rakison & Butterworth 1998; Ross 1980) or “vehicles” (Bauer *et al.* 1995; Mandler & Bauer 1988; Mandler & McDonough 1993; Mandler *et al.* 1991; Rakison & Butterworth 1998; Van de Walle 1997).

Given that categories on a continuum of taxonomic abstraction seem to be learned by infants, some research has been concerned with whether early basic-level categories cohere to form superordinate-level categories or whether basic-level categories evolve from original superordinate-level categories. Much of this work has been in response to the Roschian account of taxonomic development (e.g. Rosch *et al.* 1976). This theory suggests that categories are initially formed at the basic level and that superordinate categories develop later when the infant groups together separate basic-level representations so that hierarchical classification systems are formed (see 1.2.1.3.5., & 1.3.1.1.). However, this view is again not devoid of problems (Mandler 1997; Mandler & McDonough 1993). The more discrepancies are found between adult categories and child-basic categories, the more problematical becomes the description of basic-level categories as the first to be formed. This is particularly true in cases in which a “child-basic” category is so broad as to appear more like a superordinate category than a basic level (Mandler & Bauer 1988: 249). Carolyn Mervis (1980, 1984, 1985, 1987; Mervis & Mervis 1982) has already demonstrated this extensively. Her most typical example is the category “kitty”, which includes, for children, all felines rather than cats *per se* (Mervis 1987).³⁵

Presumably for these reasons and on the basis of evidence obtained from studies involving object manipulation by older infants (9-24 months old), Mandler, Bauer and her co-workers (e.g. Bauer *et al.* 1995; Mandler 1997; Mandler & Bauer 1988; Mandler & McDonough 1993; Mandler *et al.* 1991; see also Poulin-Dubois *et al.* 1995) proposed an alternative view. According to them, infants first form “global” categories more comparable to what are usually termed “superordinate” categories than they are to basic-level ones. The term “global” instead of “superordinate” was used to refer to these broader categories since the class-inclusion relations implied by the latter term are most likely absent in the young child’s conceptual organisation. These global categories later get differentiated to yield “basic level”-like categories. In their studies, for example, infants failed to distinguish between some basic-level

³⁵ The example of “kitty” reveals how dangerous it is to draw conclusions about the nature of early concepts on the assumption that linguistic usage fairly directly reflects underlying conceptual structure (Mervis 1987). Yet, as Brown (1958) pointed out, children’s first words largely reflect the language that is spoken to them and the fact that those words are at the basic level does not in itself provide evidence about the underlying conceptual system onto which the language is being mapped. Basic-level terms could be mapped either onto relatively undifferentiated, global concepts or onto a more detailed conceptual base.

categories such as “dogs” and “horses” but successfully formed the superordinate category “animals”.

However, with regard to the debate regarding the direction of category acquisition, it is very likely that neither position is true. Basic level categories do not always precede superordinate-level categories and superordinate-level categories do not always precede basic-level categories. Instead, it seems that infants (and children) are capable of learning many different basic-level and superordinate-level categories more or less similar to the ones adults have (Eimas & Quinn 1994; Quinn & Eimas 1996; Quinn & Johnson 1997; Quinn *et al.* 1993).

1.3.3. Categorisation in Non-Human Species.

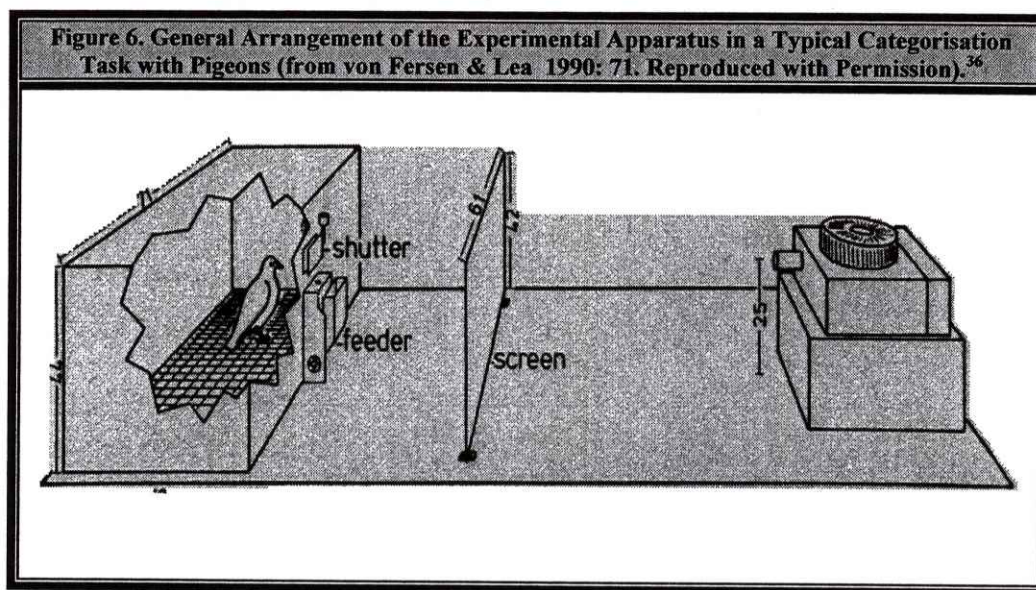
It is apparently incontrovertible that even the simplest organisms do categorise physical entities and events; for example, even the simplest creatures discriminate between noxious and nonnoxious substances, dangerous and safe environments, etc. This statement is supported, furthermore, by a growing body of experimental research showing that the ability to categorise is not limited to the human species. On the other hand, the question of whether animals can acquire concepts has nowadays been replaced by questions about what species can acquire what concepts, and under what conditions. Research on the categorisation abilities of animals has mainly focused on avians and primates.

1.3.3.1. Categorisation in Avians.

Several avian species such as bantams (Ryan 1982), budgerigars (Trillmich 1976), blue jays (Pietrewicz & Kamil 1977), chickens (e.g. Ryan & Lea 1994) or the African grey parrot (Peppenberg 1987) have been reported to categorise many natural visual categories. However, the categorisation abilities of pigeons have been studied most extensively than those of any other bird.

In a typical experiment, birds are trained to peck for reinforcement (e.g. typically access to mixed grain) whenever they see slides presented on a screen that contain members of a particular category and not to peck whenever the slides do not contain instances of the category. For example, when learning the concept “humans”, if a slide representing a scene with one or more humans appears on a screen, key pecks are reinforced (rewarded with food). On the contrary, whenever pigeons encounter a slide showing a scene without a human, their key pecks are never reinforced. Birds are thus initially trained to discriminate between two classes of slides distinguished only by the presence or absence of members of the category. In a subsequent session, birds are tested with a variety of new instances and their generalisation from a set of familiar training slides to (usually) a set of completely novel testing slides or

photographs is examined. If generalisation occurs, it indicates that their behaviour goes beyond any individually learned associations between the particular training stimuli and responses and that it is a true categorizing behaviour.



Using this technique, pigeons have been shown to form many different concepts or categories. As mentioned above, most of these involve restricted classes of visual stimuli and their much broader complements: “humans” vs. “nonhumans”, “tree” vs. “nontree”, “water” vs. “nonwater”, and so forth.

After appropriate training, pigeons can classify new exemplars of many natural categories such as “humans” in general (Astley & Wasserman 1992; Edwards & Honig 1987; Herrnstein *et al.* 1976; Malott & Siddall 1972; Siegel & Honig 1970), a particular human (Herrnstein *et al.* 1976), other pigeons (Poole & Lander 1971), fish (Herrnstein & de Villiers 1980), the common kingfisher, birds and animals (Roberts & Mazmanian 1988), bird and mammal (Cook *et al.* 1990), cats (Bhatt 1988; Bhatt *et al.* 1988; Wasserman *et al.* 1988), trees (Herrnstein 1979, 1985, Herrnstein *et al.* 1976; Vaughan 1988), other birds (McLaren *et al.* 1994), oak leaves (Cerella 1979), flowers (Astley & Wasserman 1992; Wasserman *et al.* 1988), bodies of water (Herrnstein *et al.* 1976), locations (Kendrick 1992), etc. In addition, pigeons have been shown to form categories made up of human artefacts, although these categories are more difficult to learn than natural ones (Herrnstein 1985). Pigeons accurately learned the categories “letter A” and “number 2”, when each is presented in different typefaces or handwriting (Morgan *et al.* 1976), different man-made objects (Lubow 1974), “cars” and “chairs” (Astley & Wasserman 1992; Bhatt *et al.* 1988; Wasserman *et al.* 1988), buildings (von Fersen & Lea 1990), etc.

Pigeons can also learn artificially constructed categories. Some of these studies include artificial polymorphous categories constructed to mimic the feature structure of natural categories like schematic drawings of faces (Huber & Lenz 1993, 1996), fake pigeons (e.g. Lea & Ryan 1990; Lea *et al.* 1993), or pseudobutterflies (Jitsumori 1996). Other studies have employed completely artificial categories like shapes (circles or triangles) superimposed on backgrounds of different colours (Jitsumori 1993; Lea & Harrison 1978), black and white checkerboard patterns (Aitken *et al.* 1996), three-vertical bar patterns differing in height (Pearce 1989), rectangles of different sorts (Aydin & Pearce 1994), dot patterns resembling a triangle (Watanabe 1988) or degraded figures produced by removing various segments of a triangle and a square (White *et al.* 1993).

Finally, pigeons have also been proven to learn more abstract concepts. One type of such concepts is the “same-different” concept (Cook *et al.* 1995; Cook *et al.* 1997; Santiago & Wright 1984) in which pigeons are trained to discriminate between pairs of slides containing members of the same category each from pairs of slides that contain members of different categories (see Pepperberg 1987 for similar results with the African grey parrot). Another type of such concepts is the “familiar vs. novel” concept in which pigeons are trained to discriminate between slides containing objects already encountered from slides containing new ones (MacPhail & Reilly 1989).

1.3.3.2. Categorisation in Primates.

Although a few studies have been conducted with other mammal species like dolphins (e.g. Forestell & Herrman 1988; Herman *et al.* 1989), the study of the categorisation abilities of mammals other than humans has overwhelmingly focused on primates. These include different species of monkeys and anthropoid apes like chimpanzees and orang-utans.

The categorisation tasks that primates perform are very similar to those used with birds. Primates are typically trained to press on a screen or press a lever for reinforcement (typically a sucrose or banana pellet) when they see scenes that contain examples of a category and not to press when they regard scenes that do not. After a training session, they are transferred to a test session in which they also have to generalise this discrimination to novel instances that do not exemplify the category. Again, if generalisation occurs, the animals are assumed to have learned the category under investigation.

Several species of monkeys have been studied. Again, like pigeons, monkeys have been found to learn many natural categories. Squirrel monkeys have been reported to form the categories “common kingfisher”, “bird”, and “animal” (Roberts & Mazmanian 1988), cebus

³⁶ Note that the screen between the slide projector and the operant chamber was translucent, and the pecking key was transparent. Thus, the pigeon looked through the key at the image projected on the screen. The inset shows the front view of the

monkeys the category “person” (D’Amato & Van Sant 1988), stumptailed monkeys the categories “humans” and “monkeys” (Schrier *et al.* 1984) and rhesus monkeys the categories “primate (human and nonhuman) faces”, “fruit”, “flowers”, and “trees” (Sands *et al.* 1982), “humans” (Schrier & Brady 1987; Yoshikubo 1985) and a particular species of monkey, “rhesus monkey” (Yoshikubo 1985) and hue categories (Sandell *et al.* 1979).

Studies dealing with artefactual and artificial categories are less numerous than those with natural categories. However, there is some evidence that primates can also learn such types of categories. Stumptailed monkeys, like pigeons (Morgan *et al.* 1976), were able to learn the categories “letter A” and “number 2” (Schrier *et al.* 1984), and rhesus monkeys learned artificial categories made up of arrays of symbols (e.g. triangles or circles) on coloured backgrounds (Jitsumori 1994). Finally, primates are also capable of learning more abstract categories. Rhesus monkeys (Sands *et al.* 1982; Shyan *et al.* 1987; Wright *et al.* 1984), chimpanzees and orangutans (King 1973) have been shown to learn complex “same-different” concepts and chimpanzees are able to learn the referential symbolism of many signs in American Sign Language like those corresponding to “dog”, “baby”, “flower”, etc. (Gardner & Gardner 1984).

1.3.3.3. Interspecies Comparison.

Research on the categorisation abilities of non-human species has been interested more or less explicitly in revealing the similarities and differences between the ability to categorise in humans and in non-human species (see e.g. Kuhl 1987; Mackintosh 1995 for discussions).

Although some studies have compared categorisation performance in different non-human species,³⁷ in general, researchers have been more interested in comparing categorisation data obtained from different animal species with the experimental results obtained with human beings. An important motivation for this line of research is the discovery of whether there are species-specific categorisation abilities in the different species. In this respect, the research has focused on different interesting issues.

First, the question of whether the categories animals and humans learn are equivalent or not has attracted a great deal of attention. Herrnstein, Loveland, and Cable (1976) argued that pigeons and people generalise similarly at least as far as the category “person” is concerned. Lea and Harrison (1978) demonstrate that some of the necessary conditions for polymorphous concept formation as it exists in people (i.e. the ability to respond similarly to a set of disparate stimuli not linked by any single, simple, feature, and to generalise from given instances to new ones) also exist in pigeons. Also, macaques and humans categorise the photic spectrum in a

intelligence panel.

³⁷ Stumptailed monkeys, for example, showed lower performance in learning the concept “humans” (Schrier *et al.* 1984) than had been reported for pigeons (Malot & Siddall 1972; Siegel & Honig 1971).

similar fashion (Sandell *et al.* 1979). Finally, Kluender and his co-workers (Kluender *et al.* 1987; Kluender *et al.* 1998) found that animals (Japanese quail and European starlings) and human subjects (both adults and infants) form phonetic categories in a very similar manner.

However, Herrnstein (1979) points out that although it is hard to deny pigeons a concept of sufficient generality such as “fish” (e.g. Herrnstein & deVilliers 1980) to correspond at least in part to the human concept of “fish”, this conclusion stops far short of saying that the pigeons’ concept is equivalent to the humans’. Apparently, pigeons and people converge on similar but not equivalent categories and categorisation processes. In fact, Lea and Harrison (1978: 535) argued that despite pigeons’ successful performance in categorisation tasks, it is very difficult to say that they have “learned” a concept in the same way as humans do. This is so, they claimed, because there is a persistent tradition in human concept formation experiments that “something else” is required if concept formation is to be assumed, namely, the ability to give a name to a category. Not surprisingly, “most concepts are associated with a general descriptive name or label” (Bourne 1966: 2) and apparently animals have no linguistic abilities (but see Gardner & Gardner 1984). In addition, it is important to remember that the categorising behaviour of animals is usually restricted to what has been called “primary stimulus generalisation” or the grouping of instances into a category on the grounds of their physical resemblance. However, the generalisation to new members of a category without regard to similarity, without the requirement that members of a category be perceptually similar (“secondary stimulus generalisation”, also called “mediated generalisation”) is severely limited in most non-human species (Lea 1984).

Second, the question of whether categorisation processes are simpler in animals than in humans has also received considerable attention. Although we might have a priori assumptions that categorisation abilities in animals are very simple, if not rudimentary, when compared with those of adults, some experimental findings seem to question this assumption. Apparently, animals seem to use strategies similar to those of human subjects. For example, although initial studies found inconclusive evidence as to whether pigeons were able to learn probabilistic concepts (Bhatt *et al.* 1988; Edwards & Honig 1987; Jitsumori 1993; Kluender *et al.* 1998; Lea & Harrison 1978; Pearce 1988, 1989; Watanabe 1988; White *et al.* 1993), more recent research has demonstrated the opposite. For example, some studies have argued for the formation of abstractions representing the central tendencies of schematic faces (e.g. Huber & Lenz 1996) and artificial polymorphous stimuli of different kinds (e.g. Aydin & Pearce 1994; Jitsumori 1996; von Fersen & Lea 1990). In addition, evidence has been presented that pigeons are also able to learn exemplars individually rather than abstract prototypes (e.g. Bhatt *et al.* 1988; Cook *et al.* 1990; Huber & Lenz 1996). Therefore pigeons appear to be able to use both strategies in learning, storage and classification: exemplar encoding and abstraction of a probabilistic representation.



Other researchers have suggested that pigeons' conceptual abilities are far more advanced and closer to those of human beings than hitherto suspected (see Wright *et al.* 1990 for an insightful discussion). Pigeons demonstrate, for example, an ability to successfully perform a four-category classification task in which they categorise different examples of the categories "cat", "flower", "car" and "chair" by pecking the appropriate one of four discrete response keys when shown colour slides of those objects (e.g. Bhatt *et al.* 1988; Wasserman *et al.* 1988). This ability, more complex than the mere discrimination between the presence and absence of a single class of objects more closely resembles human categorisation abilities. In addition, pigeons have occasionally bettered humans in categorisation tasks. Jitsumori (1993), for example, trained pigeons and humans to classify sets of stimuli defined out of a combination of three different features. He found out that pigeons showed complete transfer to stimuli with all three positive or negative features and even to stimuli that had one of the 3 features replaced with a novel one. However, although human subjects quickly learned the task, they failed to successfully generalise their responses to new unreinforced stimuli.

However, perhaps the most convincing result showing the somehow complex categorisation abilities of pigeons and monkeys is their capacity to learn categories at different levels of abstraction according to human taxonomies. Wasserman and his co-workers (Wasserman *et al.* 1992) found that pigeons learned to group both chairs and cars under a higher category somehow similar to the adult category "furniture". They found that some sort of secondary stimulus generalisation (i.e. non-similarity based categorisation) was available to pigeons. Similarly, pigeons appeared to have learned subordinate categories of the higher-level category "tree" in Vaughan's (1988) study in which 40 slides of trees were divided into two assortments. Also, rhesus monkeys could distinguish instances of their own species from those of a related one, the Japanese macaque (Yoshikubo 1985). The categories "rhesus monkeys" and "Japanese macaque" in Yoshikubo's (1985) study are more specific than that of "primate (human and nonhuman) faces", which rhesus monkeys have also been shown to have mastered (Sands *et al.* 1982).

The hypothesis that animals can learn categories at different levels of generalisation receives further support from Roberts and Mazmanian's (1988) study. This investigation compared categorisation data obtained from pigeons, squirrel monkeys, and human subjects in comparable tasks. The study shows that non-human animals can learn concepts that correspond to different levels of abstraction in human taxonomies but that there are differences amongst species. For example, monkeys and pigeons learned to identify novel exemplars at either the subordinate or most concrete level (the concept "common kingfisher") and the superordinate or most abstract level (i.e. the concept "animal") with pigeon's better overall performance on the concept "common kingfisher" and monkey's on the more abstract (i.e. discrimination between animals in general or no animals). However, both monkeys and pigeons learned the basic or

intermediate level of the taxonomy, represented by the category “bird”, but did relatively poorly on this task. Humans did better with this level, also quickly learned the more abstract but had difficulties with the most concrete level. The differences between pigeons, monkeys and people suggest a definite lack of isomorphy between humans and animal concepts as far as the classification of entities at different levels of generalisation in human taxonomies is concerned.

To sum up, research on categorisation in animals has uncovered striking similarities between categorisation abilities in human and nonhuman species. Research has also revealed several differences. However, it can be claimed that, despite differences, and even if nonhuman animals categorise classes of stimuli on bases other than perceptual similarity in a limited way, the level of conceptualisation that pigeons have thus far shown is sufficient for claiming the existence of striking categorisation abilities in non-human species. Clearly, many questions remain to be answered. Much more systematic work is needed in order to understand more about animal categorisation and its relationship with human categorisation but fortunately, the groundwork for this research has already been laid.

CHAPTER II:

ON THE CATEGORISATION OF SOUNDS.

2.1. The Categorisation of Sounds: Experimental Evidence.

As mentioned in section 1.1.5., human beings seem to have categories for almost anything. However, a great deal of the categorisation literature has been concerned with natural, artefactual or artificial categories like “birds”, “chairs”, “dot patterns”, etc. One of the most common characteristic of these categories is that the stimuli that instantiate them in the experimental studies so far conducted are visual. Such stimuli can be exclusively perceived through the sense of sight and identified without the aid of any other sensory mode of perception. Such stimuli are usually either pictures, drawings or replicas of certain real-world, physical entities or actual instances of them. However, natural, artefactual, or artificial categories need not be exclusively or mainly visual. If a category implies any coherent class or grouping of entities or events that occur in the world, it is obvious that there are thousands of classes of sounds that are equally encountered by categorising organisms and that could themselves be considered as categories instantiated by a variety of acoustically different events that are nevertheless treated as the same. In fact, according to Nathan (1996: 112), “sounds... are categorized in the same way as all other things in the world are”.

“Sound” categories might include more or less “natural” categories like “infants’ cries”, “dog barks”, “the sound of waves”, etc., or “artefactual” sound categories (in the sense of being produced with some man-made human artefact) like “bell rings”, “honks”, “gun shots”, etc. In addition, artificial sound categories might be exemplified by different types of sounds produced with electronic speech synthesisers. However, it seems that one of the most important source of sound categories is language. For example the different vowels that we find in words like “far”, “car”, “star”, may, to some extent, be considered as occurrences of approximately the “same” category. In this respect, it seems that the sounds of people’s speech would fall within the class of natural sound categories. The sounds of speech can also be imitated with the use of speech synthesisers to create artificially constructed sounds that mimic natural or actual sounds produced by human beings.

Sound categories differ from the typical visual categories in their sensory mode of perception. However, there seems to be a different motivation in the choice of either visual or auditory categories to conduct experimental studies, a motivation best explained by making reference to the interesting distinction (though controversial as any other binary distinction) established by Medin and Barsalou (1987) between “general knowledge categories” (GKC) and “sensory perception categories” (SPC). The former are the typical natural or artefactual categories studied by investigators interested in semantic analysis, memory organisation, abstract thought, knowledge representation and processing. These include typical natural and artefactual categories like “birds”, “chairs”, etc. The latter are those categories studied by investigators interested in sensory processes and perception. Examples of sensory perception categories include categories of speech sounds (e.g. / k /, / æ /, / t /), non-speech sounds (e.g. sobs, snores, artificially constructed sounds)¹, colours (e.g. red, blue, green)², and many of the completely artificial categories mentioned in section 1.1.5. Sensory perception categories include both visual and auditory categories.

The purpose of the rest of this section is to review all the literature that has addressed the categorisation of naturally-occurring sounds of English or artificially constructed (i.e. synthesised) sounds that mimic natural English sounds. In addition, special emphasis will be put on the categorisation techniques used.

2.1.1. The Categorisation of Sounds by Adults.

The study of the categorisation or classification that adults make of sounds has a long history in the fields of speech perception and experimental phonology. Different experimental techniques have been used to study the categorisation of sounds by adults. The most common ones are the phoneme monitoring task, the absolute identification task, the differential discrimination task in either its same-different or ABX versions, and the concept formation paradigm.

The phoneme monitoring technique is a detection task that asks subjects to push or press a button or key as fast as possible on hearing a stimulus that contains a previously specified target sound or “phoneme”.³ The target sound usually (but not necessarily) appears at the beginning of certain words or non-words presented in word lists or embedded in sentential contexts.

The paradigm was developed by Foss and his associates (e.g. Foss 1969; Foss & Lynch 1969) as a technique for measuring on-line processing difficulty during sentence

¹ These include totally arbitrary auditory categories that sounded nothing like speech (e.g. Cutting 1982; Cutting & Rosner 1974; Lotto 2000; Mattingley *et al.* 1971; Rosen & Howell 1981, 1983).

² See, for example, Goldstone (1995), Nosofsky (1988b) or Rosch (1973a, 1974b, 1975c).

comprehension. Since then, many phoneme monitoring studies have been carried out with different goals. Amongst the typical issues researchers have been concerned with are the discovery of the basic unit of speech perception or identification (e.g. Foss & Swinney 1973; Healy & Cutting 1976; Mills 1980; Savin & Bever 1980; Segui *et al.* 1981; Swinney & Prather 1980)⁴, whether reliable consonant recognition is vowel dependent (Diehl *et al.* 1987), the effect of stress on phoneme recognition (Cutler 1970), the cohesiveness of syllable onsets (Treiman *et al.* 1982), etc.⁵ However, the most important goal of phoneme-monitoring experiments has been to find out whether subjects are able to respond before the word carrying the target phoneme is identified or afterwards.

In this respect, there appear to be three main views. First, subjects respond to target phonemes only after having recognised the word in which the target is embedded (e.g. Foss & Swinney 1973; Morton & Lang 1976; Rubin *et al.* 1976). For example, Morton and Long (1976) showed that with an identical preceding context, the same initial phoneme targets contained in words having a high probability of occurrence (e.g. “he sat reading a *book* until it was time to go home for his tea”) were responded to significantly faster than those in which the word containing the phoneme was less contextually probable (e.g. “he sat reading a *bill* until it was time to go home for his tea”). Second, phoneme-monitoring responses are always based on prelexical representation of the input. According to this view, subjects base their responses only on acoustic-phonetic representation computed from the signal, and they do not exploit lexical information (e.g. Eimas & Nygaard 1992; Foss & Gersbacher 1983; Newman & Dell 1978). Finally, a more interactive view claims that information from both levels (prelexical and postlexical) plays a role in phoneme identification. Which level is the most prevalent one depends on a host of factors like task difficulty, task monotony, the phonetic similarity and nearness of sounds similar to the target one, the lexical frequency of the target-bearing word, its lexical status -whether it is a word or a nonword-, etc. (Cutler *et al.* 1987; Dell & Newman 1980; Eimas *et al.* 1990; Foss & Blank 1980; Foss & Lynch 1969; Foss *et al.* 1980; Frauenfelder & Segui 1989; Segui & Frauenfelder 1986; Segui *et al.* 1981; Stemberger *et al.* 1985). It appears now that only this third class of “interactive” models in which both lexical and prelexical information contribute to the phoneme detection process can account for the large number of otherwise contradictory experimental results found in the phoneme monitoring literature.

Although the phoneme monitoring technique has been extensively used to study the categorisation of sounds, perhaps the most widely used techniques are the absolute

³ As Foss claims, the label “phoneme” in the phrase “phoneme monitoring” has unintended theoretical implications but was simply adopted as part of lab lore (1998: 197).

⁴ Apart from phoneme-monitoring tasks, some studies have used “syllable-monitoring” tasks in which the to-be-identified target is a complete syllable (e.g. “baeb”, “saeb”) and not just a single consonantal or vocalic phoneme (e.g. Foss & Swinney 1973; Healy & Cutting 1976; McNeill & Lindig 1973; Savin & Bever 1980) or “word-monitoring” tasks in which the target is a whole word, usually made up of two syllables (e.g. Foss & Swinney 1973; McNeill & Lindig 1973).



identification paradigm or the differential discrimination paradigm (Weitzman 1993: 141).⁶ Almost every study dealing with the experimental analysis of speech perception in adults has invariably made use of one of these paradigms.

An absolute identification test is an explicit categorisation. The technique involves having subjects listen to a speech series and decide to which out of (usually) two pre-defined phonological categories each of the stimuli from the speech series belongs to (see e.g. Barclay 1972; Davis & Kuhl 1992; Ganong 1980; Healy & Cutting 1976; Lotto 2000; Miller & Micely 1955; Miller & Volaitis 1989, 1992; Oden & Massaro 1978; Samuel 1982).⁷ In a typical experiment, a synthesised speech series in which a phonetically relevant acoustic property is varied so as to range from one phonetic segment to another (e.g. the monosyllabic series / ba / to / pa /, with the / b /-/ p / voicing distinction specified by a change in voice onset time, hereafter VOT) is created.⁸ Then, listeners are presented randomised sequences of the extended series and are asked to decide and indicate (typically pressing either of two buttons or keys) whether each of the instances that they hear belongs to/sounds more like one category (e.g. / ba /) or the other (i.e. / pa /).

There are two main versions of the differential discrimination paradigm. In the *same-different* discrimination task, subjects are asked to indicate whether a pair of sounds presented one after the other are, for them, either the same or different. The pairs of synthetic sounds usually range from acoustically identical stimuli (e.g. A-A, a-a, X-X, x-x), to acoustically different stimuli selected from within the same phonetic category (e.g. A-a, X-x) to acoustically different stimuli from different phonetic categories (e.g. A-X, A-x, a-X, a-x). For example, in a speech series ranging perceptually from / ba / through / pa /, listeners have to respond “same” if they consider both stimuli in a pair as essentially the same segment (e.g. a 10 msec VOT / ba / and a 15 msec VOT / ba /) or different (e.g. a 10 msec VOT / ba / and a 70 msec VOT / pa /). The second type of differential discrimination test, the so-called *odddity* or *ABX* task, involves the selection of the “odd” sound out of a triad which consists of two acoustically different stimuli from different phonetic categories (e.g. A and B), and a third stimulus, X, which is identical either to A or to B. More specifically, subjects are instructed to indicate whether X is equal to A or B by pressing appropriately labelled buttons (e.g. Liberman *et al.* 1967).

⁵ See Connine and Titone (1996) for a review.

⁶ The identification task has sometimes been referred to as “phoneme identification” (e.g. Kidd 1989; Pitt & Samuel 1993) again with, it seems, apparently no theoretical implications.

⁷ Occasionally, three categories are used. In Miller and Volaitis’ studies (1989, 1992), subjects were shown stimuli varying in VOT from / bi / to / pi / to an exaggerated, breathy version of / pi /, designated as * / pi / and stimuli from a similar velar series in a three-choice identification task. In Barclay’s (1972) study, subjects had to identify different items as either / b /, / d / or / g /.

⁸ Although the overwhelming majority of studies have used monosyllabic non-sense CV, VC, CVC syllables, a few studies have used synthesised stimuli that “mimic” real words. More specifically, the speech series usually range from a real English word to a non-word as in the series “fish-fiss” or “kish-kiss” (McQueen 1991), “dice-tice” and “dype-type” (Connine & Clifton 1987), “dash-tash” (Ganong 1980), “duke-tuke” (Burton *et al.* 1989), “check-sheck”, “swib-swim” (Pitt & Samuel 1993), or from one real word to another as in the series “pest-best” (Connine *et al.* 1993) or “beat-wheat-wit” (Miller *et al.* 1997).

These two experimental paradigms (i.e. identification and discrimination), which have dominated speech perception since its inception and are usually employed in the same study (e.g. Barclay 1972; Carney & Widin 1976; Carney *et al.* 1977; Frazier 1976; Liberman *et al.* 1957; Mandler 1976; Miller & Eimas 1989; Pisoni & Lazarus 1974; Pisoni & Tash 1974; Samuel 1977) have been used to address different issues (see Diehl & Kluender 1987 for a review). The most remarkable ones are the issue of categorical vs. continuous perception and the location of category boundaries. On the one hand, these paradigms have contributed to a better understanding of the issue of categorical vs. continuous perception. In this respect, initial studies generally claimed that within-category discriminations were severely limited (e.g. Frazier 1976; Liberman *et al.* 1957; Mandler 1976; Pisoni & Tash 1974; see Repp 1984 for a review). However, subsequent studies demonstrated that, although between-category distinctions may be more salient, subjects are not perceptually limited and can reliably discriminate and respond differentially to stimuli that belong to the same response category (e.g. Barclay 1972; Carney *et al.* 1977; Davis & Kuhl 1992; Hodgson & Miller 1996; Miller & Volaitis 1989; Miller *et al.* 1997; Samuel 1982; Volaitis & Miller 1992). On the other hand, numerous identification and discrimination studies have helped to reveal how certain factors influence the location of category boundaries on physical-stimulus continua. It is usually the case that the endpoint stimuli contain clear exemplars of the to-be-identified (target) phoneme whereas the middle (boundary) ones are perceptually ambiguous, sounding halfway between the two phonemes. The location of category boundaries is mainly due to the influence exerted by other phonetic stimuli preceding or following the members of phonetic categories, to the internal characteristics of the to-be-categorised stimuli or to the influence of the semantic and/or syntactic context.⁹ For example, changes in category boundaries have been observed with variations in syllable-internal duration (Miller & Volaitis 1989, 1992; Miller *et al.* 1997; Summerfield 1981), sentential-level rate (Gottfried *et al.* 1990; Kidd 1989; Summerfield 1981, 1984), stimulus quality (Burton *et al.* 1989; McQueen 1991; Pitt & Samuel 1993), the lexical status of stimuli -whether they are words or non-words (Burton *et al.* 1989; Connine & Clifton 1987; Fox 1984; Ganong 1980), their lexical frequency (Connine *et al.* 1993), the existence of “lexical neighbours” (Newman *et al.* 1997), the position of the phoneme in the target word (Pitt & Samuel 1993), etc.

Finally, another useful technique used in studies on the categorisation of sounds is that which is known as concept formation, an experimental paradigm that typically consists of two parts: a training and a test session. In the training session, subjects learn to discriminate between stimuli that exemplify a to-be-learned sound category and stimuli that do not. The subjects’ task is to guess whether each stimulus they encounter is a member of the category or not, and to

⁹ A full review of all the numerous studies is out of the present concern but see Repp & Liberman (1987) for an insightful summary.

indicate whether they think it is a member or not immediately after hearing the stimulus. Subjects know whether they are right or wrong because they are provided with feedback as to whether each stimulus presented is a member of the category or not. In the test session subjects try to generalise responses to new stimuli but no feedback is provided. The specifics of this experimental technique will be discussed in detail in section 2.3.

One of the earliest applications of the concept formation experimental technique was in the field of experimental phonology to attempt to determine the psychological status of Chomsky and Halle's English vowel shift rule (e.g. Jaeger 1980a: experiment 5, 1984, 1986b; Wang 1985), of the groupings of allophones into phonemes (Jaeger 1980a: experiments 3 & 4, 1980b: experiment 2; Ohala 1983, 1986), and the binary character of feature categories like [+anterior], [+sonorant], and [+voice] (Jaeger 1980a: experiment 6; Jaeger & Ohala 1984). The findings of these experiments are difficult to summarise briefly. Suffice it to say that as far as the vowel-shift alternations are concerned, only five out of the eight originally set up by Chomsky and Halle (1968) are clearly learned as categories with subsequent generalisation to new pairs. In addition, speakers find no problems in classifying allophones as members of the same phoneme category. Finally, features like [+anterior], [+sonorant], etc., do not seem to be strictly binary but show instead some kind of gradient structure. The CF technique has also been used in the field of speech perception to investigate the question of how different vowels must be in order for human beings to learn to categorically distinguish them (Weitzman 1990a, 1990b, 1992, 1993). The most important finding of these studies is that the critical band (a range of frequencies that are integrated by the peripheral auditory system) is not a psychoacoustic boundary in learning to make vowel categorisations.

2.1.2. The Categorisation of Sounds by Children.

The gathering of experimental evidence on the categorisation of sounds by children was originally motivated by a growing interest in discovering the speech units that pre-literate and beginning readers are aware of, and the important and complex relations between phonological awareness and spelling/reading skills. For example, awareness of phonemes is thought to be especially critical for a child in order to learn an alphabetic writing system. If children are not aware, for instance, that spoken syllables like / bæk /, / bɔ:l /, / bʌg / or / buk / begin with "the same sound", they may not fully understand why the printed versions of these words begin with the same letter, why they have to spell the initial sound of those words with the same letter or intend to make similar articulatory gestures in their speech production while reading.

Studies on the categorisation of sounds by children have also used different experimental techniques, some of them ingenious versions of those used with adults. Amongst the most common ones are auditory (phoneme) recognition, common phoneme and overall phonetic

similarity tasks. These three techniques resemble identification tasks used with adults. In addition, some categorisation tasks have used different versions of the discrimination paradigm used with adults. In the rest of this section, a brief description of the techniques used and an account of the most important findings obtained will be presented.

One common technique is the so-called “auditory (phoneme) recognition” task, which is essentially an absolute identification task but one in which there is no second competing category to assign sounds to. This technique requires children to listen to a list of words and tell the experimenter whether each of them contains a (previously familiar) sound or not. For example, a child may be asked which out of a given set of words contains the sound [p] and which do not. Using this technique with synthetic speech stimuli varying in voice onset time, Wolf (1973) investigated children’s perception of voicing distinctions among the stop consonants / p /- / b /, / t /- / d / and / k /- / g /. Results showed that the boundaries between both phoneme pairs were very similar to those already obtained with adults (e.g. Lisker & Abramson 1964). Also, Treiman and Bruck (Bruck & Treiman 1990; Treiman 1985a, 1992: 73-74) found that children were less able to recognise two syllable-initial consonants (e.g. / f / and / s /) when they were part of a cluster (e.g. as in / flo / or / sna /) than when they were not (e.g. / sa /, / san /, / fo /, / fol /). In other words, it is easier for children to recognise or identify a given phoneme when it is an onset on its own instead of when it appears as part of a cluster.

A related technique used is the “common phoneme” identification task. In an experiment involving this technique, children are trained to group or classify together different words or syllables on the basis of a common phoneme contained in some of them in the same position (either in initial, medial, or final position). They have thus to detect the “odd” word or syllable that does not contain the phoneme (e.g. Bradley & Bryant 1983; Calfee *et al.* 1972; Treiman & Baron 1981; Treiman & Breaux 1982). For example, when presented with nonsense syllables like [bi], [ve], [bo], the common phoneme is / b / and it is shared by the first and the third syllables and the odd one is [ve]. Subjects’ task is to group [bi] and [bo] together because they share the same phoneme (i.e. / p /) and, in doing so, they detect that [ve] does not contain that sound. Or, if subjects hear the real CVC words [fʌn], [bʌn], [gʌn], and [pin], the first three share the common vocalic phoneme / ʌ / in medial position but [pin], the “odd” one, does not.¹⁰

The auditory phoneme recognition and common phoneme identification tasks involve attending to the “constituent sounds of words” (Bradley & Bryant 1983: 419); however, some auditory categorisation tasks require focusing on the overall sound structure of the sound sequences involved. For instance, in a task similar to the “common phoneme” identification

¹⁰ Occasionally, children are asked whether two or more words sound the same at the end in a “rhyming” task that may involve the categorisation of words on the basis of more than one phoneme in the word-final position (e.g. Calfee *et al.* 1973).



one, children are required to group words or syllables, not on the basis of a common phoneme shared by some of them, but on the basis of their overall phonetic similarity. In a syllable triad like [bis], [diz], and [bun], for instance, the common phoneme pair is [bis]-[bun], the overall similarity pair is [bis]-[diz] and the unrelated, anomalous pair is [diz]-[bun]. Children in these tasks are trained to categorise words together on the basis of their phonetic similarity rather than their sharing a common phoneme. The general agreement is that preliterate children perform better in tasks in which they are asked to respond primarily on the basis of overall similarity relations among syllables (e.g. Treiman & Baron 1981; Treiman & Breaux 1982). In other words, children more easily group together two syllables or words that are phonetically similar overall than two sharing a common phoneme but are not otherwise similar.¹¹

Although the three versions of the identification task discussed above have been extensively used with children, discrimination tasks in either the “same-different” or “abx” versions have also been employed to study their ability to categorise auditory stimuli. In one of Wolf’s (1973) experiments, each stimulus was presented paired with itself and every other stimulus (up to seven) differing in 10 msec in VOT. Children had to respond “same” if the two sounds were perceived as the same and “not the same” if the two sounds were perceived to differ in any way at all. Also, the “abx” or “oddity” version of the discrimination paradigm has been employed on several occasions (e.g. Bowey & Francis 1991; Kirtley *et al.* 1989; Stanovich *et al.* 1984). Children, for example, are usually provided with word triads in which a given phoneme is the same in all words except for one -the “odd” one- in the onset (e.g. “man”, “mug”, “peck”) or in the coda (e.g. “pin”, “gun”, “hat”) and they have to spot the odd one out. These studies have found out that children usually treat syllables in terms of onset and rime units (see Treiman 1988, 1992 for reviews) and that phonemes are easier to recognise or detect when they occur at the onset position of a word than when the phoneme is part of the rime (Bowey & Francis 1991; Kirtley *et al.* 1989; Treiman 1988).

Apart from the identification and discrimination tasks, another common procedure used to study children’s categorisation of sounds is the controlled “spelling” task, in which children have to write the letter (e.g. Fink 1974) or point to one of a set of cards with a given letter printed on it and/or pronounce it (e.g. Treiman 1985b; 1993; Treiman *et al.* 1994). Other studies have examined children’s spelling in more naturalistic situations (e.g. Read 1971, 1975, 1986).¹² The spelling behaviour of children has revealed interesting phenomena regarding how children perceive and classify the sounds of the language they are learning to read and spell. For

¹¹ To a limited extent, the categorisation of auditory stimuli by adults has ALSO been investigated with techniques like the “common phoneme” identification task and “phonetic similarity” task (Treiman & Baron 1981; Treiman & Breaux 1982). Unlike children, adults usually respond primarily on the basis of common phonemes instead of phonetic similarity.

¹² Spelling behaviour is considered as an act of categorisation because it implies treating different sounds in the same way (i.e. attempting to write the same letter). In other words, spelling different sounds with the same letter implies that those sounds are somehow treated as equivalent and elicit the same behavioural response, in this case, choosing the same letter.

example, children's spellings of the oral stops in [s]-clusters usually correspond to the lenis members of / p /- / b /, / t /- / d /, and / k /- / g / phoneme pairs (Fink 1974; Treiman 1985b; 1993). In other words, children usually categorise the stops after [s] as / b /, / d / and / g / rather than / p /, / t /, or / k / as they spell those clusters as "sb", "sd", and "sg". In addition, children's spelling of voiced alveolar flaps in words in which the adult standard spelling is "t" or "tt", is usually with the letter "d" as those flaps are phonetically more similar to voiced alveolar stops than to voiceless alveolar stops.¹³ Children spell, for instance, "pretty" as "prede", "letter" as "ladr" or "water" as "woodr" (e.g. Read 1971, 1975, 1986; Treiman *et al.* 1994). These facts clearly indicate that children's spellings sometimes reflect perceived phonetic similarities between different speech sounds that belong in different phoneme categories according to adult standard spelling and classification. The general consensus is that the explanation for such behaviour is that children who have not yet been influenced by standard spelling respond differently than adults would because they are aware of the phonetic similarity between, for example, oral stops in [s]-clusters and the lenis stops (i.e. / b /, / d /, and / g /) in word-initial position and spelled with "b", "d", and "g" respectively or between voiced alveolar flaps and voiced alveolar stops typically spelled with "d".¹⁴

Finally, a few concept formation experiments have been conducted to address the psychological reality of the vowel shift rule in children (Moskowitz 1973; Wang & Derwing 1986). These studies basically obtained the same results as those carried out with adults.

2.1.3. The Categorisation of Sounds by Infants.

The study of the categorisation of sounds by infants is another aspect of developmental psychologists' increasing interest in understanding the initial stages of both cognitive development as well as speech perception and production.

Three different experimental techniques have been used to study infants' categorisation of the sounds of English. These techniques are actually different versions of a "discrimination learning" paradigm. One experimental technique is the head-turn response task. This technique allows one to infer categorical behaviour by examining infants' discriminative head-turning responses towards a loudspeaker whenever they hear different sounds. For example, infants typically sit on a parent's lap and watch different toys held by an assistant while listening to sounds repeated over a loudspeaker located, for example, at their left. They quickly learn to produce a head-turn response toward the loudspeaker when the sound changes from one phoneme (e.g. / a /) to another (e.g. / i /) because it signals the opportunity to see a particular toy

¹³ For a longer discussion of the classification and categorisation of voiced alveolar flaps see Mompeán-González in press.



(e.g. an animated dancing bear). Once they are trained to distinguish the two sounds, generalisation of the response to novel exemplars from both of the vowel categories is tested and the infant's head-turn responses are monitored. Thus, the procedure involves conditioning an infant to produce a head-turn response for a visual reinforcer when a change from one speech sound to another speech sound occurs. The technique has provided evidence that infants are able to categorise vowel categories like / a /, / ɔ /, / i /, or / ε / (e.g. Grieser & Kuhl 1983, 1989; Kuhl 1977, 1979, 1987, 1991) or consonantal categories like / s /, / ʃ /, / f / and / θ / (Kuhl 1980), / m /, / n / (Hillenbrand 1984), and / p / (Fodor *et al.* 1975).

The second technique, the visual habituation procedure and related novelty preference paradigm permits one to observe how infants relate the auditory and visual concomitants of speech. For instance, infants are shown two filmed faces, side by side, of a woman articulating two different vocalic sounds. One face displays productions of the vowel / a /, the other of the vowel / i /. A single sound, either / a / or / i /, is presented from a loudspeaker located midway between the two facial images. The two facial images then articulate the sounds moved in perfect synchrony with one another but infants systematically prefer to look at the face that "matches" the sound. In other words, infants look longer at the face that matches the vowel they hear. They recognise that / a /-sounds go with wide-open mouths and / i / sounds with retracted lips (e.g. Kuhl 1987; Kuhl & Meltzoff, 1982, 1984). With this paradigm Miller and Eimas (1996) studied infants' categorisation of CV syllables starting with velar stops ranging from / gi / to / ki / to a breathy exaggerated version of / k /.

Finally, the high amplitude sucking procedure (or sucking-habituation technique) is a discrimination learning task that examines infants' sucking behaviour in response to auditory stimuli. At the beginning of the experimental session, infants are habituated to a given speech stimulus as it is repeatedly played. Then, after habituation occurs, infants hear acoustically different sounds. If the sounds belong in a different category, the sucking rate increases while it remains almost invariable if they hear very similar sounds. This indicates that these sounds are perceived as instances of the category (or sound) to which they were initially habituated. With this experimental paradigm, the continua / ba – pa / (Eimas *et al.* 1971), / da – ta / (Eimas 1974; Miller & Eimas 1983) or the categories / a / and / i / (Kuhl & Miller 1975) have been investigated.

Many of the studies on infant auditory categorisation have discovered that, like adults, infants are sensitive to within-category differences and that some of the members of a given speech category are perceptually more salient than others. In fact, although initial studies (e.g. Kuhl 1980; Miller & Eimas 1983) were not able to determine whether infants could perceive

¹⁴ Spelling tests have also been used with adults. Fink (1974) and Treiman (1985b) also studied the categorisation of the stops in [s]-clusters. They found out that adult subjects almost always categorised these stops as / p /, / t / and / k / because, they argued, they may be influenced by standard spelling.

within-category differences (possibly because of methodological limitations at that time), subsequent research has shown that variations in within-category stimuli have differential effects in infant's categorisation behaviour in learning both prelinguistic vowel categories (e.g. Grieser & Kuhl 1983, 1989; Kuhl 1991) and consonant categories (e.g. Miller & Eimas 1996). Despite this, between-category distinctions are generally preferred.

2.1.4. The Categorisation of Sounds by Non-Human Species.

Studies of the categorisation of sounds by non-human species have been motivated by the growing interest in determining whether the perception of speech requires special species-specific mechanisms. The general consensus today is that, at least for mammalian species, this is not necessarily the case. Most of the findings of the studies mentioned below reveal that animals usually place category boundaries, in particular those on a VOT continuum, in much the same way as humans do, and that they are also able to distinguish within-category differences although they distinguish between-category differences much better, something that also applies to adults.

Many studies on the categorisation of sounds by non-human species has used some version of the so-called "discrimination learning" paradigm similar to the one used with visual categories. There are four main types of animals that have been studied: insects, auras, birds and primates. It should be mentioned here that, although we are mainly concerned with the categorisation of the sounds of English (whether naturally-occurring or speech synthesised), the categorisation of species-specific vocalisations will also be mentioned below as this type of stimuli is also essential to obtain a broader understanding of the auditory categorisation abilities of non-human species.

A few studies have been conducted with insects. For example, field crickets have been shown to distinguish attractive from unattractive synthesised calling songs. These insects track faster on a treadmill and orient towards the source of an attractive calling song but they do not react in the same manner when they hear an unattractive one (Thorson *et al.* 1982; Weber *et al.* 1981). In addition, flying crickets orient themselves toward synthesised stimuli resembling attractive calling songs but away from bursts with carrier frequencies above 15 kHz to avoid echo-locating bats.

The sound categorisation abilities of auras (frogs and toads) have also received some attention. Gerhardt (1978) showed how female green treefrogs could discriminate the mating calls of males, which attract them during the breeding season, from non-mating calls. When female green treefrogs heard a mating call, they oriented themselves and moved towards the source of the mating call while they did not behave in a similar fashion when they heard a stimuli not considered a mating call. Also, Capranica (1966) showed how bullfrogs could

discriminate synthetic mating calls that elicit choruses of calling frogs from stimuli that do not evoke such choruses.

Other categorisation studies have looked at the abilities of avian species to discriminate sound categories. Beletsky and his co-workers (1980) showed how red-winged blackbirds could discriminate synthesised stimuli resembling species-specific songs that produce aggressive behaviour in territorial males from songs that do not produce aggressive behaviour. The measure of categorisation performance was the aggressive behaviour of the bird, shown when it heard a stimulus that elicited such behaviour and unattested when the stimulus did not elicit the behaviour. Also, Gottlieb (1980) demonstrated mallard-duck embryos' ability to discriminate synthesised maternal calls from non-maternal calls by their reaction (or absence of it) to the stimulus presented.

When studies are conducted with avian species, birds are typically trained to peck for reinforcement (e.g. mixed grain) whenever they hear positive instances of a particular category and not to do so whenever the sounds heard do not exemplify the category. Using this paradigm, the continuum [b-d-g] has been investigated with Japanese quail (Kluender *et al.* 1987) and the continuum [i-r] with European starlings (Kluender *et al.* 1998).

Finally, many studies of the categorisation of sounds by non-human animals have been conducted with mammalian species like primates and rodents.

The measures of performance in the discrimination learning tasks carried out by primates are usually their pressing a button/lever, lifting of a telegraph key or touching a loudspeaker for reinforcement (e.g. sucrose, a banana pellet, orange drink) when positive instances of a category are heard and not to do so when auditory stimuli do not exemplify the category. Rhesus monkeys have been shown to learn categories that range from relatively simple ones like / i / (Kuhl 1991) to more complex ones like auditory same-different concepts in which the paired sounds include different environmental or natural sounds like telephone rings, door buzzers, car engines, cuckoo clocks, whistles, etc. (e.g. Shyan *et al.* 1987; Wright *et al.* 1990).

Another measure of performance in such studies is heart-rate. Morse and Snowden (1975) examined the discrimination of stimuli from a synthetic [ba-da-ga] VOT continuum. They found that the heart rate of rhesus monkeys habituated to a given stimulus previously presented in a repeated way increased more when monkeys were heard stimuli from a different category than stimuli from the same category. A further version of the discrimination learning paradigm employs antiphonal vs. non-antiphonal responses as measures of categorisation performance. Using species-specific synthesised vocalisations rather than the usual synthetic stimuli resembling human phonetic categories, Masaka (1983) studied the categorisation abilities of Goeldi's monkeys. These New World macaques have forms of alarm that elicit freezing responses and others that elicit antiphonal alarm calls (i.e. mimicking calls). Masaka found that the macaques could learn to discriminate between different types of alarms observing whether a

given stimulus elicited an antiphonal or freezing response. A similar study was conducted with the pygmy marmoset, another small New World monkey with an elaborate vocal repertoire including close-mouth trills, given as a contact call in quiet, undisturbed conditions and open-mouth trills, associated with some form of aggressive or fearful behaviour. Pola and Snowdown (1978) showed that these macaques could learn to discriminate between these two types of synthesised natural trills based on the antiphonal calls produced after hearing each stimulus. Finally, in still another version of the discrimination learning paradigm known as “avoidance-conditioning”, Waters and Wilson (1976) trained rhesus monkeys to discriminate the two endpoint synthesised stimuli on two VOT labial and velar continua ranging from -140 msec VOT to $+140$ msec VOT. Rhesus monkeys were trained to cross a midline barrier in a shuttle box to avoid a shock when the positive stimulus ($+140$ VOT) was presented, and to inhibit the avoidance response to avoid footshock when the negative stimulus (-140 VOT) was presented. The correct inhibition was rewarded with food. The macaques had then to generalise to test stimuli between the two endpoints.

This version of the discrimination learning paradigm was originally used to study the categorisation abilities of mammals other than monkeys.¹⁵ In several studies with chinchillas, these rodents were also trained to cross a midline barrier in a cage to avoid a mild shock and the sounding of a buzzer when they heard a given endpoint stimulus on a VOT continuum ranging from, say, 0 msec VOT to $+80$ msec (e.g. 0 msec VOT) and to inhibit the crossing response when they heard the other endpoint of the continuum (e.g. $+80$ msec VOT). Correct inhibition of the response was then rewarded by the presentation of water from a drinking tube. When performance on these endpoint stimuli was near perfect, a generalisation paradigm was used to test the stimuli between the two endpoint stimuli on the continuum in, for example, 10 -msec steps (i.e. $+10$ msec VOT, $+20$ msec VOT, $+30$ msec VOT, etc.). Using this paradigm, chinchillas’ ability to discriminate computer-synthesised version of the vowels / a / and / i / (Burdick & Miller 1975), the syllables / da / and / ta / (Kuhl 1981; Kuhl & Miller 1975a), or / ba / - / pa / and / ka / - / ga / (Kuhl & Miller 1978) was investigated.

Chinchillas are not the only rodents whose categorisation abilities have been investigated. In still another version of the discrimination learning paradigm, including orientation and movement towards the source of the sound as a measure of performance, Ehret and Haack (1982) studied mouse-gup ultrasonic vocalisations by lactating females. When house mouse pups are in discomfort, they emit ultrasonic whistles that elicit searching and retrieving behaviour by the mother. Ehret and Haack tested mothers by playing sounds through two

¹⁵ To a very limited extent, a “classical conditioning” paradigm, a very similar technique, has been used to study the categorisation of sounds by English-speaking subjects (e.g. Jaeger 1980a: exp. 1; 1980b: exp. 1), etc. Jaeger investigated the question of whether subjects could become conditioned to [k] sounds. To do so, subjects were trained (via a mild shock) to produce a change in galvanic skin response upon hearing words with aspirated “k”. Words that did not contain the [k] were not paired with the mild shock while those with [k] sounds were.

loudspeakers and observing which the mother would orient herself towards and approach. These researchers used a variety of ultrasonic synthetic noise bands ranging from 17-5 to 50 kHz in bandwidth and 20-kHz pure tones and discovered that adult mice could distinguish between synthetic instances of maternal calls and pure tones.

To conclude this overview of the research conducted on the categorisation of sounds we should like to draw the reader's attention, as we did in section 2.1., to the diversity of the experimental techniques that have been used in order to examine the abilities of different subject populations to categorise auditory stimuli. All these techniques, summarised for convenience in table 1, have contributed to a better understanding of and more confidently held beliefs about the categorising abilities of those populations in general, and of auditory stimuli in particular.

Table 1. Categorisation Techniques Used with Different Subject Populations.	
<i>Adults</i>	<i>Children</i>
➤ Phoneme-monitoring ➤ Absolute identification ➤ Discrimination - "same-different" task - "abx" or "oddy" task ➤ Concept formation ➤ Spelling task ➤ Phonetic similarity Identification/recognition ➤ Common phoneme identification/recognition ➤ Classical conditioning	➤ Auditory phoneme recognition ➤ Common phoneme identification/recognition ➤ Phonetic similarity identification/recognition ➤ Concept formation ➤ Spelling tasks ➤ Absolute identification ➤ Discrimination - "same-different" task - "abx" or "oddy" task
<i>Infants</i>	<i>Non-Human Species</i>
➤ Discrimination Learning - Head-turn response - Visual habituation (auditory version) - Sucking-habituation	➤ Discrimination Learning - Pecking - Pressing/pushing - Orienting & moving towards the source of the sound - Heart rate - Antiphonal response - Avoidance-conditioning

2.2. The Categorisation of the Sounds of English and its Relation to the Field of Phonology.

What might categorisation and, more specifically, the categorisation of sounds have to do with phonology?.

A possible answer to this question is that, given the importance of categorisation in perception, cognition, memory, behaviour, and language (see sections 1.1.1. through 1.1.5.), the experimental study of the categorisation abilities of language learners/users might help phonologists in their goal of developing an explanatory theory of language and to arrive at a better understanding of the sources of its structural regularities. As the structure of natural languages seems to be determined to a great extent by both the characteristics of the human vocal and auditory/perceptual apparatus as well as by general cognitive factors (Nathan 1986, 1994; Pierrehumbert *et al.* 2000), categorisation, which is an important cognitive ability, might be responsible for many of the structural phenomena phonologists have usually been concerned with. The fact, for example, that different allophones or contextually-determined and acoustically different realisations of a given ‘sound’ are treated or accepted as ‘the same’ sound may be viewed as a question of categorisation. Not surprisingly, Lotto (2000: 189) claims that “purported units of speech, e.g. phonemes or features, are essentially categories” and that “the assignment of phonemic (or phonetic) identity is a process of categorisation”. In this respect, Wheeler claims that the task of a (psychological) phonology is “to discover what principles and mechanisms speakers actually use in constructing grammars from language data” (1980: 65). As much is already known about the cognitive mechanism of categorisation, its study might also aid in developing a better understanding of the structure of language and more specifically, of its sound pattern.

Another possible answer is that the data obtained from the study of the categorisation of sounds by actual speakers of the language under investigation (English in this case) might provide useful evidence in resolving controversial issues in phonological theory. For example, the phonological assignment (and categorical status) of the voiceless oral stops in [s]-clusters (e.g. “speak”, “steal”, “skin”) is traditionally a controversial issue in phonology. Some phonologists assume that those stops should be classified as / p /, / t / or / k / while others assign them to their lenis counterparts (i.e. / b /, / d / or / g /). Still, other phonologists classify them as members of other categories called archiphonemes that could be represented as / P /, / T /, or / K /.¹⁶ Be that as it may, any of these otherwise different claims represents an underlying or implicit “phoneme” theory about what units count as being “the same” and so, the solution to this phonological problem may be considered to be a question of classification or categorisation on which behavioural data obtained from the speakers of the language might shed some light.

However, taking the categorisation of sounds by actual speakers as valuable evidence to solve phonological theoretical controversies is a major assumption that is by no means shared by most phonologists and explicitly rejected by some (e.g. Bloch & Trager 1942: 40). In principle, the use of behavioural evidence obtained from actual speakers of the language is not

¹⁶ See experiment 3 for a more detailed discussion of this phonological problem.

very common in many phonological theories. Such evidence is usually disregarded or considered irrelevant to the object of study, namely the sound pattern of the language. In addition, the gathering of such evidence through experimentation is still even less usual. Finally, having recourse to actual speakers of a language to obtain evidence in order to decide which phonological analysis out of a set of different possible descriptions is more appropriate seems to imply that phonological descriptions or analyses that are theoretically valid are those that have some sort of psychological or psycholinguistic experimental support. However, the assumption that one of the aims of phonology should be to provide psychologically valid descriptions has neither been the dominant view in the field of phonology nor the most successful one. In the following sections (2.2.1. through 2.2.2.3.). I briefly outline two major views of what the task of phonology should be (descriptive and psycholinguistic) and the reasons why the second has not been sufficiently successful, the benefits that experimentation might bring to a psycholinguistic phonology, and the complementariness of the two views.

2.2.1. Descriptive Phonology vs. Psycholinguistic (Psychological) Phonology.

The question of the relationship between the patterns described by phonologists and their psychological adequacy has been a controversial topic for decades. Apparently, there are two major approaches to the issue. The first approach is that the relationship between descriptive patterns, on the one hand, and their psychological validity, on the other, is quite indirect if it is even a relevant point of discussion. According to the phonologists who endorse this view, the goal of phonological investigation is to discover, capture, and systematise all detectable phonological structures, patterns, alternations and generalisations found in language data. This view goes back to the earliest days of structuralism when phonology was regarded as a classificatory science in which any sort of mentalism was explicitly rejected (e.g. Hockett 1942; Trubetzkoy 1939). The comparative method, a rigorous way of demonstrating relations between units of language and suitable for the task of describing and cataloguing the units and patterns in language, was its methodological hallmark (Ohala 1983: 232).

However, phonologists since Chomsky (e.g. Chomsky 1964; Chomsky & Halle 1968) started to propagate the idea that phonological descriptions should go beyond the bounds of an arbitrary taxonomic or classificatory system (e.g. Campbell 1979; Derwing 1979; Fromkin 1980; Skousen 1975; Pierrehumbert 2000). These phonologists have been claiming that phonological analyses should not be mere descriptions of linguistic data. According to them, phonological analyses should represent speaker's actual underlying knowledge of the language, referred to as "competence", and thus be "descriptively adequate", that is, capable of being valid both as descriptions of the language data on which they were based, and most importantly, as reflections of the knowledge speakers have of their language. The goal of phonology should

then be to determine what speakers actually know about the sound pattern of their language and strive to represent this in phonological descriptions instead of aiming at producing consistent and thorough, though psychologically fictitious, descriptions of observable data on the language. Therefore, any phonological description pursuing that goal is also viewed as trying to capture “psychological reality” or “psychological validity” and descriptions that comply with this requirement are consequently characterised as being psychologically real, valid, adequate, etc.

Most phonologists endorsing this second view agree that the psychological reality that should be captured by phonological analyses refers to speakers’ declarative and procedural knowledge about the phonological patterns of their language.¹⁷ However, there are two different senses in which the construct of “psychological reality” can be understood (Cutler 1979; Eddington 1996, 1999). A phonological description is psychologically real in a “strong” sense when a theoretical analysis closely corresponds to or mirrors speakers’ actual mental representations and processes. In other words, the strong sense claims that the formal notations with which phonological descriptions are formulated are isomorphic to cognitive processes. On the contrary, a phonological description that is psychologically valid in a “weak” sense implies little correspondence between theoretical constructs and psychological processing or storage. What is potentially real in a phonological description that is psychologically valid in the weak sense is the function, contents, substance or general idea embodied in or captured by the analysis, not the actual “form” of the analysis.¹⁸

2.2.2. Success of the Different Approaches to the Task of Phonology.

In principle, these two views of the goal of phonological theory (i.e. the merely descriptive and the psychological or psycholinguistic) are legitimate approaches to phonological research. Defining one’s aims, methods, and object of study, a phonological investigation can either strive to search for and describe structural patterns and generalisations or to determine whether a given phonological construct or analysis has psychological validity for speakers. However, the success of these two different approaches to the study of phonology has been very different. Descriptive (classificatory) phonology appears to have had a tremendous success in

¹⁷ According to Ryle’s original distinction (1949), declarative knowledge implies the subjects’ ability to reproduce information propositionally whereas procedural knowledge is represented by means of productions and is difficult to verbalise. Once learned, procedural knowledge operates in a rapid, automatic and unconscious way. For example, speakers who fully control the phonetics of their language know when, where, and how to pronounce all of the different realisations of a phoneme category although they may not be able to reproduce this information propositionally.

¹⁸ Phonological rules, for example, are often spoken of as if they were algorithmic operations or mental processes. If this were the case, they would be considered as psychologically real in the strong sense and the step-by-step rule orderings or intermediate derivations of the phonological descriptions of phonological rules, for example, would be considered to have some correlate in the actual mental processes involved. However, phonological rules are usually defined in terms of speakers’ intuitions or tacit knowledge and not in terms of actual mental processes, which makes them candidates for being psychologically valid in the weak sense.



discovering the phonological patterns and structures in many languages like English. Any random inspection of the phonological literature will provide us with the extensive amount of phonological patterns, generalisations, restrictions, etc., that descriptive phonologists have found in language data. However, the success of the psychological approach to phonology is more dubious. Despite its exciting scientific programme, psychologically-oriented phonologists have made little progress in producing phonological descriptions that can legitimately claim to represent the actual phonological knowledge of the speakers of the language under investigation. This failure is the result of a series of gratuitous, unwarranted assumptions, as well as many unfortunate methodological misunderstandings and practices that cast serious doubts on the psychological validity of the phonological analyses that claim to capture psychological reality in some way (e.g. Derwing 1979; Eddington 1996, 1999; Fromkin 1980; McCawley 1986; Wheeler 1980). The result has been that progress in a truly psychological, psycholinguistic, mentalistic or cognitive phonology (call it what you will) has been hampered so seriously that it has been suggested that the metatheoretical concept of “psychological reality” should be completely abandoned as it appears to have been irrelevant to the practice of phonology, let alone linguistics (Black & Chiat 1981). In the rest of this section, six different but related reasons will be mentioned that explain why the psychological approach to phonology has been mostly a barren field of research.

The first reason to doubt the psychological relevance of phonological descriptions could be called the “awareness problem” (Derwing 1979; Eddington 1996; Wheeler 1980: 53, 69). This means that despite claiming to pursue psychologically valid phonological descriptions, many phonologists endorsing any of the more or less explicitly mentalistic versions of generative phonology have unconsciously continued doing phonology in a completely non-psychological way. In other words, many phonologists naively believe that their analyses reflect psychological reality but have failed to become aware, realise or comprehend that assessing the psychological validity of phonological descriptions requires methods and procedures different from the traditional structuralist ones. This lack of awareness is probably due to the influence of the comparative method used by the structuralists, a method that has made many well-meaning psychologically-oriented phonologists continue to pursue research like non-mentalistic describers of language data and practice phonology as it had been standardly done by the structuralists.¹⁹ In other words, there has been a generally unconscious inheritance of earlier views from pre-structuralist linguistics of what constitutes a phonologists’ field of enquiry and approach to research. This has made supposedly mentalistic phonologies continue to have nothing to do in the way of psychological research and the methods of the experimental cognitive psychologist while believing it does. As McCawley adverts (1986: 28, 37), simply by



taking phonology to be a branch of cognitive psychology as some psychologically-oriented phonologists do (in particular generative ones), phonological data and the methodology traditionally used do not thereby turn into psychological data and procedures despite the use of psychological jargon.

The second reason for doubting the psychological reality of phonological analyses is the “descriptive validity=psychological reality” fallacy (Eddington 1996). Many phonologists assume that, out of a number of possible descriptions for the same phonological data, the best formal analysis of the phenomenon under investigation is necessarily a psychologically significant analysis. The “best” formal analysis is usually the most general, simple, detailed, rigorous, sophisticated or economical description of a phonological phenomenon. However, this assumption simply corresponds to a generally accepted scientific principle that is applied in describing or theorising about any body of data, namely, that the simplest description consistent with the data is to be preferred (Campbell 1986; Wheeler 1980: 55). This was already adopted by strongly anti-mentalist phonologists (e.g. Hockett 1942; see also Wheeler 1980: 56 for more details) but says very little about the psychological reality of phonological analyses. However, one thing is to arrive at a general or elegant phonological analysis and another to claim that such an analysis has psychological validity because it is formally impeccable. As Black and Chiat note (1981: 42) it is a common, but mistaken, practice to refer to an analysis as “psychologically valid” when what is actually meant is that the analysis is “linguistically valid” with no psychological substance intended. In addition, a phonological analysis may be “valid” in the context of a certain theoretical phonological paradigm or “framework” but this does not mean that the analysis and the theoretical assumptions premises of the framework are psychologically real. However, a description of a body of data in the most economical and general way does not necessarily correspond to how humans ordinarily account for the same set of data. There is a growing body of experimental evidence in the cognitive sciences that supports this assumption (see for example Wheeler 1980: 54-74). The experimental findings obtained on taxonomic levels against the classical view (see 1.2.1.3.5.) testify to the fact that elegance, simplicity or generalisation does not always correspond to the way in which human knowledge is organised. However, there is an increasing body of experimental phonological literature that also shows that the argument that speakers always prefer the “simplest” solution is inaccurate (Jaeger 1986b).

The third reason is the “the-discovery-of-a-pattern-implies-its-psychological-reality” fallacy. There is an implicit assumption made by many psychologically-oriented phonologists that the discovery of the existence of phonological patterns and generalisations constitutes proof that speakers have some sort of declarative or procedural knowledge of them (Eddington 1996,

¹⁹ The comparative method, when applied to a single language yielded the method of *internal reconstruction* and this technique is the basis for the structuralist phonologist’s criterion of complementary distribution used in finding phonemes and the

1999; McCawley 1986; Wheeler 1980). However, the fact that an alternation, constraint, or structure can be found and described utilising formal methodology is not grounds for supposing that it is somehow represented in native speakers' minds, nor that they have any sort of conscious or tacit knowledge of it, or account for these structures in the same way as phonologists do. There is usually insufficient justification that phonological analyses are more than mere descriptions but actually reflect some sort of psychological knowledge (other than the linguist's) that can be generalised to a given speech community. A phonological generalisation may, for example, have arisen by chance, or be the result of diachronic phonological or morphological changes no longer in use and without any synchronic psychological relevance (Eddington 1999: 15). Also, certain alternations may be due to purely articulatory or aerodynamic influences. As the data from which generalisations or structural patterns are obtained have been produced by humans, it is possible that humans have knowledge of, or utilise them (Derwing 1979: 116). However, their existence merely shows that they are available to be potentially known or used.

In fact, a great deal of psycholinguistic evidence is accumulating that undercuts the assumption that a language user knows everything the linguist knows. Speakers will account for some of these regularities, but not all (Ohala 1990).²⁰ Believing the opposite (i.e. that speakers know everything the linguists does) is precisely one of the most widespread pitfalls of psychologically-oriented phonologists that can be called "projection". As Ohala says (1990), phonologists typically project the extensive knowledge they have about the many different phonological phenomena of a language (including historical derivational relationships between words, the morphemic structure of complex words, knowledge of cross-linguistic sound patterns, etc.) onto the mental phonological system of linguistically-naïve native speakers.

The fourth reason to question the psychological validity of phonological analyses has to do with the fact that many phonologists attempt to produce phonological analyses that claim to represent speaker's phonological knowledge, but they carry out such analyses in almost complete isolation from those speakers (Derwing 1980; Eddington 1996, 1999). Consequently, many critics wonder how phonological descriptions can profess to be psychologically real if they are arrived at with minimal or no recourse to actual speakers and when the latter are consulted, data are usually based on very few speakers (e.g. one or two subjects). Part of the reasons that psychologically-oriented phonologists' fail to fully incorporate data from actual

generative phonologists' positing of common "underlying" phonological forms for related morphemes (Ohala 1983: 232).

²⁰ A clear example of this is the psychological validity of the vowel-shift rules of the original Chomsky & Halle analysis (Chomsky & Halle 1968). A substantial number of psycholinguistic studies including oral production tests (e.g. Ambruster 1978; Eddington 2001; Ohala 1974, Steinberg & Krohn 1975; Myerson 1976; Wang 1985), preference tests (e.g. Ambruster 1978; Myerson 1976, Wang 1985), recall tests (e.g. Cena 1978; Myerson 1976), and concept formation tasks (Jaeger 1980a, 1984, 1986b; Moskowitz 1973; Wang & Derwing 1986) have shown that only five of the alternations resulting from the Great Vowel Shift as described by Chomsky and Halle (1968) are psychologically significant for contemporary English speakers, that is to say, these alternations, in contrast to others, play a significant role in experiments involving concept formation, learning, memory, and pronunciation preference. These are precisely the alternations with the most frequent number of related pairs (Cena 1978; Ohala & Ohala 1987; Pierrehumbert *et al.* 2000).



speakers of the language is that they still frequently view language as a kind of isolated “natural object” which should be investigated independently of the psychology of its actual speakers and hearers, in much the same way as the structuralists did (Derwing 1979, 1980). In fact structuralist phonologists explicitly rejected having recourse to speakers as a mean to validate their analyses (e.g. Bloch & Trager 1942: 40). Instead, phonologists’ descriptions apparently characterise the phonological knowledge of some virtual or ideal speaker that more closely resembles their own expert knowledge than that of actual speakers’. Consequently, critics claim that in order to determine what is actually known or utilised by the speakers, the focus of the research must turn back towards the speakers themselves.

The fifth reason for questioning the psychological validity of many phonological analyses is that the evidence on which they are based is not only limited and narrow but also ambiguous and inconclusive as far as validating psychological claims is concerned (Eddington 1996, 1999). Such evidence has usually gone under the heading of “internal evidence” or “formal evidence”, and it consists of data that phonologists glean from a corpus of utterances, pronouncing dictionaries, the examination and transcription of carefully monitored speech, etc. The facts from these sources include synchronic generalisations or patterns based on the surface regularities, phonetic, phonemic or morphophonemic alternations found in the language data, and their distributions in the linguistic system, as well as what structures or elements are not found in a particular context (e.g. segmental and sequential constraints, restrictions which exist in the distribution of the phonological elements). In addition, internal evidence includes facts about “typological predominance” or “expectedness” (e.g. certain phonological situations are “more expected” typologically across language and certain types of diachronic change are more expected than others).

As mentioned above, and despite sporadic arguments to the contrary (e.g. Chomsky 1980: 12), the general consensus today is that internal evidence cannot satisfactorily answer questions about the psychological validity of phonological descriptions. Consequently many critics call for other non-traditional sources of evidence typically overlooked by phonologists that fall under the heading of “external evidence” or “substantive evidence” (Bertinetto 1992; Campbell 1979, 1986; Cutler 1979; Davidsen-Nielsen 1975; Derwing 1979; Eddington 1996, 1999; Fromkin 1980; Jaeger 1980a; Lass 1984: 214-216; McCawley 1986; Nearey 1981; Ohala 1986; Pierrehumbert *et al.* 2000; Skousen 1975; Wheeler 1980). These critics claim that external evidence is not some sort of supplementary information to verify the psychological validity of phonological descriptions, but that it is primary and more valid or revealing than internal evidence in supporting or validating claims about the psychological reality of theoretical descriptions and constructs.

External evidence includes data obtained from speech errors, word games, language acquisition, speech defects, etc.²¹ Unlike internal evidence, which involves language used in unexceptional ways such as careful, monitored speech, external evidence is gathered from actual language use, especially in unusual and exceptional ways and situations and is either observed in natural, everyday speech production or perception or induced under experimental conditions. External evidence is more highly valued than internal evidence because it is assumed that if a given phonological construct has some psychological relevance it should have some kind of observable, perceivable behavioural consequences in speakers' actual use and manipulation of the sound pattern of their language that may be observed either in "naturalistic" or in experimentally-induced conditions.

Perhaps, the wisest stance to take on the issue of the (admittedly artificial)²² dichotomy "internal" vs. "external" evidence is that all evidence is good and useful in order to extract information about speakers' knowledge of the sound pattern of their language, but that some types of evidence are simply more useful or telling of cognitive processing and products than others. In other words, it appears that there is a continuum in the quality of evidence (Eddington 1996; Ohala 1986). Internal evidence, for example, is relevant to the search for psychological reality in that it determines what structures exist in the language, and are thus available to be potentially internalised by the speakers of the language. It is the first step in arriving at psychologically valid analyses because it dictates which observed patterns, of the infinite observations that could be made, are made. Internal evidence is, however, merely suggestive and rather ambiguous to interpret in mentalistic terms, and rarely conclusive. On the contrary, external evidence, in particular that which is induced experimentally (Eddington 1996; Ohala 1986), seems to have a greater "resolving" power to decide which phonological descriptions are significant for speakers because, as mentioned above, it shows speakers behaving in ways where they must call upon their knowledge of the sound patterns of their language in overt and revealing ways. It appears then that phonological descriptions that claim to be psychologically significant must not be founded on internal evidence alone, but must crucially include external evidence as well.

Finally, one of the most serious reasons to doubt the psychological validity of phonological analyses is their unfalsifiable character. One of the distinguishing features of

²¹ Many sources of external evidence have been cited over the years. These include speech errors, word games, first language acquisition, speech defects or disorders like aphasia, problems in second language acquisition, metrics and verse, construction of orthographies and spelling errors, observations on how speakers adapt borrowed words into their languages, speakers' verbal reports or judgements on a given phenomenon, etc. Any of these types of evidence can be obtained under natural conditions in which the spontaneous linguistic behaviour of subjects is observed or they can be gathered under experimental conditions with all sorts of different perception and production experiments. For example, several studies deal with spontaneously observed slips of the tongue or speech errors (Cutler 1979; Fromkin 1973; McKay 1972; Stemberger 1983) whereas other studies create the conditions to elicit such errors (e.g. Davidsen-Nielsen 1975; Stemberger 1985; Wells-Jensen 2000). The same applies to word games, which have been studied in naturalistic conditions (e.g. Campbell 1979) but also experimentally (e.g. Campbell 1986; Hombert 1986; Treiman 1983, 1984, 1986).

²² In fact the term "external" evidence is unsatisfactory and such evidence is in fact external from the standpoint of a theory which ignores it (Derwing 1979).

empirical sciences in opposition to non-empirical ones is their falsifiability. In this respect, it would be desirable that any claims about which phonological analyses have psychological validity could be verified or refuted empirically. Otherwise, phonological descriptions that never go beyond the theoretical stage can never guarantee that they actually represent what is psychologically valid for speakers. The testability of phonological theories is therefore essential (Bertinetto 1992; Derwing 1979; Derwing & Nearey 1986). In this respect, it appears that experimentation is the main means of hypothesis testing in phonology as in any other discipline. Experiments provide an ideal opportunity to show whether a particular phonological analysis is right or wrong from a psychological standpoint. Unfortunately, purportedly psychological phonological descriptions make many assumptions, speculations and conjectures about internalised phonological systems, but usually fail to state them as explicit hypotheses that could be tested empirically. Even the phonologists who are most concerned with psychological reality generally fail to be sufficiently experimental in their approach. If a psychological phonology is to be pursued, phonologists should not be exclusively occupied analysing language data except insofar as such analyses are made for the express purpose of providing concrete and explicit hypotheses about speakers' internalised phonological systems that are subject to being empirically confirmed or refuted.

Unfortunately, much of what has traditionally been done in phonology and is currently written in contemporary approaches to phonology is the product of nonexperimental research that is never empirically tested (Eddington 1996, 1999; Wheeler 1980). This means that it is not derived from data obtained from real subjects' performance in contrived and controlled tasks that examine the phenomena under investigation. Instead, most supposedly mentalistic analyses are the product of theoretical assumptions and logical reasoning about the consequences of a theoretical framework in which a given analyst works.²³ This makes phonology a discipline more similar to other nonempirical ones such as formal logic, pure mathematics, and philosophy than to empirical sciences like cognitive psychology, psycholinguistics, etc. Phonologists' unacceptance of standard scientific experimental methods for validating theories isolates them from the scientific mainstream (Ohala & Jaeger 1986b: 6). The problem with the psychologically-oriented phonologists who continue to regard phonology as an autonomous discipline in the sense that its methods are different from other empirical sciences, which is, by the way, the general trend in linguistics (Ohala 1983: 232). Recalling the quotation with which the present dissertation started:

Phonology can benefit by embracing the data and methods from fields as diverse as acoustics, psychology, and ethology. Phonology does not lose its identity by this; what sets phonology off from other disciplines is its *questions*, its *ends*, not its *methods* or *means*.

²³ The recent history of phonological theory has been marked by the invention of many frameworks, such as Lexical Phonology, Declarative Phonology, Government Phonology, or Optimality Theory. Frameworks are packages of theoretical assumptions about the fundamental nature of language.

2.2.2.1. Criticisms about the Use of Experiments in Phonology.

If the use of carefully designed and conducted experiments appears to be an appropriate method to validate or invalidate claims about the psychological pertinence of phonological analyses, it is paradoxical that most psychologically-oriented phonologists have, in general, been either negative or at least indifferent to experiments (Mompeán-González 2001b).

There may be a host of reasons for the apparent reluctance of phonologists who claim that their analyses reflect psychological reality from conducting experiments. Apart from an insufficient exposure to the field of cognitive psychology, the most likely reason seems to be phonologists' unfamiliarity with the nature of experiments and how experimental research proceeds (Wheeler 1980: 66). For this reason, phonologists probably continue doing what they feel most comfortable with, namely, analysing language data (Wheeler 1980: 57).²⁴ In addition, the popular misconception that experimentation always involves complex instrumentation, procedures, and statistical analyses (Ohala 1986: 10; Ohala & Jaeger 1986 1986b: 2), may deter many phonologists from embarking upon experimental research. The lack of experiments is, however, puzzling given the fact that the majority of phonologists are academics who regularly use the equivalent of experiments as part of their teaching practice (Ohala 1986: 11). Phonologists, for example, make their students record words or listen to different sounds and identify them. Moreover, this reluctance and sometimes overt rejection has gone together with certain criticisms of experiments that can be, and have been easily contested (e.g. Derwing 1979; Eddington 1999; Nearey 1981; Ohala 1995; Ohala & Jaeger 1986b). Such criticisms usually reveal phonologists' insufficient acquaintance with the experimental method as it functions in the empirical sciences. Six common criticisms and replies to such criticisms will be discussed below.

First, some phonologists opposed to the use of experiments usually argue that experiments often produce conflicting results about a given phenomenon or produce no evidence supporting the existence of a well-described, apparently real phenomenon. On the one hand, the argument is usually made that phonological theory cannot depend on (external) experimental evidence in order to validate or test phonological hypotheses because it is often the case that experimental results are contradictory. Ironically, phonologists who criticise this usually take into account external evidence when it supports their theories but reject or ignore it

²⁴ A typical example can be found in Kenstowicz and Kisseberth (1977). At the beginning of their book on generative phonology, they claim that external evidence is needed in order to find out what kind of phonological analyses have psychological validity for speakers and that internal data cannot fully provide a satisfactory account. However, they argue that they are "forced" to use the latter because adequate external evidence is lacking. Therefore, they write an entire book based on a type of evidence they acknowledge is insufficient instead of contributing to obtain valuable external evidence.

when it does not (Bertinetto 1992; Eddington 1996; Nearey 1981). However, conflicting results occur in *all* fields of science which incorporate experiments (Campbell 1986: 172; Eddington 1999: 20). Contradictory results are the result of wrong interpretations, poor methodological practice, or the presence of unanticipated confounding variables (Campbell 1986; Derwing 1979). The existence of such factors should merely make researchers look into the phenomena under investigation in more detail and refine their experimental methods and procedures. For example, conflicting results may be obtained in two different experiments because an unpredictable uncontrolled variable or a methodological flaw made the results of experiment B contradict the findings of experiment A. If this is the case, a better designed experiment that neutralises the uncontrolled variable suspected of causing interference and overcomes the technical or methodological flaw can be conducted. Such an experiment may then confirm the results of either experiment A or those of B. In short, the answer to an experiment suspected of being somehow flawed and producing conflicting results with previously acquired evidence is conducting a better-controlled experiment (Ohala 1995: 721; Ohala 1986: 10; Ohala & Jaeger 1986b: 5). Bad practice does not mean that the entire enterprise is misguided (Campbell 1986: 172). On the other hand, it may be the case that successive experiments fail to provide any evidence for the existence of a frequently described phonological phenomenon. Critics of experiments could argue that experiments may fail to provide evidence for something that is out “there” in the data (e.g. Gussman, cited by Fromkin 1980; Kiparsky 1975). However, if there is abundant experimental evidence that a phonological pattern, generalisation, etc., does not play a part in storage, recognition, or production, there should be no problem in denying it any psychological significance no matter how time-honoured, traditional, or frequent the construct is in the specialised literature (Eddington 1999; Mohanan 1986).

Second, it has been claimed that the unusual circumstances of the experimental situation (e.g. being in a laboratory, using a given apparatus, the feeling of being examined, etc.) may cause subjects to give unusual responses that they would not under normal circumstances. In other words, a particular experimental technique may yield data which are caused by the technique, rather than reflecting the experimental subjects’ control of the phenomenon, which is of real theoretical interest (Derwing 1979: 125; Ohala 1986: 10). This would cast doubt on the validity of the results outside of the experimental setting. In other words, the “external validity” of the results obtained (i.e. the extent to which they are applicable to entities or people other than those tested in the experiment) would be questionable. Although experimental artefacts may crop up in any experiment, the advantage of experimental artefacts over other “accidents of observation” is that they are experiment-specific and they can be controlled for in two ways. On the one hand, the same phenomenon can be tested again with a better-controlled or refined version of the experimental paradigm that avoids the methodological or technical flaw detected



in the previous experiment.²⁵ In addition, a fruitful approach in experimentation is to use different experimental techniques to cross-validate results and rule out experiment-specific artefacts. If several different experiments yield similar results, the latter are more likely to be due to actual mental processes, and not to something inherent in the experiment (Derwing 1979, 126-127; Derwing & Nearey 1986: 190-191; Eddington 1999; Nearey 1981). On the other hand, the results can be compared with external evidence obtained in more natural or ecological circumstances. If experimental results are in accordance with conclusions drawn from observations of naturally-occurring phenomena believed to show the psychological validity of a given phonological construct, then the validity of the experimental results are more strongly supported.

Third, a common objection to the use of experiments in phonology is that they provide observable *performance* data obtained in experimental conditions, but one cannot be really sure whether subjects' performance is a reflection of speakers' underlying *competence*. In this respect, generative phonologists have been disturbingly reluctant to deal with experimental evidence due to a fear that all experiments are fraught with performance effects that will totally obscure the actual phonological competence (Eddington 1999; Nearey 1981). However, in the empirical sciences, experimental behavioural data obtained by means of carefully-controlled experiments, the results of which are replicated experiment after experiment, can be safely considered a reflection of the psychological reality of phonological constructs and processes, at least in the weak sense. It is true that experiments are limited in that they may not be used to decide between competing conceptual analyses which are essentially notational variants of each other. However, they may provide clues about the mentalistic import of phonological analyses in the weak sense discussed above.

Fourth, some phonologists implicitly or explicitly assume that well-developed theories or frameworks and extensive knowledge are pre-requisites to the use of experiments. In principle, it seems that the phonologists who think in these terms believe that a theoretical framework should be constructed and extensively elaborated before any of their hypotheses are experimentally tested. However, one cannot but doubt whether this is an excuse for postponing the experimental testing of such hypotheses, something phonologists may not be willing to undertake. In general, phonologists fail to realise that in the empirical sciences such as cognitive psychology and physics, theory building and experimentation are inseparably connected. As Eddington (1999: 18) states: "it would be absurd for psychologists to denounce the utility of psychological experiments simply because they felt that too little is known about perception,

²⁵ For example, if the presence of a tape recorder is claimed to make subjects self-conscious and thus give unnatural responses to the experimenter's questions, then a hidden microphone can be used. If the laboratory environment is to blame as it is a somehow artificial setting and intimidating, then the studies can be done in the street or in subject's homes (Ohala 1986: 10). However, experimenters should make certain that if an experiment is conducted in the field, the problem remains of presenting the task to native speakers in a way that will be meaningful and not intimidating to them, so that pre-existing rather than situation-specific categorisation may be tapped (Jaeger 1980a: 343).

learning, or cognition”; and he continues to argue that “the truth is that much of what is known about physics and psychology is a direct result of theory-based experimentation, and could not have been established in any other way” (p. 18). Clearly stated theory should naturally lead to and precede empirical research and descriptive analyses are the necessary forerunner of empirical investigation, but such theories need not be extensively elaborated. The only prerequisite for an experiment is a hypothesis which is consistent with the bulk of the existing scientific knowledge and which is stated in such a way as to specify what outcome would support the hypothesis and what outcome would disprove it. Once a hypothesis has been confirmed or refuted, the theory is then modified, which in turn leads to better hypotheses and more experimentation.

Fifth, some phonologists claim that experimentation proceeds very slowly, and in this way it retards progress in a discipline like phonology. The assumption is that if phonological theory had to wait for experimental support for every notion it proposes, slow progress would be made (Ohala 1986: 12). However, it is doubtful whether any progress is really made in psychological phonological theories that never test their hypotheses empirically. In fact, we only have to see the scores of different generative theories spun out of the same linguistic data to be convinced that such “progress” is probably an illusion. Just as a particular experiment may not shed all the light there is on a given phenomenon, and more experiments may be needed that clarify the initial results, non-empirical methods involve much less effort but the result is the same, only tentative, temporary “truths” (Derwing 1979; Ohala 1986: 12; Ohala & Jaeger 1986a: 4-5). Real scientific progress would follow if an equal amount of imagination and enthusiasm were spent in the design and conduct of experiments as is currently spent formulating the theoretical hypotheses that most phonologists never test.

Finally, some maintain the experimental method is not appropriate for phonology because linguistic behaviour is dependent on so many factors that it is virtually impossible to control them all in contrived, experimental settings. However, this only means that it may be more difficult to do experiments in linguistics (or any behavioural science), not that it is any less necessary (Ohala & Jaeger 1986a: 6).

2.2.2.2. Advantages of the Use of Experiments in Phonology.

The phonologists who advocate the search for psychologically valid phonological analyses but paradoxically criticise experiments as methodological tools for validating or invalidating their claims usually fail to notice the benefits of experimentation. In fact, external evidence obtained experimentally has been claimed to be the most useful information in testing the psychological adequacy of phonological analyses (e.g. Bertinetto 1992; Eddington 1999; Nearey 1981; Ohala 1986, 1995; Ohala & Jaeger 1986b; Nearey 1981; Wheeler 1980). In the

rest of this section, some of the advantages of the use of experiments in phonology will be discussed.

In the first place, the use of experiments contributes to the methodological reorientation that a truly psychological phonology needs in order to be considered a veritable empirical science. This was, by the way, the most noticeable failure of the so-called Chomskyan revolution, which once promised to develop a truly psychological phonology. However, classical generative phonology merely entailed a terminological reorientation in the direction of the psychologisation of linguistic jargon and aims, but no corresponding methodological revolution accompanied these changes (Derwing 1979: 17). In addition, experimentation is not associated with any particular phonological theory or framework. Thus it can be embraced by any phonologist working in any particular phonological framework, but who is interested in the psycholinguistic pertinence of phonological analyses. In other words, experimentation requires no particular theoretical affiliations.²⁶

Second, experiments provide an ideal opportunity to control for the number and characteristics of actual speakers and to focus on their performance on specific experimenter-designed tasks. The gathering of experimental evidence from a relatively large number of subjects avoids the strategy of trying to infer something about conceptual structure from a systematic analysis of linguistic structures or anecdotal information from a few subjects. In other words, the primary limitation shared by most psychologically-oriented phonological research, namely the problem of making psychological claims based on individual analyst's own intuitions as to what is actually psychologically valid for the speakers of a speech community, is avoided by having recourse to a relatively large homogeneous group of experimental participants. Any characterisation of what constitutes the phonological knowledge of the speakers of a language should ultimately come from data obtained from a significant number of linguistically naive and relatively homogeneous speakers of the language, the number and characteristics of which can be pre-defined for a given experiment.

Third, experiments make the phenomena under investigation or observation occur under controlled conditions created by the experimenter when he so wishes (Campbell 1986: 169-170; Nearey 1981; Ohala 1986, 1995; Ohala & Jaeger 1986b: 2; Weitzman 1993: 141). In this way, the influence of anticipated and potentially distorting factors or variables that might render their evidential value ambiguous can be neutralised (Ohala 1986: 10-11; Wheeler 1980: 68). Typically, this is done by directly contriving the situation under which the observation is made. This reduces the ambiguity of the evidence to a minimum. Thus, experiments allow control over the quality and quantity of the evidence gathered relevant to the crucial theoretical questions.

²⁶ Phonological investigation with an explicit commitment to the use of experiments has usually been referred to as "experimental phonology" (e.g. Ohala & Jaeger 1986a) or "laboratory phonology" (Pierrehumbert *et al.* 1996, 2000).

This is the opposite to external evidence obtained in naturalistic or “ecological” conditions, in which the experimenter observes behaviour without disturbing the environment.

Fourth, experiments provide inexhaustible empirical evidence of the highest quality on the declarative and procedural knowledge that speakers possess (McCawley 1986: 38; Nearey 1981; Ohala 1986: 10, 19).²⁷ This is because a particular phonological question can be investigated experimentally on countless occasions with a given experimental technique that can be progressively refined and because it can be investigated using different experimental techniques. To the extent that researchers spend their efforts on a specific problem, the evidence obtained will accumulate and, in time, shed more light on the problem under investigation. Additionally, experiments provide high-quality evidence because they control for the possible variables affecting the data. In this respect, if the data obtained are suspect of being influenced by unpredicted variables, new data of even greater quality can be obtained from subsequent experiments that control for that variable. The refinement of experimental techniques may also contribute to the gathering of increasingly better evidence. In addition, it helps the discipline gain experience in recognising previously overlooked sources of error and in finding ways to compensate for them (Ohala & Jaeger 1986b: 3). Consequently, conflicting results will, in principle, cease to appear at some stage of the process of data gathering (Campbell 1986: 172).

Fifth, experiments help to determine which of the proposed constructs and analyses of a given phenomenon have psychological validity for speakers, and which do not (Derwing & Nearey 1981; Eddington 1996, 1999). Given a set of linguistic data, there will be an unlimited number of ways to account for the data, and there is no *a priori* method for determining which way the speakers might choose. Experimentation helps to determine which regularities that phonologists perceive in their data correspond to the regularities that speakers are actually aware of, and which regularities play no part in their linguistic processing (Campbell 1979; Skousen 1975). Experiments allow phonological descriptions to be empirically tested and allow them to be subject to potential falsification (i.e. verification or refutation) as psychologically valid descriptions in the weak sense.²⁸ In other words, experiments serve to verify or refute the hypotheses put forward by theoretical phonology and allow one to choose unambiguously between competing analyses.

²⁷ Speakers’ judgements over the permissibility of certain phonological sequences is a classic example (e.g. Greenberg & Jenkins 1964; Ohala 1986; Ohala & Ohala 1986), morphophonological issues that go from experiments of the past and plural formation deal with relatively “close to the surface” morphological alternations (Derwing & Baker 1977; Derwing & Nearey 1986), to more abstract morphophonological rules like alternations in verb tenses (e.g. Bybee & Moder 1983), derivationally related words (Derwing & Baker 1977; McCawley 1986; Ohala & Ohala 1987), including the controversial “vowel shift” rules in English (see note X), speakers’ awareness of the morphemic constituents of complex derived words (Ohala 1986), morpheme structure constraints (Ohala & Ohala 1986), the internal structure of syllables with its parts like onsets and rimes (Treiman 1983, 1984), etc.

²⁸ A clear example is the long-standing debate between the linear view of the syllable, which regard the syllable as a linear string of phonemes with no internal structure (e.g. Hooper 1972) and the hierarchical view which claims that the syllable has a hierarchical internal organisation with units intermediate in size between the syllable and the phoneme, namely, onsets and rimes (Fudge 1969, 1987, 1989; Selkirk 1982). A substantial amount of experimental research reviewed by Treiman (1988, 1992) has shown that the hierarchical view of the syllable is the psychologically valid for speakers and not the linear view.



Finally, experimentation contributes to both the revision and refinement of existing phonological theories, as well as to the development and construction of new theories. Experimentation can aid researchers to revise and refine existing phonological theories because it helps in choosing between different theoretical phonological analyses as psychologically pertinent ones (Ohala & Jaeger 1986b: 2, 4). Experimental evidence is most efficient in allowing us to weed out unsupported hypotheses so that the discipline can spend its subsequent efforts and time on the hypotheses that receive experimental confirmation. Although psycholinguistic experiments are usually considered appropriate for testing and evaluating previous theoretical positions, the role of experiments is not merely to test prior theories, but also to contribute to the development of new or revised theories which can be submitted to a new round of experimental tests. In this way, experimentation may play an active role, not only in the evaluation, but also in the future development of phonological theory (Derwing 1979: 128, 131). The role of experiments in the revision, refinement, and development of phonological theories may thus lead to some actual scientific progress. This is so because the use of experimentation makes phonology exhibit what Ohala (1990, 1995; Ohala & Jaeger 1986b) calls “fruitfulness” and “convergence”. This means that although some of the alternative answers to phonological questions may eventually be rejected or modified, they are successful or fruitful if even in their error they stimulate a directed search for new data and for the formulation of new generalisations.

In short, several advantages about the use of experiments as instruments for validating the psychological pertinence of phonological analyses have been mentioned. These advantages, summarised in table 2, apply to other fields of research, not just phonology. However, they seem particularly relevant in the latter field precisely because of its traditional neglect of experiments and occasional misunderstanding of the way experimentation proceeds and the progress it brings to the field.

Table 2. Advantages of the Use of Experiments in Phonology.

➤ Contribute to the methodological reorientation of an empirical psychological phonology. Require no particular theoretical affiliations.	➤ Provide control for the number and characteristics of participants and to focus on their performance
➤ Make the phenomena under investigation occur under controlled conditions. Neutralise the influence of anticipated and potentially distorting variables. Allow control over the quality and quantity of the evidence gathered relevant to the theoretical questions.	➤ Provide inexhaustible empirical evidence of the highest quality
➤ Help to determine which of the proposed theoretical constructs and analyses of a given phenomenon have psychological validity for speakers and choose amongst competing analyses. Permit phonological descriptions to be empirically tested and be subject to potential falsification (i.e. verification or refutation)	➤ Contribute to both the revision and refinement of existing phonological theories, as well as to the development and construction of new theories.

2.2.2.3. The Complementariness of Descriptive and Psycholinguistic Phonology.

Given the misunderstandings in the past of what a truly psychological phonology implies, it is desirable that the goals of any phonological analysis and the methods used to attain them should be made explicit. Irrespective of the results of a particular investigation, a good analysis (and a suitable evaluation of such analysis) may only be attained if it makes explicit its domain (descriptive or psychological) and the methods used. Otherwise, progress in any of these two approaches to phonology will be held back, if not halted, to the extent that these two goals are confounded and no distinction is made between them (Wheeler 1980: 74).

However, the two approaches to phonology discussed are neither mutually exclusive nor irreconcilable. In fact, these two different roads can converge and can even be considered complementary. In this sense, it may be possible to provide a unified account of both the structure of a language, as well as what declarative and procedural knowledge speakers have of their native language or acquired languages (Eddington 1996). In other words, the emphasis on

distinguishing between a purely descriptive phonology and a psychological phonology, whose goal is to determine which phonological constructs or descriptions have psychological validity for speakers, does not mean that there is not, or should not be, an interaction between them. Descriptive phonologists could continue to systematise the patterns found in a language within a formal approach while psycholinguists or psychologically-oriented phonologists could verify whether the phenomena described are in fact meaningful to speakers. Moreover, a psychological phonology actually needs the work of descriptive phonologists. In fact, previously stated theoretical descriptions naturally lead to, and precede empirical research because experiments have little meaning unless motivated by theoretical questions (Eddington 1999). Theory is, for the experimentalist, a starting point for an empirical investigation. As Wand and Derwing state:

For the experimentalist... theoretical controversy, as the driving force that leads to better tests and clearer answers, is the meat that keeps the enterprise alive, not the poison that brings it all to an agonising and inglorious end

H. Samuel Wang & Bruce L. Derwing (1986: 113).

However, it seems that a committed psychological phonology should strive to embrace the tools and methodological procedures of those empirical sciences devoted to the study of human cognition.²⁹

The present dissertation takes the view that phonological descriptions should show psychological relevance and that phonological questions and competing phonological analyses should be tested empirically in order to attain psychologically plausible phonological theories.

2.3. Concept Formation as an Experimental Paradigm.

In order to investigate how speakers of a language categorise the sounds of their language, and thus obtain experimental evidence to propose phonological analyses that reflect what is psychologically valid for speakers, useful experimental techniques should be selected. Over the years, different experimental techniques have been used to study the categorisation of sounds (see sections 2.1.1 through 2.1.4.).

The concept formation experimental paradigm appears to be particularly suitable for one of the basic goals of the present investigation (namely, to obtain experimental behavioural evidence from language speakers' categorisation of the sounds of English to provide psychologically plausible accounts of phonological issues). In fact, its usefulness for assessing the psychological reality of phonological categories and revealing how linguistically native speakers classify the elements of their language has already been demonstrated (e.g. Jaeger 1980a: 381, 1986: 216; Ohala 1986: 18). Therefore, the technique was chosen as the experimental tool to be used in the experiments reported in chapter 3. The name of the

technique, (i.e. “concept formation”) is usually taken to designate a well-known experimental technique, not to suggest an abstract process (Baker *et al.* 1973: 207; Kendler 1961: 447).³⁰ This technique was originally and extensively used in psychology (e.g. Bolton 1977; Bourne & Bunderson 1963; Bruner *et al.* 1956; Dominowski 1970; Hull 1920; Smoke 1932) and was used to answer questions about the mental representation that subjects abstracted in the process of learning a category,³¹ the processes by means of which subjects identify, during the experimental session, the target concept,³² etc.

However, more recently, the technique has already been applied to linguistic questions other than phonological ones (e.g. Baker *et al.* 1973; Jaeger 1980a: exp.4). It has also been employed in studies on the categorisation of the sounds of English to address different phonological and/or phonetic questions (see 2.1.1 & 2.1.2.).³³ The technique has proved useful in obtaining information about within-category differences amongst various category members and about the boundaries between categories (e.g. Jaeger 1980a: 67, 1986: 212) as well as about the psychological status of the categories formed (e.g. Derwing & Wang 1995: 363; Jaeger 1980a: 67; 1986: 212; Ohala 1983: 238-239)³⁴ or even about the hierarchical structure of phonological taxonomies (Jaeger 1980a: 67, 1986: 212).

²⁹ For a similar discussion of the complementarity of both approaches on the web see LINGUIST List 4.934 (<http://www.linguistlist.org/issues/4/4-934.html>) and LINGUIST List 4.952 (<http://www.linguistlist.org/issues/4/4-952.html>).

³⁰ However, the term “concept formation” has usually been associated in the past with a specific mental process associated with the learning of a new category created by the experimenter in opposition to “concept attainment”, “concept learning”, “concept acquisition” or “concept identification”, or the process of discovering a category which the subject *already* has in his/her mind (Bruner *et al.* 1956; Hunt 1962). Despite this distinction, the dividing line between the supposedly different mental processes that these names imply is very hard to draw and the psychological reality of the distinction can be questioned (e.g. Bolton 1977; Picas 1966).

³¹ For example, probabilistic models became popular because they offered explanations of such phenomena as better (i.e. faster) categorisation of previously unexperienced instances representing the central tendency of the category under investigation than of novel instances not representing such central tendencies (e.g. Hartley & Homa 1981; Homa & Vosburgh 1976; Homa *et al.* 1973; Posner & Keele 1968, 1970; Strange *et al.* 1970). However, later research showed that an exemplar model of categorisation could equally account for such results (Hintzman 1986; Hintzman & Ludlam 1980).

³² This was done with the reception procedure (e.g. see section 2.4.). The assumption in those CF experiments was that performance in such a task was problem solving: the subject had to solve the “problem” of identifying an already familiar concept to him/her whose actual identity was not revealed at any moment before or during the CF task. Subject’s inspection of such stimuli would lead them to solve the problem by formulating and testing different hypotheses as experimental trials succeeded one another (Bourne 1982: 3; Bruner 1956: 55; Dominowski 1970: 152, 169-172; Hulse *et al.* 1980: 272-278; Rosch 1973a: 329, 1973b: 113). As many researchers mention (e.g. Baker *et al.* 1973: 207; Bourne 1982: 3; Neisser & Weene 1962: 640; Rosch 1973a: 328, 1973b: 112, 1975d: 180; Sutcliffe 1993: 57-58), the typical CF experiments of the information processing era used completely artificial categories in which the stimulus features and the categories to be learned were usually already well-known by the subject. For example, all subjects would know what was meant by red circles or by either red circles or blue squares and so on. What the subject had to discover was which particular concept the experimenter had pre-defined.

³³ This technique has also been used to study the phonological categories of other languages like Italian (Bertinetto 1992), Japanese (Jaeger 1980a: experiment 2) or Taiwanese (Derwing & Wang 1995).

³⁴ Subjects’ formation of a category is taken to imply that either the category pre-existed in the subject’s mind or that at least the components or attributes of that category pre-existed and are grouped together. For example, if subjects are asked to form a category like “large red octagons” (the shape of stop signs), it is reasonable to assume that for those subjects who are automobile drivers, this will be a pre-established category formed on the basis of prior experience; however, if subjects are asked to form the category “large red rectangles”, it is more likely that they have no such pre-existing category and that their performance would be based on their knowledge of the concepts “large”, “red”, and “rectangle”. Moreover, if a subject does not learn a category at all, it may be considered that neither the category pre-existed in his/her mind nor that the attributes or combination of attributes involved in the category under investigation. In this respect, Jaeger claimed (1980a: 67; 1986: 212-213) several ways in which the psychological pre-existence of categories (i.e. their prior existence in the mind) can be inferred from subjects’ behaviour in CF experiments. First, the probable lack of correlation between the ease with which a category is formed and whether it was known previously to subjects and second, the way subjects generalise a category to stimuli differing from those used to exemplify it. If subjects quickly and consistently include in the category certain tokens which differ qualitatively from those taught as being positive instances, then this could be a reflection of a previously existing category which contains a broader range of members than those presented as positive instances.



In the following sections (2.3.1. through 2.4.) we will review the different parts of the experimental paradigm, the data that can be obtained in a CF experiment, and the different versions of the technique in an attempt to better understand the experimental technique with which the experiments reported in this dissertation were conducted.

2.3.1. Parts of a Concept Formation Experiment.

Although CF is typically considered as a two-stage paradigm consisting of a ‘training’ (or ‘learning’) stage followed by a ‘test’ (or ‘study’) stage, CF experiments usually consist of four different phases: instructions, a learning session, a test session, and a post-experimental interview (Jaeger 1980a: 66-76; 1986: 214-221).

2.3.1.1. The Instructions.

The instructions necessary to perform a CF task can be presented visually or in spoken form. Instructions inform the subject about the nature of the experiment. First, the instructions ask the informants to focus their attention on discovering the presence or absence of some specific phenomenon in the stimuli to which they will be exposed in the experimental task. In other words, instructions inform subjects that the specific phenomenon appears in some items and not in others. The stimuli that contain the phenomenon form a class or category while those that do not are not instances of that category. Second, instructions ask subjects to respond to each stimulus in the manner specified after they have come up with an idea regarding which stimuli contain the phenomenon and which do not. Third, the instructions inform the subjects that they will be provided with feedback in an initial session (the “training” session) as to whether or not the stimuli that they are presented exemplify the phenomenon they have to discover. Finally, the subjects are told that after the initial session there is a second session (the “test” session) in which they will continue to be provided with stimuli that exemplify the phenomenon and others that do not. However, they are informed that feedback is not provided (or is provided to a limited extent) in this session.

In addition, there are three important aspects related to the instructions used in CF experiments that should be discussed. The first refers to the amount of information they convey about the to-be-learned category, the second refers to the moment during the experimental session at which instructions are presented, and the third has to do with the homogeneity of the instructions used across comparable CF experiments.

In the first place, experimenters have to be careful with the amount of information provided in their instructions (e.g. Ohala 1983: 237). On the one hand, if too much information about the to-be-identified phenomenon is given, subjects may base their responses on the

experimenter's implied category rather than on their own intuitive classifications. On the other hand, if too little or no information is provided, informants may either be incapable of forming any category at all, or may form an incorrect one. If the to-be-identified concept were not sufficiently described or not described at all, and subjects had to identify it without any previous description, the task could become very hard because there are many possible dimensions any stimulus may be categorised for, and there is no guarantee that informants would ignore irrelevant dimensions to the concept under investigation (Dominowski 1970: 153).³⁵

In the second place, it is important to decide whether all the information that the instructions convey should be presented at the same time or at different moments during the experimental session. In general, this amounts to presenting all the instructions before the training (or learning) session or presenting them at two different moments. This latter alternative amounts to giving the subjects the necessary instructions to perform the learning session at the beginning of that session and the instruction to perform the test session during the break between the learning and the test sessions respectively. If the former mode of presentation is adopted, experimenters should make certain that subjects understand that the experimental task has two different parts, and that a change from the first to the second will involve some variation, namely the restriction or elimination of feedback from the experimenter. If the latter mode of presentation is preferred, subjects may forget the criteria they used to form the category in the first part of the experiment. This may be particularly likely to happen with very difficult concepts. Consequently, this method of presentation should be avoided with apparently difficult concepts.

Finally, it should be noted that if a group of informants is to participate in a series of CF experiments, as in the case of the present studies, the instructions should be as similar as possible throughout so that all experimental subjects across different CF tasks receive the same amount of information and instructions.

2.3.1.2. The Training (or Learning) Session.

The aim of the training session (also called "study" or "learning" session)³⁶, is to teach the experimental subjects the phenomenon under investigation. This is done by training them to classify a set of items or stimuli into different groups or categories that have been pre-defined by the experimenter. For the sake of simplicity, let us assume that a single concept is to be formed or learned, and that there are two categories involved: the to-be-learned category made

³⁵ However, even when instructions appear to be relatively satisfying to the experimenter, informants may form categories based on any number of (possibly redundant) attributes besides the one(s) the experimenter has in mind.

³⁶ As Bourne puts it (1966: 20), the label "learning" is used for those tasks and experiments in which the emphasis and central interest lies in the acquisition of differential responses for formerly confusable attributes or of some complex behavior strategy which implements a formerly unfamiliar rule for grouping. For Hunt (1962: 2): concept learning is "acquisition, or utilization, or both, of a common *identifying* response to dissimilar stimuli".

up of stimuli belonging to that category, and a second category defined as not including the members of the first category.³⁷ In the learning session, subjects are trained to respond to a particular type of stimuli containing the phenomenon under investigation and exemplifying the category (i.e. positive stimuli) in one way, and to respond to another type of stimuli that do not include that phenomenon and do not exemplify the category, in another way.³⁸ This kind of learning is usually termed “intentional” because subjects are explicitly informed that they have the task of learning a given category and they perform such a task consciously and deliberately.³⁹

The learning session of a CF experiment takes place only after informants report having understood the instructions. In it, there are three critical events, stimulus presentation, response, and informative feedback. These three events, occurring in that order, constitute one trial (Bourne 1966: 5-6). A training session usually consists of a relatively large number of trials so that the experimental subjects have sufficient opportunity to solve the problem.

In a learning session, two sources of information are available to the subject, namely, the stimuli themselves and the informative feedback given by the experimenter. As mentioned above, stimuli may be positive or negative. Positive instances include clear, uncontroversial exemplars of the category. Negative instances should be made up of non-interfering (i.e. “neutral”) stimuli as well as of negative stimuli with some possibly interfering factors which are to be excluded from the concept. Negative interfering stimuli serve to guarantee that subjects do not inadvertently form some unwanted category using possibly misleading features.⁴⁰

Before the second source of information that subjects can examine is presented (i.e. the feedback), a decisional response must be made by the subject as a reaction to the stimulus. Thus, after each stimulus is presented, and the subject has some notion of what the category involves, the subject’s task consists in trying to give the correct responses, as instructed, after which the correct response is indicated with the provision of feedback.

Like the stimulus, feedback provides the subject with an external source of information relevant to the way the subject has responded and should respond on future trials. Feedback

³⁷ However, not all experimental tasks are quite this simple and well structured. There are many CF experiments which require the subject to learn several concepts concurrently; that is, the subject may be asked to sort the patterns into 5, 6, or 12 or any number of different categories and to produce a different verbal or motor response to each of these categories.

³⁸ The training or learning session can be carried out by all-positive instances or by all-negative instances, although a mixture of both is the most frequently used procedure in CF tasks. Hovland and Weiss (1953) found that more subjects attained the correct concept when it was transmitted by all-positive instances than by all-negative instances. Mixed positive and negative instances appear to be intermediate between all-positive and all-negative series in difficulty of learning.

³⁹ “Intentional” learning is the opposite of incidental learning or the unconscious acquisition of category knowledge (see e.g. Kemler Nelson 1984: 734-735). The overwhelming emphasis in laboratory studies has been on how categories are learned under intentional conditions. What subjects learn about categories under these circumstances may be very different from what they learn when such intentional strategies are not activated. It must be emphasised, though, that most of our categories have likely been acquired under incidental conditions.

⁴⁰ There are several studies and discussions which have assessed the relative differences in information between negative and positive items (e.g. Bourne 1966; Dominowski 1970; Hovland & Weiss 1953; Jaeger 1980a; Smoke 1932). The general consensus is that negative instances are more difficult (but not useless) to use for purposes of attaining a CF task. This is due to subjects’ difficulty in using negative information about the concept, their difficulty of assimilating information concerning what the concept “is” as compared with assimilating information concerning what it “is not”.

informs subjects about the status of each instance they are exposed to (whether it is a positive or negative instance of the category).

The temporal relationships among the three events of a trial (stimulus, response, and informative feedback) are particularly crucial in order to perform a CF task successfully. In this respect, the overall length of a trial may be broken down into three major components: the **pre-stimulus interval**, the **stimulus interval**, and the **post-stimulus interval**. The “pre-stimulus interval” is the period of time between an indication or warning signal that a forthcoming stimulus is going to be presented and the actual presentation of such stimulus. The “stimulus interval” is the period of time during which the stimulus is available to the subject for inspection. Typically, the subject will be required to make his response within this interval or shortly afterwards. The “post-stimulus interval” is the time between the end of the stimulus and the beginning of the next trial. The post-stimulus interval can be divided into three major components, namely the “reaction time interval” or time that the subject takes to respond to the stimulus in the manner specified, the “delay of informative feedback” (also called “feedback interval”), or period of time between the subject’s response and the presentation of feedback and finally, the “postfeedback interval” or the time between the occurrence of feedback and the presentation of the next stimulus.

Research suggests that variation in these intervals is relevant to some extent (see Bourne 1966: 69-72; Dominowski 1970: 154; Pikas 1966: 130). First, although delay of informative feedback has shown no significant effect on learning retardation (e.g. Bourne 1957; Bourne & Bunderson 1963), it would be unrealistic to conclude that delay of feedback is generally insignificant (e.g. Bourne 1966: 70). Longer delays than the ones typically used in experiments (e.g. more than 8 seconds) might produce significant retarding influences. Second, the length of postfeedback intervals is particularly critical probably because it is the period of time that subjects have at their disposal to study, assimilate, or otherwise process the information they have received from both the stimulus presentation and its corresponding feedback. In this respect, as it indubitably takes time to scan and to assimilate the information available on any single trial, if time is limited, performance may suffer. Therefore allowing more than a few seconds (i.e. more than two or three) after feedback improves performance. There is a marked improvement in performance as the postfeedback interval increases in duration. However, as both stimulus and feedback are not available during the postfeedback interval, informants might begin to forget about them when the interval becomes too long. In fact, relatively long intervals (ca 25 sec.) have been shown to retard learning the most while moderate intervals facilitate it (Bourne 1957; Bourne & Bunderson 1963; Bourne *et al.* 1965).

2.3.1.3. The Test Session.

The test session (also called the “transfer” session)⁴¹ is an optional part of a CF study (Weitzman 1993: 142). Its use depends on the specific goals of the study although most CF experiments include it. The subjects’ task in the test phase is essentially the same as in the training stage, that is, one of categorising stimuli presented one at a time, except there is no feedback. In other words, it is a sort of “absolute identification” test of the kind traditionally used in speech perception experiments. The test sessions of CF experiments have typically had at least one of the following three basic aims.

The first one is to determine whether informants have actually learned the category. In principle, if subjects reached criterion in the learning session, they should have no problems in continuing to provide correct responses to positive and negative stimuli of the type presented in the learning session. However, in order to guarantee that subjects have actually learned the category, the test session usually utilises so-called “**control**” tokens. These are clearly positive or negative instances of the category. However, they contain some attribute not yet encountered by the informants. Control stimuli are checks on the possibility that informants have not formed a category different from that intended by the experimenter, or that they may have just memorised the members of the category taught in the learning session. If subjects respond incorrectly to control words, then it is not clear what category they have actually formed. Subjects might reach criterion by consistently applying some incorrect criteria or simply memorising the category members. As a consequence, the informants’ responses to control tokens included in the test session have to be evaluated before deciding that they have actually formed the category correctly.⁴² The same applies to responses to positive and negative instances (as mentioned earlier). As Hunt puts it (1962: 41), from the subjects’ assignments, “the experimenter may be able to infer the concept which the subject has learned”. Thus the use of control tokens in the test sessions and their subsequent analysis becomes particularly important. If subjects generalise their responses to these new cases correctly, their behaviour more clearly indicates that they have actually learned the category.

The second aim of the test session is to ascertain how informants classify instances whose category membership may be doubted for some reason. These stimuli are called “**test**” tokens and they represent new cases which might be considered as category members, but whose actual

⁴¹ Transfer of training may be described as the influence of prior learning or experience in one task on performance in another. It refers to the utilisation of habits, associations, or knowledge acquired under one set of circumstances in a new, later situation.

⁴² In fact, Jaeger showed that, in her (1980a) experiments, one informant who reached criterion did not form the category correctly (as seen by random answers during the test) and another informant who did not reach criterion but seemed to have formed the category correctly, performed correctly on most of the learned and control pairs in the test session (Jaeger 1986a: 228).

category membership is unclear.⁴³ Test tokens provide the experimenter with information about the boundaries of categories formed by the subjects during the learning session. Examination of subjects' inclusion or exclusion of the test words in the target category gives information about the external structure and composition of the category and its boundaries.

These two aims justify why feedback as to the correctness of subjects' responses to all stimulus types is not provided. It is precisely subjects' unguided responses to all stimulus types which are of interest.

Finally, a third aim of the test session in some CF experiments has typically been to evaluate the nature of the category information acquired during experience with the training items, that is, whether a summary representation, or exemplars, or both, have been abstracted. Many studies have manipulated the nature of the training items in order to differentiate between alternative models (see section 2.3.1.).

2.3.1.4. The Post-Experimental Interview.

The final part of a CF session takes place after the test is over, and it is an interview conducted with the experimental subjects. The post-experimental session, like the test session, is optional and is considered as supplementary evidence to the behavioural data obtained in the learning and test sessions. The aim of this interview is to elicit subjects' spontaneous reports of how they performed the task and to get them to express all that they know about the phenomenon or category under investigation. This is accomplished by asking subjects different questions and discussing their answers with them. For example, subjects may be asked what they thought the difference was between the members of the category and the non-members, what they would call the category, etc.

It is important to bear in mind, however, that the interview method may not tell us everything the subjects know about a category or concept, but only what subjects *can* express about the category or concept. In fact, this might be less than what is actually revealed by their behavioural evidence. The general consensus is that although much of what is acquired may eventually be made available to conscious expression, often not everything that is stored can be expressed (Reber 1989: 231). The information subjects express is probably what is most accessible and salient for them. For this reason, the post-experimental interview may help to shed light on the behavioural evidence obtained during the experiment.

⁴³ In a sense, both test and control instances share an important characteristics: they contain some specific aspect not previously encountered in the training session. However, control items are clear positive or negative instances while the category

2.3.2. Interpretation of the Data.

Once a typical CF experiment has been performed in the ways specified above, experimenters have at their disposal different sources of information or measures of overall performance. These include the percentage or number of informants who form the category correctly, the total number of correct responses in the learning session, their trials-to-criterion scores, the time required to reach a solution, the proportion of correct to incorrect responses to different stimulus types, the reaction times to stimulus presentation (if measured), etc. In addition, if a post-experimental interview is conducted, subjects' verbalised knowledge about the category under investigation is available. A post-experimental analysis of these sources of data helps shed light on the questions under investigation.

2.3.2.1. Number of Subjects Who Form the Category.

One way to measure the difficulty or availability of a category is to find out how many subjects are able to form the category correctly. In Jaeger's words:

A category which 100% of the subjects can form correctly can be hypothesised to be a category of more general salience or accessibility than a category which only 30% of the subjects can form, for example.

Jeri J. Jaeger (1980a: 72).

This number is, in turn, inferred from the number of informants who reach criterion in the learning session. In CF experiments, a criterion is pre-established to decide whether informants have formed the category correctly in the learning session. Criterion is usually defined in terms of a specific, fixed number of successive correct trials, and/or a fixed number of trials with N or fewer errors at any point in the learning phase.⁴⁴ The subjects who are able to meet or satisfy this criterion are considered to have successfully formed the required concept and any mistake made after that point is considered a performance error (Wang & Derwing 1986: 107).

However, another way to decide whether subjects reach criterion is to examine the number of overall cumulative errors from the beginning of a CF task until the very end (Weitzman 1993: 142). For example, if a subject's cumulative errors continue to rise at a more or less linear rate until the end of, say, a 60-trial learning session, that subject has clearly not learned to make the category distinctions. However, if after so many trials, the curve of cumulative errors starts becoming asymptotic, until very few, if any errors are occurring, this is *prima facie* evidence that the subject has acquired mastery of the category under investigation.

membership of test stimulus is uncertain.

⁴⁴ In error-free trials in succession the number was 10 trials in Wang and Derwing's (1986)'s study, 13 in Dennis et al's (1973) study. In a fixed number of trials with N or fewer errors, the criterion was 15 trials in a row with 2 or fewer errors (Jaeger 1980a, 1980b, 1984, 1986; Ohala 1983, 1986) or 14 trials with 2 or fewer errors (Derwing & Wang 1995). Such criteria are usually

For example, it is very difficult to provide 37 correct responses at random in a learning session that contains 60 trials ($P\text{-value} = 0.03 < 0.05$).

The number of subjects who reach criterion is a very important measure of actual category learning. However, in order to make certain that subjects have actually learned the category the number of correct answers to both positive and negative stimuli (including controls) in the test session should also be observed. If subjects have actually learned the category, they should have no problem generalising their responses to both positive and negative stimuli in the test session. However, the fact that this is usually so should not be overlooked while deciding whether a subject has actually learned the category or not. If the subjects provide many incorrect answers to such stimulus types in the test session (i.e. positive and negative), it is doubtful whether they have actually learned the category intended by the experimenter, and their answers to test stimuli in the test session are of questionable value. Therefore, before a final decision on how many subjects have actually learned the category is adopted, such response patterns should be analysed. In addition, further evidence that subjects have learned the category intended by the experimenter is also obtained during the post-experimental interview when the subjects' are asked to name the category.

The data from the subjects who perform a CF task but are unable to successfully form the concept are usually examined separately from those who reach criterion. Such data are not included with the results of those who attained the concept, even though their analysis may provide useful information regarding why subjects did not learn the category (simply because it was somehow difficult, because they focused on something different, etc.).

2.3.2.2. Trial-to-Criterion Scores and Time to Solution.

The total number of correct responses in the learning session, the average number of trials to reach criterion and the time taken to reach such criterion also indicate, like the percentage of informants able to form the category, whether a given category is more or less difficult.⁴⁵ A category for which the average number of trials to criterion is, for example, 30, out of 60 total trials, will presumably have been an easier category to learn than one for which the average number of trials to criterion is 55. If several CF experiments are compared, observing the three measures (i.e. number of subjects who reach criterion, average number of trials to criterion, and time to solution) may reveal which categories are more difficult to attain than others.

established taking into account binomial distributions and obtaining a number of successive responses out of a given set of trials that it is very difficult to achieve by chance.

⁴⁵ The average number of trials to criterion, number of correct responses, and time to solution provide essentially the same performance index. In the first place, correlations between them are invariably high, ranging around .90 (e.g. Bourne 1957). If errors are large in number, trials and time to solution are also found to be large. Furthermore, it is almost always the case that these measures are affected in the same way by important independent variables. For example, Bulgarella and Archer (1962) reported that errors, trials, and time increase in parallel fashion with task difficulty. In general, then, the outcome of most experiments can be summarised in terms of any one of these measures alone.

When a series of experiments is being performed, the criterion chosen should be held constant across experiments so that this measure can be compared among the different tasks. By comparing the number of trials that it takes to subjects to reach criterion across various conditions or by comparing the total number of errors in a fixed number of trials, good measures of the relative learnability of the various categories can be obtained (Weitzman 1992: 116). In this respect, differences in the number of trials needed to reach criterion in different CF experiments would indicate that some concepts are more difficult to form or less available or salient to consciousness than others. A category which has a very high mean trials-to-criterion score would be posited to be rather low on the hierarchy of concepts, while one which is formed rapidly by most informants would be higher on the hierarchy.

Furthermore, if trials to criterion across several experiments are to be compared, the order of presentation of positive and negative tokens should be the same in all experiments. This is desirable as different studies have found out that the order in which information is presented affects the learning process (e.g. how hypotheses are tested, retained, or rejected, etc.), as well as the possible representation of that knowledge in memory (e.g. Bruner *et al.* 1956; Elio & Anderson 1984; Hovland & Weiss 1953).

2.3.2.3. Subjects' Correct vs. Incorrect Responses to Different Stimulus Types.

Response patterns to different types of stimuli are of primary importance. In the learning session, the pattern of errors in subjects' responses to both positive and negative interfering and non-interfering tokens can provide information about which members are relatively easy to categorise and which cause more interference.

In the test session, correct answers to both positive and negative instances indicate whether subjects have actually learned the category. In addition, correct answers to control words indicate that the subject is able to generalise the concept formed in the learning session to new clear member or non-members of the category not previously encountered in the learning session. As mentioned above, numerous incorrect answers to positive and negative stimuli (including controls) indicate that, although criterion was reached in the learning session, informants have probably formed a category different from that intended by the experimenter. Correct responses to positive and negative stimuli at this stage are important because reaching criterion in the learning session and actually learning the category do not necessarily correlate. Finally, subjects' answers to test instances reveal whether the category includes potentially ambiguous category members or not, help chart the boundaries of the category under investigation, and clarify the category status of such instances.

2.3.2.4. Subjects' Reaction Times.

In CF tasks, the time it takes an informant to respond affirmatively or negatively to each stimulus (i.e. his or her reaction time), can shed light on the subjects' categorisation processes and perception of the experimental stimuli. For example, longer RTs may indicate less confidence in one's responses; thus, positive tokens which consistently elicit long RTs are probably less clear examples of the category, and negative tokens which elicit long RTs probably have some salient attribute in common with the positive tokens (see Jaeger 1980a: 75, 219-220 for a longer discussion).

2.3.2.5. Subjects' Verbal Reports in the Post-Experimental Interview.

The post-experimental interview reveals all the information that subjects happen to be able to spontaneously verbalise or be induced to express about the category under investigation. This is again useful information to decide whether the subjects have learned the category or not.

One would expect anyone who knows a concept to be able to provide a reasonably explicit description or name for it (Bourne 1966: 4); it has occasionally been suggested that an excellent indication of a conceptually relevant grouping of entities is "the fact that it has an easily elicitable and widely recognised (standardised) name" (Dougherty 1978: 68). Although the ability to discriminate and sort stimuli into categories and the subsequent generalisation to new instances not previously encountered is usually sufficient to assume that a concept has been learned, there is "a persistent tradition that something more is required if we are to talk of concept formation" (Lea & Harrison 1978: 535). This finds expression in a number of ways, typically in the custom of asking subjects in experiments for a verbal expression of the concept they have learned, or the basis on which they have found the category, that is, their ability to name or describe the category (Bourne 1966: 4; Lea & Harrison 1978: 535). In this respect, the notion of codability is a useful one in order to interpret subjects' category-naming behaviour in CF experiments (Jaeger 1980a, 1980b, 1986). It seems that categories that are higher on the availability-to-consciousness hierarchy will be more codable in four different ways: 1) they will be named with a single word or with very few words; 2) subjects will respond rapidly when asked to name them; 3) they will agree with one another as to the correct name of the categories; and 4) a single subject will be consistent in the name he or she gives to the categories at different times.⁴⁶ However, it must be pointed out that although subjects who form a category are usually supposed to be able to name, describe, or provide some information about the strategies they use (Jaeger 1980a) or the attributes that they consider relevant (Bourne 1966: 4) to classify something as a category member or not, it is possible to learn and extend a concept to

⁴⁶ In Jaeger's (1980a, 1980b) experiments, evidence was gathered on the first and third of these points.

new cases correctly without being able to define it, name it or even explain on what basis they are making generalisations (Kendler 1961: 448). Such concepts are, perhaps, probably very low in the accessibility hierarchy (Jaeger 1980a; 1986).

2.4. Versions of the Concept Formation Experimental Paradigm.

Although a CF experiment usually has a basic two-stage structure made up of the learning session plus the test session, there are as many slightly different versions of the CF technique as allowed by the imagination or resources of the experimenter who uses it (Jaeger 1986a :214). Bourne expressed the idea in the following way:

A paradigm is a general plan or method for conducting research. It is not, however, an inflexible prescription for what an experimenter should do. On the contrary, its operations are typically quite modifiable and adaptable to the requirements of each new experimental problem. Further, it consists of little more than a skeleton, to which the particular manipulations and measurements of variables unique to the experiment may be attached.

Lyle E. Bourne (1966: 5).

Apparently, the primary difference between versions of the CF experimental paradigm lies in a) the technical devices employed, b) what subjects have to do, c) where they are required to do it, d) when they are required to do it, and e) the amount and type of information they receive during the experimental task.

First, a CF experiment can be conducted with tape-recorders or computers. Most of the CF studies in the fields of experimental phonology or speech perception have used tape-recorders (e.g. Derwing & Wang 1995; Jaeger 1980a: experiments 4, 5, & 6; Jaeger & Ohala 1984; Ohala 1983, 1986; Wang & Derwing 1986), although a few have used computers (Jaeger 1980a: experiments 2 & 3; Weitzman 1992). The fact that most CF have been conducted with tape-recorders may simply be a reflection of the reliance on this method at a time (the 1980s) when the use of computers was not so widespread as it is today.

Second, as far as subjects' task is concerned, there are different variations over exactly what subjects are required to do. For example, subjects may be required to perform both the learning and test session, as most CF studies illustrate (e.g. Bertinetto 1992; Jaeger 1980a; Ohala 1983, 1986: groups 3 & 4) or they may only be requested to conduct a learning session (e.g. Derwing & Wang 1995; Ohala 1986: groups 1 & 2). Also, subjects may be asked to perform the learning session until they reach criterion (e.g. Jaeger 1980a: experiments 4, 5, & 6)⁴⁷ or to continue to perform the learning session until the end, even if they have reached

⁴⁷ If subjects have been informed that the task consists of two parts (i.e. the learning and the test session), it is recommendable to avoid telling the subject whether he or she has reached criterion or not. Instead, it is good to transfer subjects to the test session even if they have not reached criterion for two reasons. First, this does not discourage subjects who, if informed that the test has two

criterion (Jaeger 1980a: experiments 2 & 3), which is the usual case. In addition, subjects' responses in both the learning and test sessions may be verbal like saying "yes" or "no" (see e.g. Bertinetto 1992; Derwing & Wang 1995; Jaeger 1980a: experiments 4, 5 & 6; 1980b: experiment 2, 1984; Jaeger & Ohala 1984; Ohala 1983, 1986; Wang & Derwing 1986)⁴⁸ or motor, like pushing or pressing either of two buttons or keys, etc. (see e.g. Jaeger 1980a: experiments 2 & 3; Weitzman 1992). In the former case, the version is usually referred to as "CF with a verbal response" while in the latter case it is referred to as "CF with a motor response" or "automated concept formation" (Jaeger 1980a). Finally, subjects may be interviewed after the test session or not. If they are interviewed, they may be asked to produce oral responses, which may be recorded either on a tape or on paper. In the latter case, it may be written down by the experimenter who interviews the subjects or by the subjects themselves. An alternative method is to give the experimental participants a written questionnaire or oral questions can be combined with a written questionnaire.

Third, the setting in which a CF experiment is performed can also vary. CF experiments have typically been performed in the laboratory possibly because of its specific function to perform experiments and/or its advantages for the experimental situation (e.g. background noise is better controlled for in a laboratory setting). Laboratory settings are suitable for conducting CF experiments with different subjects performing the task simultaneously (e.g. Derwing & Wang 1995), although many studies make subjects perform CF tasks individually (e.g. Jaeger 1980a), which is the usual case. To our knowledge, no CF has been conducted in the field. This simply reflects researchers' preference to conduct CF experiments in laboratories or sound-attenuated rooms, not the inviability of doing so in the field. In fact, Jaeger (1980a: 350, 380) claimed that CF investigations outside the laboratory would be desirable with speakers from a broad range of languages (including unwritten ones). However, at the beginning of the 1980s, as Jaeger acknowledged (1980a: 343), the computer technology necessary to conduct CF experiments outside the laboratory was available but rather expensive, and CF experiments could only be conducted with tape-recorders. However, with the advent of affordable laptop computers a CF experiment using such a technology (either in its motor-response or verbal-response versions) could be easily conducted outside the laboratory.

Fourth, CF tasks can also vary in relation to the moment at which the experimental subjects have to perform some of the specific subtasks of the experimental session. For example, all instructions can be read at the beginning of the experimental session or the information necessary to perform the learning session can be presented before it begins and the directions necessary to conduct the test session in the inter-session period. Most CF experiments

parts, notice that they are not given the opportunity to do the second part; and second, they may have actually learned the category even if they did not reach criterion (see note 76).

⁴⁸ Verbal responses usually imply subjects' affirmative (e.g. "yes", "correct", "right") or negative ("no", "incorrect", "wrong") statements as to the category membership of the stimuli presented.

have provided the experimental subjects with all the instructional materials at the very beginning of the experiment. In addition, the pacing of the task or the time subjects have to perform it may also vary. In CF experiments and similar tasks, informants may work at their own pace or at a pace established by the experimenter. In other words, stimuli may be presented at given intervals of time (irrespective of whether subjects provide answers or not) or, on the contrary, the presentation of stimuli may follow the subjects' pace. This basically amounts to performing a CF experiment which uses a reception procedure or a selection procedure. In a selection procedure, informants are allowed to choose the stimuli about which they will obtain information and usually (but not necessarily) take the time they need to choose stimuli. On the contrary, a reception procedure, the experimenter controls which stimuli are shown and determines the feedback interval and postfeedback intervals, forcing the subject to respond in some brief period of time (Bourne 1966: 5-8; Dominowski 1970: 153; Hulse *et al.* 1980: 267-268).⁴⁹

Finally, experiments can also vary in relation to the information subjects receive during the experimental session. For example, stimuli may be presented one at a time (the successive presentation method) or simultaneously (Bourne 1966: 6). In other words, subjects encounter stimuli in a sequential order (e.g. Baker *et al.* 1973), in which case they must keep track of the stimuli presented or they can see all stimuli at once (e.g. Dennis *et al.* 1973). CF experiments with visual stimuli have typically used the successive presentation method. CF tasks with auditory stimuli have to use that method out of necessity. In addition, the type and amount of feedback given in a CF may vary from one experiment to another. For instance, feedback may be visual (e.g. green and red lights) or auditory (a voice saying "yes", "no"). Auditory feedback usually consists of a voice saying "yes" or "no" (e.g. Ohala 1986), "correct" or "incorrect", "right" or "wrong". Visual stimuli may consist of a green light or a red light that turns on/off to signal the category status of a given stimulus. In Weitzman's (1992) CF study, subjects received feedback in the form of a smiling face (correct) or a frowning face (incorrect) that appeared on the computer screen. In addition, although feedback is generally provided for *all* the stimuli in the learning session and for *none* in the test session, feedback in the test session may occasionally be provided for some of the negative and positive stimuli of the type presented in the learning session depending on the difficulty of the concept. If the concept is supposed to be relatively simple, no feedback is provided as it is unlikely that informants will have trouble transferring it (i.e. remembering it) from the learning to the test session. However, if the concept is considered relatively difficult, lack of feedback in the test session may cause informants to

⁴⁹ The reception procedure is older (see e.g. Hull 1920) and is the typical presentation method in CF experiments. The selection procedure owes its development to the work of Bruner and his co-workers (Bruner *et al.* 1956) and their interest investigating the strategies subjects follow to attain or form a concept. In principle, the selection procedure is the most suitable to know about subjects' hypotheses because it allows the subject to select his own instances in a problem and experimenters can determine from stimulus selections (and corresponding hypotheses) whether or not the subject is using any systematic plan or attack

forget what criteria they were previously using to guide their responses. As a consequence, subjects' responses in the test session may become increasingly random and it may be desirable to provide feedback for at least some percentage of tokens which are neither control nor test items.

Many more variations might be mentioned or imagined. However, no matter how many variations a CF task may involve, it is more or less safe to conclude that irrespective of the version of the CF experimental paradigm that is used, the success of the experiment may not depend so much on the version itself but on a host of factors that also have to be controlled for.

or strategy in the problem. In reception experiments, on the contrary, the subject is in a sense at the mercy of the experimenter for information about the solution.

CHAPTER III.

EXPERIMENTAL EVIDENCE FROM THE CATEGORISATION OF SOUNDS: A PSYCHOLINGUISTIC PHONOLOGICAL INVESTIGATION.



3. 1. Introduction.

The previous chapter can be considered as a review of the extensive body of evidence that shows that categorisation is ubiquitous in the speech mode as well as in the visual mode perception, as shown in Chapter I. However, in Chapter II it was also argued that the categorisation of sounds by experimental subjects would provide useful evidence to solve theoretical controversies in phonological theory with an explicit psychological commitment. An experimental technique was then selected and discussed. The aim of chapter III is to conduct different experiments to obtain evidence on specific phonological and/or phonetic controversies to propose psycholinguistically plausible phonological analyses.

3.1.1. Overview of the Present Experiments.

The purpose of the present section is to describe different issues common to the four experiments reported in this chapter. This will make the overall motivation, design and procedure of each experiment easier to understand and will spare the reader unnecessary repetitions in the course of reading the specifics of each CF experiment.

3.1.1.1. Aims of the Experiments and Selection of the Issues under Investigation.

The explicit general aim of the experiments discussed in the present chapter is to obtain experimental evidence to provide an account of certain phonological issues that closely corresponds to the knowledge speakers have about the sound pattern of their language. In other words, the purpose of the four experiments is to propose phonological descriptions or accounts that represent speakers' actual knowledge of the phenomena under investigation as revealed behaviourally by their categorising performance.

When a researcher wishes to conduct an experiment in phonology, the phenomena under investigation have to be carefully chosen and must be somehow relevant to the field. In this respect, it is advisable to follow the advice of Pierrehumbert and her co-workers when they state that:

... Given the rapid pace of change in theoretical linguistics, and the great expense and labor of laboratory research, the shrewd experimentalist will not devote an experiment to even the most central claim of any single linguistic framework. Instead he or she will look for a topic which represents a source of tension across many frameworks, or which has remained unsolved by traditional methods over many decades.

Janet Pierrehumbert, Beckman, Mary E., & D. Robert Ladd (2000: 279)

In this respect, the choice or selection of the specific phenomena under investigation was the result of the casual observation that theorists usually disagree on what the phonological classification of some very specific phonetic entities should be. These phenomena are the categorisation of the so-called “semi-vowels” or “semi-consonants”, the classification of the English affricates / tʃ / and / dʒ /, the categorisation of bilabial stops after word-initial [s], and the psychological status of the linguistic unit called “allophone”, exemplified by the allophone [p^h]. The specific categories taught in the four experiments reported below were “consonantal sounds”, “oral plosive sounds”, “/ p /”, and “[k^h]”. The introductory sections of each of the four experiments reported below make explicit what the controversies are in relation to the categories investigated.

3.1.1.2. Selection of Subjects.

The selection of the experimental participants was guided by the principle of obtaining the greatest degree of homogeneity. 80 British native speakers of English between the ages of 18-45 (mean age 23 yrs) were selected for participation in these studies. There were 38 men and 42 women. The subjects were recruited either on the University of Murcia Campus or in different places in the town of Murcia. Some of them were recruited by word of mouth or through an advertisement that briefly explained that some research on the sounds of English was being conducted and that native British adults with no hearing problems were needed.

Before the beginning of the experiment, subjects were given a questionnaire to evaluate the homogeneity of the group. Information about the name, age, place of birth and different questions about the educational (including phonological/phonetic) background of the experimental subjects was elicited. The questionnaire revealed that they spoke some other language (Spanish mainly) to some extent although none of them was completely bilingual. All subjects were university students when they performed the experiments or had been university students in the past and were professionals or tourists. Very few of them had received formal instruction in phonetics and/or phonology (less than 8%). If that was the case, subjects had, at

the most, studied a brief introductory course that lasted for no longer than 6 months some time ago and reported remembering very little about it. The post-experimental interview later revealed that they did not remember anything derived from their formal instructions in phonetics and/or phonology about the specific category they studied experimentally. For this reason, the whole group could be characterised as phonetically naive. Subjects reported no history of a speech or hearing disorder. Subjects were paid 6,01 € (1000 pts.) for their participation.

3.1.1.3. Arrangement of the Experimental Tasks.

While designing the present experiments it became of primary importance to decide whether every subject should perform the four CF experiments or, on the contrary, only one experiment. As far as the former option is concerned, it seems that although having the same subjects perform a series of CF experiments and compare their performance in these tasks would shed some light on the relative difficulty in learning different concepts, two possible problems have to be taken into account. First, as the estimated duration of each experiment was 30/35 minutes (including familiarisation with the set of instructions, performance of both the learning and test sessions, and the post-experimental interview), subjects might take at least 120 minutes to complete the four experiments. As a consequence, subjects' increasing fatigue might have a negative influence on their overall performance (Jaeger 1980a: 186-187). One possible solution to this problem is to make each learning session shorter. If the learning session finishes once subjects have reached criterion, this would reduce the overall time. In addition, if instructions are the same throughout the experiments, subjects could read them only once (at the beginning of the first CF experiment), which would also reduce the overall time needed to perform all the experiments.

The second problem has to do with subjects' increasing familiarity with the experimental paradigm, which might affect the ease with which a particular category could be formed. In this respect, if a group of subjects is to perform a series of CF experiments, the order in which the tasks appear might have some influence on subjects' overall performance. Subjects do progressively better on succeeding tasks and poorer on the initial one(s). In fact, improvement in performance over successive tasks has actually been observed in CF experiments (e.g. Haygood & Bourne 1965). The general trend is toward improved performance with practice and smaller differences between conditions. As Jaeger puts it:

The fact that subjects are learning about the experimental format during the first CF session could make their performance somewhat weaker on the first task.

Jeri J. Jaeger (1980a: 187).

One possible solution to minimise or neutralise acquired familiarity with the experimental paradigm may be to include all the various possible orders in which the different experiments might be performed (e.g. 1234, 1243, 1324, 1342, etc.) so that no unique specific order of presentation privileges any category or makes it be at a disadvantage in relation to the others. However, this implies 24 combinations and therefore 24 different CF experiments. Unfortunately, to obtain an acceptable number of subjects (e.g. 10 people) for each group was somewhat problematic due to the author's limited resources.

For these reasons, it was decided that four experimental groups should be created, that each of the eighty subjects should be randomly assigned to each of the four groups and that each group would perform only one CF experiment. In this way, familiarity with the experimental design would be the same for every subject. In addition, subjects should not become excessively tired of performing the task and fatigue might not have a negative influence on subjects' performance or at least have the same effect for every subject.¹ Therefore, the number of participants in each experimental condition was twenty, which is almost the mean number of subjects employed in previous CF experiments.²

3.1.1.4. Stimuli, Materials, and Apparatus.

The stimuli were 400 carefully chosen monosyllabic real English words of different syllable structures. The words had been divided into four lists of 100 words each. The words were nouns, adjectives or verbs occasionally preceded by a definite or indefinite article, pronoun, etc. (e.g. "a kind"). In general, in compiling the word list proper names, learned or scientific words, very low-frequency words and anglicised foreign words were somehow arbitrarily rejected. Words were relatively common English words. The reason why only words were used and not larger linguistic units like sentences, etc. was twofold. First, the phonological word has usually been considered as the basis for phonemic analysis given that the analysis of larger sections is a great deal more complicated (Gimson 1980: 50) and second, most CF studies have used such a unit.

All the stimuli of the experiments were produced by a female 22-year-old native speaker of English from the south of England. She was a temporary language assistant at the university of Murcia and had been in Spain for over a year. In the judgement of different University of Murcia staff the speaker had not acquired marked Spanish features in her pronunciation. In the spontaneous judgement of different experimental participants in the post-experimental interview, the speaker had a standard English Pronunciation. A speaker with such an accent was

¹ In order to familiarise subjects with the CF paradigm, some studies include pre-test (or practice test) to guarantee that subjects know how to perform the task (e.g. Derwing & Wang 1995).

selected on purpose because this accentual variety, known as BBC English or RP, has traditionally been considered as the most widely understood. Native speakers of English may vary widely as to their accent but, as they apparently understand speech spoken with the so-called BBC English accent without problems (Mompeán-González & Hernández-Campoy 2001), it was considered appropriate that the experimental participants should be exposed to this accentual variety. The speaker did not know the ultimate experimental purpose of the recording but was simply told that the recording might be useful for future investigations. Before recording the stimuli, the speaker was instructed to pronounce the words as naturally as possible (neither quickly nor slowly) with no stress variations and to repeat any word that she happened to mispronounce or pronounce with a remarkably different pitch contour. Even so, after hearing the recording once finished, pronunciations considered to be poor tokens of their respective words were recorded again until they were deemed acceptable.

The stimuli were digitally recorded and stored on a hard disk file in an EcoKey Computer (pentium 200) using the DARTPRO audio processing programme (distributed by Dartech, Inc.) in a sound-attenuated room with a noise level between 30 and 35 dBA, as determined by sound level metre measurements. An Electrovoice 635A microphone attached to a pair of Yamaha YH-1 headphones and placed 10-12 inches from the subject's mouth in a plane approximately 2 inches above the upper lip was also used to record the stimuli.

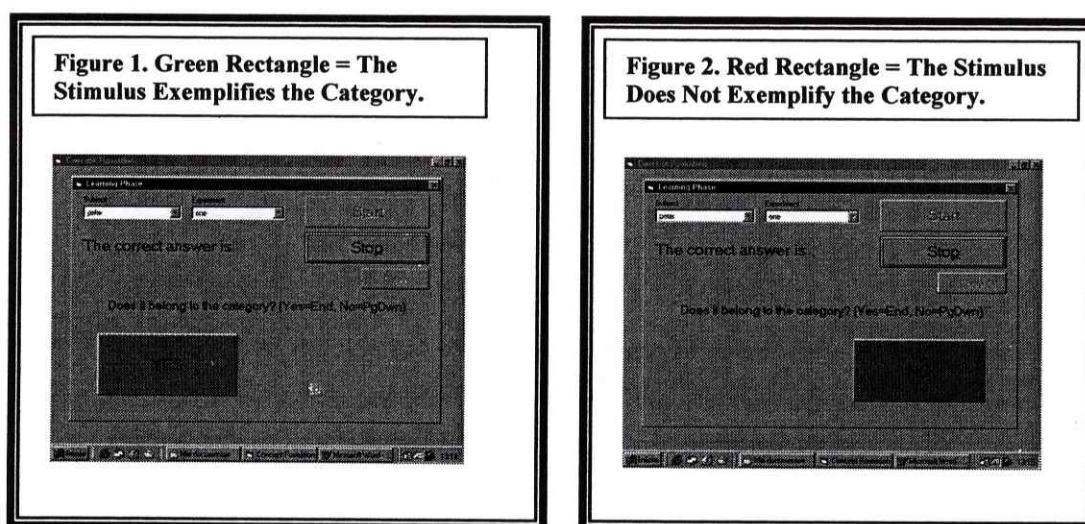
In a post-recording laboratory working session, the experimenter extracted each monosyllabic word from each of the 100-word recordings and created a specific audio file in which the time between the beginning of the sound file and the beginning of the word and between the end of the word and the end of the audio file was 5 msc respectively. As there were four 100-word lists, 400 audio files were created in this manner and stored on the hard disk. Definite or indefinite articles or personal pronouns were eliminated in the post-recording creation of sound files.

All the CF experiments and experimental events in this study were controlled by the aforementioned computer in which a software programme specifically designed to perform the present CF experiments had been installed. This programme was designed to run the learning and test sessions of the experiments automatically. The programme permitted to predetermine, at will, two time intervals for each experimental trial, namely, the time between the beginning of the stimulus and the presentation of feedback (i.e. stimulus interval, reaction time intervals, and delay of informative feedback) and the post-feedback interval. In addition, the programme allowed to retrieve any number of audio files previously stored on the hard disk, to order them sequentially to design a given CF experiment, and to indicate the status (positive, negative or test instance) of the stimuli in relation to the category under investigation.

² The number of subjects in some CF experiments is, for example, 7 (Jaeger 1980a: experiments 6 & 7), 9 (Jaeger 1980a: experiment 2), 10 (Ohala 1983, 1986), 15 (Jaeger 1980a: experiment 5; 1984), 18 (Jaeger 1980a: experiment 4), 20 (Derwing &



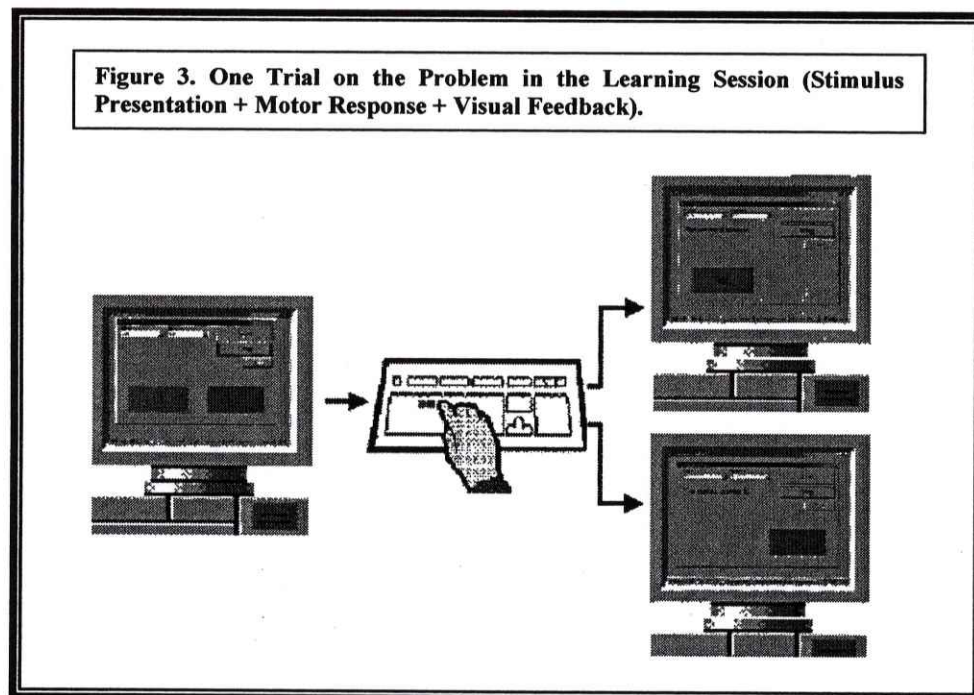
At the very beginning of each CF experiment, every subject, sitting in front of the computer, could see two rectangles on the screen: a red rectangle with the word “no” written inside and a green rectangle with the word “yes”. The two rectangles appeared in the lower part of the screen next to one each other. When subjects were ready to begin the task (the **learning session**), they had to click on a button labelled “start” in the top right hand side corner of the screen. When this was done, subjects listened to the stimuli over headphones. All stimuli in each experiment were presented binaurally over Yamaha YH-1 stereo headphones at a comfortable listening level (about 68/74 dB SPL-B) in a sound-attenuated room. It is important to stress the fact that subjects *never* had any access to their orthographic representations. When the stimulus happened to be a positive member of the category, the red rectangle immediately disappeared and the green one remained to indicate that the stimulus was a positive example. The same would apply, *mutatis mutandis*, when stimulus would turn out to be a negative member of the category. Figures 1 and 2 show the two different looks of the screen of the computer when feedback was being provided.



After the last trial in the learning session, the programme stopped. Subjects were then automatically transferred to the **test session**, which they began after clicking on a “start” button also in the top right hand side corner of the screen as soon as they were ready. This part of the experiment was exactly the same as the learning session except for the fact that subjects were never provided with feedback. Therefore, the two rectangles, still on display on the screen, remained static throughout this part of the experiment.

On both the learning and the test sessions, subjects pressed either of two appropriately coloured computer keys (either a red key or a green key) on the top left-hand side corner of the computer keyboard to indicate their responses. So, their responses were motor and the reason

why the specific version of the CF experimental paradigm used in the present experiments is the so-called “concept formation with a motor-response” or “automated concept formation” (see figure 3). The programme automatically recorded subjects’ responses to all stimulus types in both the learning and the test session. The programme recorded “P” whenever the green key had been pressed, which meant that the subject thought that the stimulus exemplified the concept, “N” when the red key had been pressed, which indicated that the subject thought the stimulus did not, and “X” when no answers were provided at all. The programme also recorded subjects’ reaction times. The beginning of the stimulus triggered a computer interrupt which initiated response time counting.



1.1.5. Specific Procedural Aspects of the Different Parts of the CF Experiments.

As mentioned earlier, the CF experiments were conducted in a sound-attenuated room. Subjects were tested individually. The experimenter gave every subject the instructions, which were read quietly with no time pressure (ca 5 minutes on average). Then the experimenter left and remained out of sight at some other part of the sound-attenuated room separated from the experimental work station through a window during both the learning and test sessions, which together lasted around 15 minutes. The experimenter remained outside the view of the experimental subjects to avoid inducing feelings of being watched in them and evoke an unsatisfactory level of self-consciousness in the experimental subjects. At the end of the test

session, the experimenter, informed by a short tone emitted by the computer at the end of the experiment, returned to conduct the post-experimental interview.

As far as instructions are concerned, an effort was made to render them as similar as possible. In fact, they were the same for the first two experiments, the same except for the description of the stimulus in the third experiment and essentially the same except for a few variations in the second paragraph in the fourth experiment. Instructions were presented in written form. A further effort was made to provide neither too much nor too little information in the description of the category. For the first two experiments, the description of the to-be-learned category was “a certain type of sound at the beginning of the word”.³ For the third and fourth experiment, the description was “a certain type of consonantal sound somewhere in the word”. The explicit reference to sound was very important in the description of the category. If the word “sound” (or any similar expressions) had not been used and more general information had been provided for any of the categories under investigation (e.g. “some of the words have something in common”), subjects would probably have pursued similarities of meaning, spelling, stress, grammar or whatever notion might have happened to strike them first. However, by mentioning “sound”, subjects’ attention was believed to be exclusively directed to pronunciation or phonetic aspects. In addition, if more specific information had been provided for, say, experiment two (e.g. “a hard or “explosive” sound”, etc.), subjects could have based their responses on the experimenter’s implied category rather than on their own intuitive classifications. In experiments 3 and 4 it was decided that the adjective “consonantal” should precede sound because the consonantal sounds that exemplified the category could appear in any position in words and this might distract subjects to think that the sound in question could be a kind of vowel. In other words, describing the target category as “a kind of sound” again would have been slightly more general information than deemed appropriate.

In addition, subjects read all the necessary instructions to perform both the learning and the test sessions at the beginning so that no break in continuity between both parts might cause any forgetting of the category and/or the criteria used to learn it. Subjects’ acceptable paraphrasing of the contents of the instructions and the task they had to perform guaranteed that they had understood the instructions well. Consequently, it was not considered to be necessary to provide subjects with instructional information between the learning and test sessions. Needless to say, subjects were not allowed to speak while performing the experiments. The instructions and the experimenter emphasised that subjects should concentrate on how the words that they heard were spoken, rather than how they were spelt.

Both the learning and the test sessions involved a standard CF reception procedure, the most common version of the concept formation experimental paradigm. As regards the learning

session, there were 32 positive and 28 negative instances in all experiments. Consequently, there were 60 trials altogether in each learning session. The proportion of positive and negative instances was almost equated in these experiments. Positive instances included a wide range of acceptable manifestations of the categories under investigation. Most of the negative tokens were non-interfering although some of them were interfering because of their having some potentially misleading phonetic and/or orthographic characteristic. **Only experiment 4 did not contain any interfering item.** In addition, as trials to criterion across several experiments had to be compared, the order of presentation of positive and negative tokens and feedback schedule was the same in the learning session as is the case with previous experiments with similar purposes (e.g. Derwing & Wang 1995; Jaeger 1980a). In addition, the interval between the beginning of the stimulus and the provision of feedback lasted 3 seconds and the postfeedback interval lasted 5 seconds in each experiment. A 1000-Hz tone of 100 msec duration appeared at the onset of each trial 1 second before the presentation of the stimulus (see figure 4 for a summary of the time intervals, experimental events, and actual time in a trial on the problem in the learning session).⁴ Finally, it should be mentioned that subjects were made to hear the entire learning session. This part of the experiment did not stop when the subject had reached criterion. In this way a large number of responses to all types of words would be available for evaluation.

Figure 4. Time Intervals, Experimental Events, and Time in One Trial on the Problem in the Learning Session.							
TIME INTERVAL		PRE-STIMULUS INTERVAL	STIMULUS INTERVAL	POST-STIMULUS INTERVAL			(NEXT) WARNING SIGNAL
				REACTION TIME + DELAY OF INFORMATIVE FEEDBACK		POST-FEEDBACK INTERVAL	
EVENT	WARNING SIGNAL tone	---	STIMULUS [k ^h ɪl]	SUBJECT'S RESPONSE Push green/red button on the keyboard	FEEDBACK Visual green/red rectangle	---	---
TIME	100 msec	1 sec	3 seconds		5 seconds	1 sec	

In the test session, there were 19 positive and 12 negative instances in all experiments except for experiment 4 in which there were 22 positive instances and 18 negative ones. These

³ This description is precisely the one Jaeger used for the phoneme / k / (Jaeger 1980a: 69, 1980b: 246) and one which she claimed would leave them free to define for themselves what constituted the notion of "a sound" and which aspect(s) of the sounds of the words occurs in all positive tokens.

⁴ Although initially it was thought that a timer would show the time between the end of feedback provision and the presentation of the following stimulus (i.e. the post-feedback interval), it was finally decided that this might distract subjects' attention. Consequently, as in Jaeger's (1980a, 1980b) experiments, the presentation of each stimulus word was preceded by a short tone approximately 1000 msec before each word.

instances included 3/4 control tokens. In addition, there were 9 test instances except in experiment 4 in which no test instances appeared. No feedback was provided at all for any kind of stimuli in the test session of the present CF experiments. Each trial consisted of 1000-Hz tone of 100 msec duration followed by a second of silence, the presentation of the stimulus and 5 seconds between the beginning of the stimulus and the occurrence of the next tone indicating the imminent presentation of the next stimulus (see figure 5 for a summary of the time intervals, experimental events, and actual time in a trial on the problem in the test session). The order of presentation of positive, negative and test tokens was the same across the four experiments.

Figure 5. Time Intervals, Experimental Events, and Actual Time in a Trial on the Problem in the Test Session.					
TIME INTERVALS		PRE-STIMULUS INTERVAL	STIMULUS INTERVAL	POST-STIMULUS INTERVAL (SUBJECT'S REACTION TIME+TIME LEFT FOR NEXT TRIAL)	
EVENT	WARNING SIGNAL	---	STIMULUS	SUBJECT'S RESPONSE	(NEXT) WARNING SIGNAL
	tone	---	[k ^h ɪl]	Push green/red button on the keyboard	tone
TIME	100 msec	1 sec	5 seconds		1 sec

Although in some CF experiments subjects who do not reach criterion are not transferred to the test session, all subjects in the present studies carried out that part of the experiment (as well as the post-experimental interview) irrespective of whether they had reached criterion or not.

In the post-experimental interview, subjects were asked different questions about their performance in the task including the name they would give to the category, the examples they remembered, the articulatory gestures used to pronounce the sound(s) in question, their subjective auditory impression of how the stimuli sounded like (e.g. harsh, soft, etc.), etc. Subjects had to write this information on a piece of paper. The experimenter also wrote comments on a separate piece of paper.

3.1.1.6. Data Collection and Interpretation.

The following sources of data were collected in this investigation for each of the four experiments: subjects' cumulative errors from the beginning of the CF in the learning session, their correct, incorrect and no answers to both positive and negative stimuli in both the learning

session and test session and their answers to test words. In addition, verbal data were obtained in the post-experimental interview. These data were analysed using the software programmes Microsoft Excel 97 and the statistical packages SPSS 10.0.7, and BMDP (1985 version) as well as the programme Minitab 13.31.

The criterion to decide whether a subject had attained the learning session any of the four different CF tasks was set at 37 correct responses out of the 60 experimental trials. This specific number was selected because it is almost impossible that a subject might reach that number responding at random ($P\text{-value: } 0.03 < 0.05$).

Although the programme used permitted recording subjects' reaction times automatically (and in fact it did), reaction times were not taken into account in this study for different reasons. In the first place, the aim of the present experiments was not to discover within-category differences processing differences between different realisations of the categories under investigation but simply to "chart", so to speak, the boundaries of those categories for the speakers who learned them. In the second place, as Jaeger also acknowledged in her (1980a) CF experiments (Jaeger 1980a: 361), the present experimental investigation was not ideally suited for looking at such within-category differences, which might probably exist. This is so because reaction times are dependent on many different factors like lexical frequency (e.g. Connine *et al.* 1993), the position of the target sound in the word, the phonetic context in which it is embedded, for example RTS to word-initial targets are somewhat longer than anywhere in the word (word-internal positions) (van Ooyen *et al.* 1992: 102), orthographic factors (e.g. Jaeger 1980a), etc. For example, in experiments 1 and 2 the position of the target sounds was the same (i.e. at the beginning of the word) but neither the lexical frequency of the word was controlled for nor the vocalic environment. Subjects, for example, might take longer to decide that a given word begins with a sound that belongs to a category simply because it takes longer for them to identify the word.

3.2. Experiment 1. The Category “Consonantal Sounds” in English.

3.2.1. Design: Rationale, Purpose of the Experiment, Research Questions, and Hypotheses.

The terms “vowel” and “consonant” have been in use for many centuries (Gimson 1980: 27) and they can ultimately be traced back to the Greek grammarians (Abercrombie 1967: 39; Trubetzkoy 1969: 223). These terms have been extensively used in phonetics and phonology for over a hundred and twenty years now but considerable controversy has surrounded their definition, characteristics, and their respective inventories.

In phonetics, vowels and consonants have been traditionally distinguished on the basis of different articulatory, acoustic, or phonetic contextual features.

According to many phoneticians (and phonologists as well), the most important articulatory characteristic distinguishing vowels and consonants is whether the air passage is obstructed or not. In the production of consonants, an obstruction results from a decided narrowing or a complete closure of the air passage at some point in the oral tract. On the contrary, the production of vowels requires no partial or complete obstruction as there is no contact or close approximation of the articulators (e.g. Abercrombie 1967: 55; Bloomfield 1979: 102; Christophersen 1956: 25, Gimson 1980: 28, 35; Gleason 1955: 23; Jakobson 1968: 100; Jones 1984: 12, 1989: 24; Ladefoged 1993: 6, 11-12; Ladefoged & Maddieson 1996: 281-282; Pike 1943: 67; Roach 1983: 3; Trubetzkoy 1969: 84; van Ooyen *et al.* 1992: 102). Thus, in the production of vowels the air issues in a continuous stream through the pharynx and mouth, it passes out quite freely in an unimpeded way over the centre line of the tongue, and it is never prevented or hindered from escaping.⁵

Friction and sonority are the most frequently mentioned acoustic features used to set consonants and vowels apart. Sounds in which some obstruction causes audible frictional noise are considered to be consonants while sounds in which no audible friction occurs are considered as vowels (e.g. Bloomfield 1979: 102; Jones 1984: 12, 1989: 23, 208; Pike 1943: 70; Trim & O'Connor 1953: 103). Another important acoustic criterion is the inherent sonority of sounds regardless of their surroundings, which is an acoustic (not a contextual) criterion. Pike (1943: 69-70) claims that vowels are sounds which are naturally more sonorous and resonant than the consonants. Jones (1989: 24) also uses the relative sonority or “carrying power” of sounds as a distinguishing factor between consonants and vowels. According to Jones, vowels are more

⁵Although the air passage is not obstructed for vowels, these can be varied due to the position of the lips and the tongue. By means of the tongue, the most important speech organ for the purpose of modifying vowels, we can form many different resonance chambers and so pronounce many different vowels.

sonorous than consonants when pronounced in a normal manner because they carry better or can be heard at a greater distance, when pronounced with the same length, stress, or voice-pitch.⁶

Finally, a few contextual criteria have been used to define these two classes of sounds. These criteria refer to the relationship of speech segments to neighbouring sounds in larger phonetic units like the syllable. For example, the two main views of the phonetic syllable, the “prominence” and the “pulse” theories, converge on what sounds should be classed as consonants and which ones as vowels. According to these theories, vowels are the most prominent element of each syllable and they carry the chest pulse while consonants are less prominent and either release or arrest the chest pulse.⁷

Another contextual phonetic criterion usually cited is the pronunciation of certain sounds at the end of unstressed syllables before vowels or consonants. In English, the pronunciation of the preposition “to”, the definite article “the”, and indefinite article “a/an” is usually mentioned as a contextual test to decide whether the following speech sound is phonetically a vowel or a consonant (e.g. Brown 1977: 50-51; Gimson 1980: 54, 212; Pike 1943: 74; Roach 1983: 50-51). The preposition “to”, [t^hu:] in citation form, is pronounced [tə] before consonants (e.g. [tə mi:]), but [tu] before vowels (e.g. [tu ə'laʊ]). Also, the definite article “the”, pronounced [ði:] in citation form, is rendered [ðə] whenever the next sound is a consonant (e.g. [ðə p^het]) but it is [ði] when followed by a vowel (e.g. [ði eɪk]). Similarly, the indefinite article is pronounced [ə] preceding consonants (e.g. [ə set]) but [ən], with an epenthetic nasal, before vowels (e.g. [ən ɑ:m]). Consequently, sounds that can be preceded by [tə], [ðə], or [ə] may be classed as consonants and those preceded by [tu], [ði], or [ən] as vowels.⁸

In phonology, where phonemes have traditionally been regarded as structural units that function in certain ways in the system in which they belong, irrespective of how they sound, phonological vowels and consonants are often distinguished on the basis of two different but related criteria. The first criterion has to do with the role of phonemes in the structural unit called phonological syllable and the second with their distribution. According to these two criteria, vowels are central in the syllables and function as their nuclei while consonants are marginal and function as syllable onsets and codas (e.g. Abercrombie 1967: 79; Gimson 1980: 51-52; Hockett 1955: 78; Ladefoged & Maddieson 1996: 282; O'Connor 1973: 199; van Ooyen

⁶ Another acoustic feature that might distinguish vowels from consonant is that vowels are characterised by well-defined two or three formant structure while the latter are typically not (Gimson 1980: 213; O'Connor 1973: 199). However, such an argument has been less often adduced to draw the line between vowels and consonants.

⁷ According to the prominence theory, vowels occupy the single peak of prominence that each syllable has while consonants, which have relatively weak prominence, occupy syllable margins or “valleys” (e.g. Bloch & Trager 1942: 22-24; Gimson 1980: 51-52; Jones 1989: 24-25, 55; Kreidler 1989: 26-28; Pike 1947: 60). According to the pulse theory, which is concerned with the muscular activity controlling lung movement during speech, the stream of air expelled by the chest-pulse can be released by accessory articulatory movements and arrested or cut down by other articulatory movements, bringing the syllable to an end. Thus, consonantal sounds act typically as the onset or “releasing” factor and the closure or “arresting” factor while vowels are nuclear to the syllable and render the chest pulse audible (e.g. Abercrombie 1967: 39-41; Gimson 1980: 52; Pike 1947: 60).

⁸ Although not usually mentioned, the auxiliary “do” in interrogative questions follows a similar pattern to that of the preposition “to” (i.e. [du:] in citation form, [də] before consonants, and [dʊ] before vowels) and it could also be added to the list (see, for example Monroy-Casas 1998: 44)



et al. 1992: 102). Also, vowels and consonants are two groups of sounds that have widely different distributions. According to this second criterion, the syllable is again the best structural unit that most economically expresses the combinatory latitudes of vowels and consonants within a given language. Vowels occur in syllable nuclei and are preceded and followed by a consonant unit or permitted consonant combination (e.g. O'Connor 1973: 199; Pike 1947: 60, 235, 254; Roach 1983: 4; Trim & O'Connor 1953: 122).

In short, from a purely phonetic standpoint, we can think of consonants and vowels as two groups of sounds that can be distinguished by a given set of articulatory, auditory or phonetic contextual features. In phonological terms, vowels and consonants may be conceived of as phonological units or groups of units that function in different ways in the linguistic system and have widely different distributions.⁹

At this point, it should be mentioned that irrespective of the criterion or criteria chosen, there is never an exact correspondence between the phonetic and phonological lists of vowels and consonants in a given language (Abercrombie 1967: 79; Gimson 1980: 29; Gleason 1955: 340; Trim & O'Connor 1953: 104; van Ooijen *et al.* 1992: 102). In other words, the inventories of phonetic and/or phonological vowels (and consonants) generally coincide but the overlapping is never complete. This is not surprising since, after all, phonology and phonetics have typically pursued different goals. However, a more serious problem is that even in the same domain (either phonetics or phonology), theorists usually disagree on the status of certain segments or units. This is so because no matter which criteria are adopted for defining vowels and consonants (in the phonetic or phonological sense) and how exhaustive these criteria are believed to be, the latter often turn out to be unsatisfactory because there seems to be a number of dubious cases that do not neatly catalogue themselves as they seem to comply with the previously-defined characteristics of both consonants and vowels. Such cases, which are usually grouped into ad-hoc categories like semi-vowel, semi-consonant, etc.,¹⁰ question the widespread assumption that a clear-cut vowel-consonant delineation with every segment or unit belonging in one or the other category exists (e.g. Jones 1989: 23). As such assumption is the norm, rather than the exception, the controversial cases are usually parcelled out between the obvious classes on the grounds of their similarity (but never complete match with the features of typical examples of the categories "consonants" or "vowels").

Three of these controversial segments in General British pronunciation are the so-called aspirate /h/ and the so-called glides or approximants /j/ and /w/.

⁹ For a number of answers to the question of how vowels and consonants can be distinguished on the web visit <http://www.linguistlist.org/~ask-ling/archive-most-recent/msg01429.html>.

¹⁰ The emergence of a wide variety of phonetic terms for designating these controversial items testifies to the difficulty of drawing a sharp line between both categories. Amongst others, the following terms have been used: demi-vowel, demi-consonant, vocalic consonant, consonantal vowel, consonantoid, vocaloid, covowel, etc.

/ h /¹¹ is articulated with the same tongue position and lip shape as the vowel that follows. For instance, during the [h] in “he”, “hut”, “hard”, all the speech organs except the glottis (i.e. lips, tongue, soft palate, etc.) are in the position for [i:], [ʌ], and [a:] respectively. In fact, / h / is simply a strong voiceless version or equivalent of whatever vowel comes next. It might then be actually more accurate to transcribe words like “hard” as [qɑ:d], although [h] is sufficient indication (Jones 1989: 201). This state of affairs implies that the realisations of / h / do not generally have a close enough tongue position or glottis to produce the kind of local friction typical of fricatives. However, the realisations of / h / occasionally imply some local friction, as when / h / is followed by a sound that needs a close front or palatal tongue position, as in “heed” / hi:d / or “huge” / hju: /, in which case a voiceless palatal fricative [ç] (e.g. “huge” [çu:dʒ]) is often heard. In addition, some audible friction may be produced in the larynx when / h / is pronounced with strong exhaling-force, the glottis being narrowed and the vocal cords brought close together, but not so close that they are set in vibration. The result is a whispered vowel, another possible pronunciation of / h /. Finally, / h / can also occur with glottal friction but voiced, especially when it appears between voiced sound or vowels, as in “alcohol”, “behind”, etc. In these cases, the whisper position of the glottis is combined with that for voice. The result is a voiced glottal fricative sound [ɦ].¹²

/ j / is usually realised as a voiced palatal.¹³ In its production, the vocal cords vibrate while the front of the tongue is raised towards the hard palate to almost the same position as for [i:] (see figure 6), although the approximation of the front of the tongue to the hard palate is not so close so as to cause local friction. In addition, the lips are generally neutral/spread (unrounded) but may anticipate the lip-rounding of a following vowel. Also, / j / entails a rapid vocalic glide from the initial palatal approximation to the position of the tongue required by whatever vowel comes next, which is generally more open than / j /, except when [j] is followed by [i:], even in which case / j / rarely involves local friction. However, in some circumstances there may be some local friction, as when / j / is followed by a rather close variety of / i: /, as in “yeast”, when / j / is preceded by / h /, as in “hue”, “human”, “huge”, etc., in which case a voiceless palatal fricative (i.e. [ç]) is frequently heard, or after accented / p t k /, as in “pew”, “tune”, or “cute”. In these latter cases / j / is not only fricative but also partly or wholly voiceless.

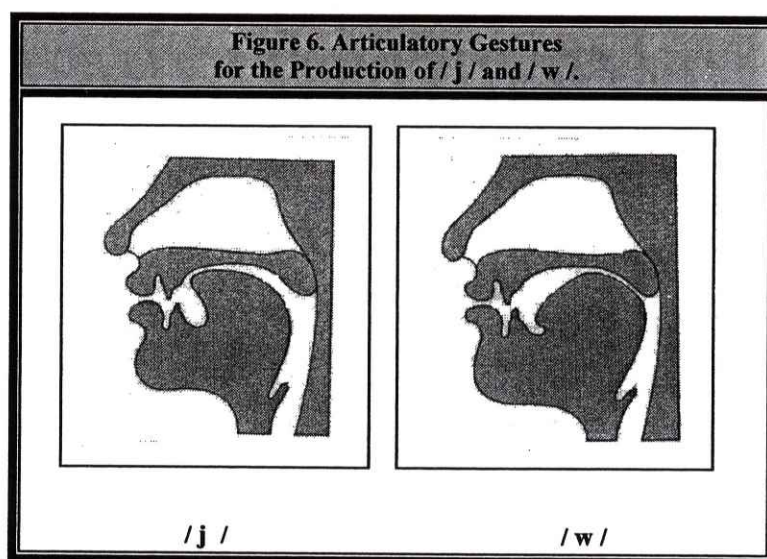
¹¹ For the details given below see, amongst others, Abercrombie (1967: 59), Brown (1977: 25), Christophersen (1956: 132), Gimson (1980: 191), Gleason (1955: 21), Hockett (1955: 33), Jones (1984: 115-117, 1989: 201-202, 210-211), Kreidler (1989: 68), Ladefoged (1993: 62), MacCarthy (1965: 140), O'Connor (1973: 143-144), Pike (1943: 71-72), and Roach (1983: 40-42).

¹² As in the case of the voiceless glottal fricative, this is done in a breathy way, allowing more air to escape through them than in normal voice.

¹³ For the details given below about / j / and / w / see, amongst others, Bloch and Trager (1942: 23-24), Brown (1977: 23-25), Christophersen (1956: 128, 130), Gimson (1980: 212-215), Jones (1984: 117-118, 1989: 25, 206, 209-210), Kleidler (1989: 26-28), Ladefoged (1993: 61), O'Connor (1973: 68, 148, 150-151), Pike (1943: 72), Roach (1983: 43-44, 51), and Trager (1942: 211).



/ w / can be described as a voiced labio-velar. In its production, the vocal cords normally vibrate while the back of the tongue is raised towards the soft palate to a close position, as in [u:] (see figure 6), although not so much so as to produce local friction. The lips are more or less rounded and contracted round a small central opening (degree of lip-rounding and contraction depending on the following consonant). Also, / w / involves, like / j /, a rapid vocalic glide from a close or half-close back rounded position to the position of whatever vowel follows which has a steadier duration and is usually more open, except for the sequence [wu:] (e.g. “woo”). However, even in this case, / w / involves no (or very little) local friction. Generally frictionless, when / w / is preceded by accented / t, k / as in “twice”, “quick”, the devoicing is complete and the resulting segment is a voiceless labio-velar fricative (i.e. [w̥] or [ʍ]).¹⁴ In these latter cases, a high vowel position rather close to the roof of the mouth and the voiceless air-stream is strong enough to cause local friction.



The phonetic descriptions of the [h], [j], and [w] clearly indicate that they have some characteristics similar to vowels. First, / h / is a voiceless version of the following vowel with the same tongue position and lip shape as the vowel that follows it.¹⁵ Also, except for one feature (i.e. lack of prominence in a syllable), [j] and [w] share most of the characteristics of

¹⁴ Some speakers of English occasionally pronounce / h / before / w /, particularly when they are speaking clearly or emphatically (e.g. “which” [hʍɪtʃ], “whether” [hʍeðə]) in which case [w] is usually devoiced. On these occasions, phonological contrasts are established between words like “which” and “witch”, “whether” and “weather”, etc. Sometimes, a voiceless labio-velar fricative [ʍ] may be heard instead of [hʍ]. These alternative pronunciations of [w] are rapidly disappearing in RP and most speakers have / w / in such words (see Gimson 1980: 217; Jones 1984: 118, 1989: 208; Ladefoged 1993: 62; Roach 1983: 41-42; Wells 1982b: 285).

¹⁵ As Abercrombie says (1967: 58), vowels are usually thought of as being essentially voiced. In fact, the etymology of the word (ultimately from the Latin vox “voice”) might lead one to believe that this must be so. However, although a vowel may be taken to be voiced if nothing is said to the contrary, vowels also occur voiceless, whispered, etc. (Hockett 1955: 33). Presumably the criterion of sonority is partly responsible for the entrance of voicing into vowel definitions (e.g. O’Connor 1973: 199; Trim & O’Connor 1953: 103; van Ooyen *et al.* 1992: 102). Conversely, the criterion of friction has been partly responsible for the elimination of voiceless or whispered vowels (e.g. Jones 1989: 201) from those definitions (Pike 1943: 72).

typical vowels (e.g. no decided obstruction of the air passage, free passage of the air through the centre of the oral tract, no friction, voiced, etc.). In fact, [j] and [w] have usually been described with the parameters used for vowels. So, depending on the openness of the following sound, [j] is said to be articulated with a half-close to close front vowel tongue position and [w] with a half-close to close back vowel tongue position. For these reasons, most phoneticians consider them as more vowel-like than consonantal. The name approximants, sometimes given to these speech sounds (e.g. Ladefoged 1993: 10, 60-61, 280) refers to an articulation in which one articulatory is close to another but without the tract being narrowed to such an extent that a turbulent airstream is produced.

While in phonetics the speech segments [h], [j], and [w] are usually classified as vowels, in phonology, the phonemes / h /, / j /, and / w /¹⁶ are generally classified as phonological consonants for three different reasons that, have to do with the function, distribution, and patterning of those phonemes in the linguistic system. First, / h /, / j /, and / w / are classed together with the consonants because they behave as marginal units in the syllable, like any other consonantal phoneme. More specifically, / h /, / j /, and / w / occur in the onset position of the syllable either initially or in an initial cluster preceding a syllabic sound (they cannot occur, however, in the coda position in syllables). They can be described as non-syllabic in the phonological sense (Abercrombie 1967: 79; Gimson 1980: 31, 53, 212; Gleason 1955: 35; Trager 1942: 221-222).

Second, / h /, / j /, and / w / seem to be in the same distributional group as the consonants according to the pronunciation of the preposition “to”, the definite article, and the indefinite articles before words starting with those segments in the same tone-unit. The pronunciation of those function words is the same when they precede [h], [j], or [w] ([l] and [r] as well) as it is when they precede any other clear member of the category consonant in phonological terms like / p /, / s /, etc. (e.g. Gimson 1980: 44; Kaye 1989: 24; Roach 1983: 50-51).

Finally, Gimson (1980: 94-95) gives another reason why / j / and / w / (the argument could also be applied to [h]) should be considered as consonants in the phonological sense. According to Gimson, [j] and [w] regularly occur in positions preceding most of the vowel phonemes (see table 1). Given this and, since [j] and [w] are often purely vocalic from a phonetic standpoint, it is possible to consider their combination with other vowels as constituting rising diphthongs when [j] and [w] are followed by monophthongal vocalic

¹⁶ Trager (1942) suggests that, despite the fact that [j] and [w] are phonetically similar to close front and close back vowels and they seem to be merely the non-syllabic varieties of these vowels and to be in complementary distribution with them, they are not because they contrast with high vowels directly as in “woos” vs. “ooze”, “yeast” vs. “east”, “oof” vs. “wool”, “if” vs. “Yiddish”. Phonological contrast makes [j] and [w] acquire the status of phonemes. The same reasoning may be applied to [h] (e.g. “art” vs. “heart”, “hit” vs. “it”, etc.).



segments (see table 1, columns A. & C.),¹⁷ or as triphthongs when [j] and [w] are followed by a vocalic glide (table 1, columns B. & D.). Nevertheless, Gimson argues, if such combinations were treated as phonological diphthongs or triphthongs, the inventory of basic vowel phonemes would increase enormously, something that most phonologists might presumably reject. Therefore, the principle that we might call “pattern economy” in the linguistic system seems to justify considering / j / and / w / as consonants occurring in syllable onsets rather than as complex vocalic units occurring in syllabic nuclei.¹⁸

Table 1. Phonetic Sequences of [j] and [w] + Monophthongs and Diphthongs in RP.							
A. [j] + MONOPHTHONG		B. [j] + DIPHTHONG		C. [w] + MONOPHTHONG		D. [w] + DIPHTHONG	
[j + ɪ]	“Yiddish”	[j + eɪ]	“yea”	[w + ɪ]	“wit”	[w + eɪ]	“way”
[j + e]	“yet”	[j + aɪ]	---	[w + e]	“wet”	[w + aɪ]	“wide”
[j + æ]	“yap”	[j + ɔɪ]	“yoick”	[w + æ]	“wax”	[w + ɔɪ]	“quoit”
[j + ɒ]	“yacht”	[j + əʊ]	“yeoman”	[w + ɒ]	“watch”	[w + əʊ]	“woe”
[j + ʊ]	“you” ¹⁹	[j + aʊ]	“yowl”	[w + ʊ]	“wood”	[w + aʊ]	“wound”
[j + ʌ]	“young”	[j + ɪə]	“year”	[w + ʌ]	“won”	[w + ɪə]	“weir”
[j + ə]	“failure”	[j + eə]	“Yare”	[w + ə]	“were” ²⁰	[w + eə]	“wear”
[j + i:]	“yeast”	[j + uə]	“you’re”	[w + i:]	“week”	[w + uə]	
[j + u:]	“youth”			[w + u:]	“woo”		
[j + a:]	“yard”			[w + a:]	“waft”		
[j + ɔ:]	“yawn”			[w + ɔ:]	“wall”		
[j + ɜ:]	“yearn”			[w + ɜ:]	“word”		

To briefly summarise, the phonetic characteristics of [h], [j], and [w] are more similar to those of vowels but the function and distribution of the phonemes / h /, / j /, and / w / places those phonemes within the class of the phonological consonants.

The long-standing question of the phonetic and phonological definition of vowels and consonants, the drawing line between both categories, and the category status of controversial segments or units presents itself as an attractive case for the experimentalist interested in the way that language users actually conceive of the categories “vowels” and “consonants” and in reflecting that conception in phonological analyses. If these categories are actually psycholinguistically relevant to speakers, it might be interesting to find out about how speakers conceptualise them and how they classify some of the traditionally controversial segments. With this idea in mind, a CF experiment was designed to teach subjects the category “consonantal sounds” (or “consonantal sound” if a more abstract concept is preferred), exemplified by uncontroversial examples of consonants in either a phonetic or phonological definition of the term as positive instances. The counterparts of the former stimuli (i.e. the negative ones) were

¹⁷ In these cases, the second element rather than the first is the more prominent, although the reverse is the usual pattern in English diphthongs.

¹⁸ Gimson does not give this phenomenon a specific name.

¹⁹ Weakly stressed.

²⁰ Weakly stressed.

uncontroversial examples of the category “vowel” in (again) either a phonetic or phonological sense

The purpose of the experiment was then to provide evidence on the ability of subjects to classify different members of the category “consonantal sounds” and thus assess the truth in the statement made by van Ooijen and his co-workers (1992: 102) that “language users usually have some awareness of the distinction” between vowels and consonants. The purpose was also to analyse how subjects classify and treat the three controversial segments [h], [w], and [j]. The specific research questions that the present experiment addressed were:

- 1) Can native speakers of English form the category “consonantal sounds”?
- 2) If so, how do speakers classify [h], [j], and [w] word-initially and followed by vowels?²¹

By consonantal sounds it is to be understood the group of speech sounds that, according to the phonetic and phonological criteria mentioned above, qualify as clear uncontroversial examples of consonants in either the phonological or phonetic senses. The same applies, *mutatis mutandis*, to the (contrastive) category “vocalic sounds”.

The first hypothesis entertained was that subjects would learn the category “consonantal sounds” because we believed that the articulatory, acoustic (and perhaps contextual) differences between vowels and consonants may be considerable enough for subjects to be aware of these two categories of speech sounds, which early induced awareness through orthography in early childhood may also promote. In addition, there is some experimental evidence that the distinction between vowels and consonants is somehow important for speakers at least at the perceptual level.²² The second hypothesis was the subjects would consider [h], [w], and [j] as members of the category “consonantal sounds”. The second hypothesis is based on the fact that, although subjects no doubt pay attention to the phonetic structure of sounds when they classify sounds, when speakers have the choice to assign speech sounds to categories on either distributional or phonetic criteria, distributional factors usually override phonetic ones (e.g. Ohala 1983).

²¹ It was convenient to restrict our investigation to the initial position in words. The question of whether the second element of diphthongs like / əu au ai ei ɔɪ /, as in “go”, “how”, “high”, “hey”, “boy”, vocalic glides in phonetic terms, is a different issue. Although some researchers claim that pre-vocalic and post-vocalic [j w] do not imply a difference in their function in the language (e.g. Gleason 1955: 38), most phonologists claim that it is questionable whether / j w / should be considered as separate, final contrastive units (Bloch & Trager 1942: 23; Gimson 1979: 45, 95, 212; Trubetzkoy 1969: 50; Wells 1982a: 49-50) instead, and that [j w] are not admitted word-finally as / j w / but are instead part of the phonological diphthong.

²² There is already some evidence obtained with the phoneme monitoring experimental technique on the existence possible different perceptual functions for the two types of sounds. It has been shown that vowels produce longer reaction times than consonants, suggesting that they are harder to perceive (Cutler *et al.* 1990; Mehler *et al.* 1981; van Ooyen *et al.* 1991, 1992) and vowels are detected faster than semivowels. Position in the syllable is ruled out as the factor producing the difference between vowels and consonants. In addition, studies of spontaneous slips of the ear show that consonants are misperceived more often than vowels (Bond & Garnes 1980). In particular, vowels in stressed syllables seem to be accurately perceived.



3.2.2. Method.

3.2.2.1. Subjects.

Twenty subjects between the ages of 18 and 37 (mean age 25 yrs) and selected randomly from the group of 80 recruited for the CF experiments reported in this work took part in this experiment. There were 8 men and 12. All of them were linguistically naive. The subjects had the characteristics mentioned in 3.1.1.2.

3.2.2.2. Stimuli.

The stimuli for the present experiment consisted of 100 monosyllabic English words. These had a consonant-vowel-consonant (CVC) canonical syllable structure in the case of positive instances and a vowel-consonant (VC) syllable structure in the case of negative instances. Consequently, the target sounds that subjects had to detect (i.e. consonantal sounds) were at the beginning of each word. The stimuli had a mean duration of 597 msc.

In the learning session, the positive instances (32 items in total) included 16 different types of consonantal phonemes in word-initial position (2 instances per phoneme). The consonants included examples of oral plosives, which involve a complete obstruction of the air passage accompanied by a velic closure (/ p b t d k g /), fricatives, which involve a decided narrowing of the air passage (/ s z f v θ ð /), affricates, characterised by a complete obstruction followed by an affricative release (/ tʃ dʒ /), and nasals, with simultaneous oral closure and velic opening (/ m n /).²³ Negative instances (28 items in total) contained a variety of different vocalic phonemes including monophthongs (/ ʌ ɑ: u: ɜ: /, 1 instance each; / ɪ æ /, 2 instances each; and / i: e ɒ /, 3 instances each), diphthongs (/ ɔɪ aɪ /, 1 instance each; / aʊ /, 2 instances; and / eɪ /, 6 instances), and triphthongs (/ aʊə /, 1 instance). There were 5 phonologically short monophthongs (/ ʌ ɪ æ e ɒ /). These had a relative steady quality. There were 6 types of vowels with changing quality: 5 diphthongs (1 closing-backing: / aʊ /; 3 closing-fronting / ɔɪ /, / aɪ /, & / eɪ /) and 1 centring triphthong (/ aʊə /). All negative stimuli were non-interfering except for one, the word “hours”. As the initial letter (i.e. “h”) of this particular word is typically referred to as a consonant letter but it is silent, a possible orthographic interference might be produced

²³ No monosyllabic words including word-initial / ʒ / and / ŋ / were included as these consonantal phonemes constitute phonological segmental constraints in English at the beginning of words.

and so, the word was considered as a potentially interfering one because of its conventional orthographic representation.²⁴

In the test session, positive instances (19 items in total) included all the consonantal phonemes already encountered in the learning session (1 example per phoneme). In addition, 3 stimuli included the voiceless palato-alveolar fricative / ʃ / in initial position, a consonant not previously encountered that served as a control. Negative instances (12 items in total) included examples of the vocalic phonemes presented in the learning session (/ ɪ ʌ ɑ: ɒ /, 1 instance each; / ɜ: eɪ /, 2 instances each) as well as 2 vocalic phonemes not previously encountered (i.e. / ɔ: / & / əʊ /, 2 instances each), which were negative controls. The test tokens (9 items in total) included three items starting with [w] (“wit”, “wall”, “once”)²⁵, three beginning with initial [j] (“youth”, “yet”, “use”), and three items with initial [h] (“hot”, “heat”, “hard”).

The positive stimuli of these experiments were words spelled with one letter in most cases except for those beginning with / θ /, / ð /, / ʃ /, and / tʃ / (spelled “th”, “th”, “sh”, and “ch” respectively). The negative stimuli were spelled with single “vowel letters” except for a few digraphs (e.g. “ea” = / i: /, “ai” = / eɪ /, etc.). The test stimuli starting with [h] were spelled with the letter “h”, the most common spelling of such sound in initial position.²⁶ The test items starting with [j] were spelled with “y”, its most common word-initial spelling, in the case of “youth” and “yet”, and with “u”, another (less frequent) spelling, as in “use”.²⁷ The spelling of [w] was with the letter “w”, in “wit” and “wall”, the ordinary spelling of the sound,²⁸ and by “o”, a less frequent spelling, in the case of “once”.

Tables 2 offers a summary of the types of stimuli used and table 3 the actual list of items and their category status.

²⁴ It was difficult to find any other orthographically interfering item spelled with “h” or any other interfering letter. The only other example found, “heirs”, was discarded as it was considered that subjects might understand the word “airs” on hearing [eəz] and the orthographic interference might disappear.

²⁵ The word originally selected was “one”. However, since its pronunciation could evoke the word “won”, “once” was finally chosen despite having a syllable structure dissimilar to that of any other stimulus (i.e. [w]VCC) with a two-consonant cluster in the coda.

²⁶ / h / is also the phonological value of the digraph “wh” in words like “who”, “whose”, “whom”, “whole”, “whore”, and their derivatives. Also, the letter “h” is silent in “hour”, “heir”, “honour”, “honest”, and their derivatives. It is also often silent in unstressed syllables in pronouns (e.g. “him”, “her”), notably in names ending in “-ham”, such as “Balham”, “Buckingham”, and in a few words like “vehicle” or “annihilate”.

²⁷ The spelling “y” also appears in medial positions (e.g. “vineyard”). In addition, the letters “i” and “e” often have the value [j] when the following sound is [ə] (e.g. “onion”, “simultaneous”). In words spelt with “u”, “ue”, “ui”, “ew”, or “eu” and representing / u: /, / j / is sometimes inserted before / u: / (e.g. “uniform”, “few”) and sometimes not (e.g. “rule”, “chew”, etc.).

²⁸ This is so even when “w” is preceded by a consonant (e.g. “twelve”, “dwindle”). However, when “w” is followed by “r” (e.g. “wrong”, “wrist”, etc.), it is silent. Also, “u” is generally pronounced [w] when preceded by “q” (e.g. “quite”, “quick”, “equal”, etc.) and often when preceded by “g” in unstressed syllables (e.g. “language”). The letter “o” is an exceptional spelling for [w] and it occurs in “one”, “once”, “choir”.



Table 2. Types of Stimuli and Number of Items per Subtype of Stimuli Used in Experiment 1.				
Type of Stimulus	LEARNING SESSION		TEST SESSION	
	Subtype of Stimulus	Number of Items	Subtype of Stimulus	Number of Items
POSITIVE ITEMS	Plosives / p b t d k g /	12	Plosives / p b t d k g /	6
	Fricatives / s z f v θ ð /	12	Fricatives / s z f v θ ð <u>f</u> / ²⁹	9
	Affricates / tʃ dʒ /	4	Affricates / tʃ dʒ /	2
	Nasals / m n /	4	Nasals / m n /	2
	TOTAL	(32)	TOTAL	(19)
NEGATIVE ITEMS	<i>Non-interfering:</i> Monophthongs / ʌ æ ɑ: u: i: ɒ ɪ e ɜ: /	17	<i>Non-interfering:</i> Monophthongs / ɪ ʌ ɑ: ɒ ɜ: <u>ɔ:</u> /	8
	Diphthongs / ɔɪ aɪ aʊ eɪ /	10	Diphthongs / eɪ <u>əʊ</u> /	4
	<i>Interfering:</i> Orthographically / aʊə / ("hour")	1		
	TOTAL	(28)	TOTAL	(12)
TEST ITEMS	----	---	[h] [j] [w] TOTAL	3 3 3 (9)

²⁹ The underlined phonemes in bold letters appearing in the tables represent control stimuli.

Table 3. Stimulus List for Experiment 1:
Category "Consonantal Sound(s)".

Order	Stimulus	Positive (+) Negative (-)	Order	Stimulus	Positive (+) Negative (-)
LEARNING SESSION					
1.	path	+	53.	change	+
2.	ash	-	54.	aim	-
3.	boom	+	55.	jug	+
4.	ache	-	56.	ebb	-
5.	toad	+	57.	ace	-
6.	duck	+	58.	mug	+
7.	kid	+	59.	itch	-
8.	up	-	60.	name	+
9.	give	+	TEST SESSION		
10.	eat	-	1.	pet	+
11.	seethe	+	2.	bathe	+
12.	zone	+	3.	tooth	+
13.	edge	-	4.	ode	- control
14.	fish	+	5.	deep	+
15.	van	+	6.	hot	test
16.	at	-	7.	eight	-
17.	egg	-	8.	ape	-
18.	ill	-	9.	cab	+
19.	thing	+	10.	guess	+
20.	off	-	11.	oath	- control
21.	that	+	12.	fetch	+
22.	each	-	13.	heat	test
23.	hours	-int	14.	us	-
24.	cheese	+	15.	vague	+
25.	job	+	16.	all	- control
26.	miss	+	17.	seed	+
27.	out	-	18.	youth	test
28.	neck	+	19.	ship	+ control
29.	eve	-	20.	wit	test
30.	on	-	21.	zoom	+
31.	pub	+	22.	earn	-
32.	aid	-	23.	thin	+
33.	beach	+	24.	then	+
34.	teach	+	25.	urge	-
35.	oil	-	26.	shock	+ control
36.	odd	-	27.	orb	- control
37.	dove	+	28.	once	test
38.	arm	-	29.	chase	+
39.	call	+	30.	yet	test
40.	goose	+	31.	of	-
41.	ale	-	32.	judge	+
42.	safe	+	33.	wall	test
43.	earth	-	34.	if	-
44.	zip	+	35.	hard	test
45.	fang	+	36.	shell	+ control
46.	owl	-	37.	map	+
47.	ooze	-	38.	net	+
48.	ice	-	39.	use	test
49.	vet	+	40.	art	-
50.	thick	+			
51.	age	-			
52.	those	+			

3.2.2.3. Procedure.

3.2.2.3.1. Instructions.

Subjects were given a sheet of instructions asking them to perform a CF task in which they had to focus their attention on the initial sound of each word they heard. The instructions told them that some words contained “a certain type of sound in the initial position of the word” that they had to identify while other words lacked that certain type of sound at the beginning of the word as described in 3.1.1.4 and 3.1.1.5. The words that contained the to-be-identified type of sound, subjects were told, were to be associated with the colour “green” while those that lacked that type of sound were associated with the colour “red” as shown in two rectangles on the screen of the computer that they would use to perform the task. Subjects were told that after hearing each word, they would be provided with an answer as to whether or not the word had included the to-be-identified type of sound in the initial position. If it included the type of sound in question, a red rectangle on the screen would disappear so that the presence of the green one would indicate that the sound had contained the to-be-identified type of sound. If, on the contrary, the green rectangle disappeared and the red one remained on the screen, this might indicate that the word did not include the sound. The instructions also told the subjects to begin responding (by pressing either a red or a green key on the keyboard) once they had some idea of what that “certain type of sound” was as soon as they heard each new word. Subsequent disappearance of either the red or the green rectangles on the screen would tell them whether the word had contained an instance of the to-be-identified type of sound. Subjects were also informed that after a certain amount of trials, feedback would be no longer provided. The actual instructions were as follows:

Please read these instructions carefully.

You are going to participate in an experiment that has been designed to study certain aspects about the sounds of English. For this experiment you are going to use the computer, keyboard, and headphones on this table.

You are going to be listening to a list of words over the headphones. Some of these words contain a certain type of sound in the initial position of the word, while other words do not. Your task is to pay attention to the words and try to figure out the type of sound some words have but others lack at their beginning.

At first, you will not know which type of sound we are talking about. Here is how you can figure it out. There are two rectangles, a green one and a red one, on the screen of the computer. There are also two keys or buttons, a green one and a red one, on the keyboard. Both the rectangles and the keys mean the same. Green means “the word contains the type of sound involved in its initial position”. Red means “the word does not contain the type of sound in its initial position”.

Whenever you hear a word, one of the two rectangles on the screen will remain on display while the other will disappear after a few seconds. When the word begins with the type of sound you have to figure out, the green rectangle will stay but the red one will disappear. When the word does not begin with the sound, the red rectangle will stay but the green one will disappear. Rectangles disappear automatically. (Every rectangle that disappears will eventually reappear just before a short tone which indicates that a new word is about to come).

However, we are not only interested in showing you which words contain or do not contain a certain type of sound by the disappearance of one of the two rectangles. We are also interested in your guesses as to whether the words do or do not contain the type of sound and how you show it. So, when you have some idea

of what the type of sound is, you should start pressing the keys on the keyboard according to what you think. You'll find out whether you are right or wrong by looking at the behaviour of the rectangles.

For example, if you hear a word and then press the green key on the keyboard, you'll be indicating that you think that the word contains the sound. Then, if the green rectangle stays while the red one disappears, you'll know you were right: the word did contain the sound as you had thought. On the contrary, if the red rectangle stays but the green one disappears, you'll know you were wrong: the word did *not* contain the sound.

Conversely, if you hear another word and then press the red key, you'll be indicating that you think that the word does not contain the type of sound. Then, if the green rectangle stays while the red one disappears, you'll know you were wrong: the word did contain the sound, unlike what you had thought. On the contrary, if the red rectangle stays while the green one disappears, you'll know you were right: the word did not contain the sound as your answer had originally indicated.

Because you should answer according to your first, immediate impression, there are only have 3 seconds between the end of the word and the moment in which either the green or red rectangles on the screen disappear. In other words, you only have three seconds to press either the green or the red keys on the keyboard to indicate your decisions.

It is important to remember that what you have been told to do is the main part of the experiment and it lasts around 8 minutes. This part has been designed to let you find out the type of sound some words share while others lack in the position indicated (i.e. word-initially). Your responses will not be counted in this part. So, you should not worry about making mistakes or trying out different ideas.

There is a second part in this experiment. It lasts 4 minutes. In this new part, which will start automatically after you finish the first part, you have to do exactly the same thing as in the first part. There is, however, a very important difference between both. In the second part the green and red rectangles will always be on display. No rectangle will disappear at any moment. So, you will not know whether you are answering correctly or incorrectly. This should be no problem as you got a lot of practice in the initial part and by now you may already know which type of sound we were talking about. Nevertheless, you should press either the green or the red keys on the keyboard to indicate, according to you, whether the words you hear do or do not contain the sound you tried to figure out in the first part.

In both parts you should use the forefinger of your dominant hand to press either the red or green keys on the keyboard.

Please do read these instructions again. At the end, ask any questions you have. It is extremely important that you understand the instructions perfectly.

Once subjects had finished reading the instructions, the experimenter approached them and asked whether they had understood the instructions. Subjects' paraphrasing of the sequences of events during the experimental task, and interactive discussion with the experimenter guaranteed that they had actually understood them. Next, the experimenter told the subjects to put on headphones. The learning session started when the experimenter had gone out of the subjects' sight as indicated in 3.1.1.5.

3.2.2.3.2. Training Session, Test Session, and Post-Experimental Interview.

The training and test sessions proceeded as described in 3.1.1.5. Subjects were run individually in a sound-treated room. After the computer-based task finished, a post-experimental interview was conducted. No incidences were registered.

3.2.3. Results.

3.2.3.1. Training and Test Sessions.

Of the 20 subjects who participated in this experiment 12 were considered to have formed the category correctly (6 men and 6 women), with a mean of 48.67 correct responses (range 37-59, s.d.= 6.50) in the learning session.

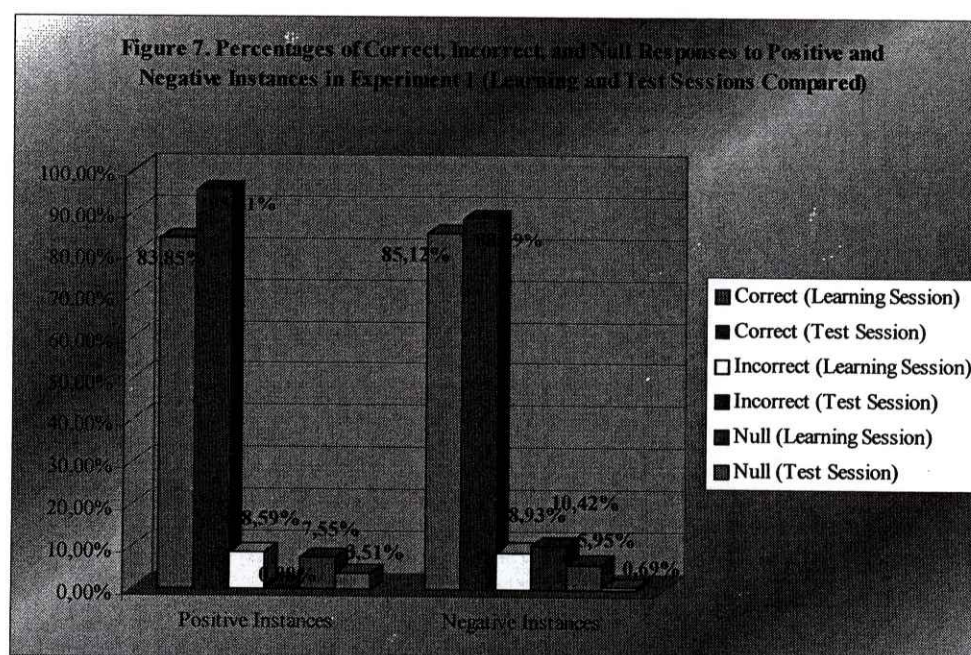
The numbers of correct (C), incorrect (I), and null responses (NR) to both positive and negative stimuli in the learning session are shown in Table 4. The table also shows the percentage of correct responses to each stimulus type, the number of items per type of stimulus, and the number of responses elicited, which results from multiplying the number of items by the number of subjects who the pre-established 7-correct-response criterion in the learning session. The same information obtained from subjects' performance in the test session is shown in Table 5.

Table 4. Number of Correct/Incorrect/Null Responses and Percent Correct Responses to Positive and Negative Stimuli in Experiment 1 (Learning Session).						
Stimulus Type	Type of Response			% Correct Responses	Items	Number of Responses Elicited
	C	I	NR			
Positive Instances	322	33	29	83.85%	32	384
Negative Instances	286	30	20	85.12%	28	336
TOTAL	608	63	49	84.44%	60	720

Table 5. Number of Correct/Incorrect/Null Responses and Percent Correct Responses to Positive and Negative Stimuli in Experiment 1 (Test Session).						
Stimulus Type	Type of Response			% Correct Responses	Items	Number of Responses Elicited
	C	I	NR			
Positive Instances	218	2	8	95.61%	19	228
Negative Instances	128	15	1	88.89%	12	144
TOTAL	346	17	9	93.01%	31	372

A close comparison of both tables shows that percentages of correct responses to both positive and negative stimuli substantially increase in the test session when compared with the percentages in the learning session. In other words, correct responses were significantly more frequent in the test session than in the learning session (84.44% versus 93.01%; $P\text{-value} = 0.000 < 0.05$ by a contrast of proportions). This increase in subjects' accuracy reveals that they were performing better in the test session than in the learning session, in which they were already doing fairly well. Subjects' better performance in the test session is also reflected in figure 7, in which the increase of the percentage of correct responses to both positive and negative stimuli, the decrease of the percentage of incorrect and null responses to positive stimuli in both sessions, and the decrease of null responses in the test session from the learning to the test

session can be observed. The only slightly different result from this trend is the percentage of incorrect responses to negative stimuli, which is higher in the test session than in the learning session. These facts suggested a possible relationship between type of stimulus and type of response. In this respect there was not a significant relationship in the learning session between type of stimulus (i.e. negative, positive) and type of response (correct, incorrect, null) ($P\text{-value} = 0.69 > 0.05$ by a test of independence or chi-square test). However, a significant relationship was obtained in the test session between type of stimulus and type of response ($P\text{-value} = 0.00 < 0.05$ by a test of independence or chi-square test). Focusing on the corrected residuals of the test, it became clear that the more positive stimuli subjects encountered, the fewer the errors they made, and the more negative stimuli encountered, the more errors they made.

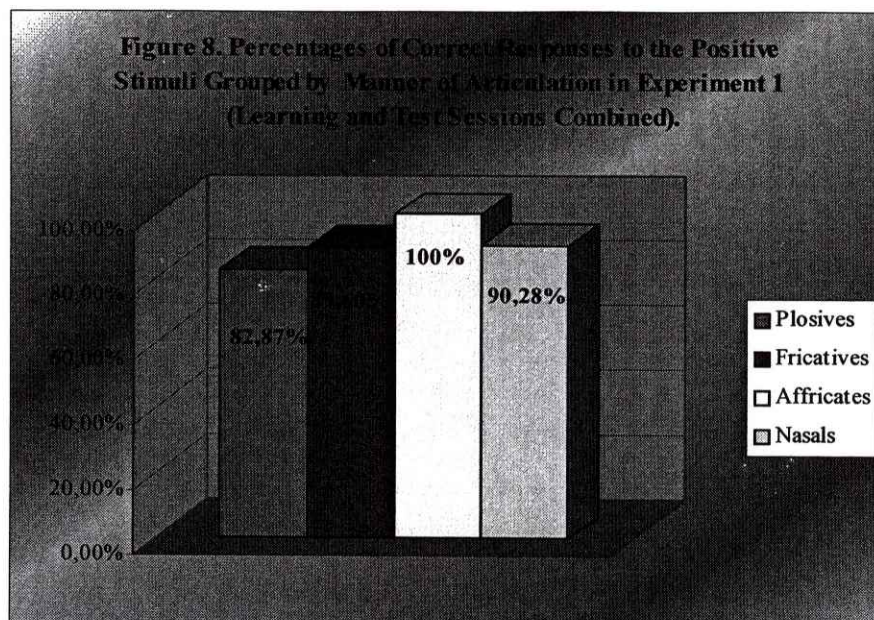


Collapsing the values of the variable “type of response” (i.e. correct, incorrect, and null response) into two categories (correct response vs. incorrect/null response), it was shown again that there was not a significant relationship between type of stimulus and type of response in the learning session ($P\text{-value}: 0.71 > 0.05$ by a test of independence using the statistic Yate’s corrected chi-square test). On the contrary, and focusing again on the corrected residuals, it was found out that, in the test session, the number of positive stimuli were again in inverse proportion to the number of incorrect/null responses and in direct proportion to the number of correct responses. On the contrary, negative stimuli were in inverse proportion to the number of correct responses and in direct proportion to the number of incorrect/null responses, all relationships being statistically significant ($P\text{-value}: 0.02 < 0.05$ by a test of independence using the statistic Yate’s corrected chi-square test). The fact that subjects’ incorrect (or incorrect/null)

responses are in inverse proportion to the number of negative stimuli in both the learning and the test sessions may have to do with the already demonstrated phenomenon of a poorer performance in a concept formation task when the number of negative instances increases (e.g. Hovland & Weiss 1953; see also section 2.3.1.2.). In other words, the more positive instances a concept learner encounters, the easier and faster the learning will be but negative instances hinder learning and it is not surprising that they are responsible for higher rates of incorrect responses.

Table 6 shows the numbers of correct, incorrect, and null responses to the positive stimuli of both the learning and test sessions (combined). The stimuli are divided into four groups of phonemes according to the manner of articulation that characterises the realisations of each phoneme in the experimental task (i.e. plosive, fricative, affricate, or nasal). The percentage of correct responses to each type of phoneme and each “manner-of-articulation” group, the number of items per type of phoneme or category defined by manner of articulation, and the number of responses elicited in each case are also indicated. The percentages of correct responses to the four different “manner-of-articulation” subtypes of positive stimuli in both the learning and the test sessions (combined) can also be seen in figure 8.

Table 6. Number of Correct/Incorrect/Null Responses and Percent Correct Responses to Positive Stimuli Grouped by Manner of Articulation and Type of Phoneme Type in Experiment 1 (Learning and Test Sessions Combined).							
Manner of Articulation	Phoneme	Type of Response			% Correct Responses	Items	Number of Responses Elicited
		C	I	NR			
Plosives	/p/	30	1	5	83.33%	3	36
	/b/	29	3	4	80.55%	3	36
	/t/	32	2	2	88.89%	3	36
	/d/	29	4	3	80.55%	3	36
	/k/	28	4	4	77.78%	3	36
	/g/	31	4	1	86.11%	3	36
SUBTOTAL		(179)	(18)	(19)	(82.87%)	(18)	(216)
Fricatives	/s/	30	3	3	83.33%	3	36
	/z/	30	1	5	83.33%	3	36
	/f/	31	3	2	86.11%	3	36
	/v/	31	3	2	86.11%	3	36
	/θ/	31	2	5	86.11%	3	36
	/ð/	35	1	0	97.22%	3	36
	/ʃ/	36	0	0	100%	3	36
SUBTOTAL		(224)	(13)	(15)	(89.6%)	(21)	(252)
Affricates	/tʃ/	36	0	0	100%	3	36
	/dʒ/	36	0	0	100%	3	36
SUBTOTAL		(72)	(0)	(0)	(100%)	(6)	(72)
Nasals	/m/	32	3	1	88.89%	3	36
	/n/	33	1	2	91.67%	3	36
SUBTOTAL		(65)	(4)	(3)	(90.28%)	(6)	(72)
TOTAL		540	35	37	88.24%	51	612



These latter percentages were compared with a one-tailed (type of stimulus) analysis of variance and it was found out that the differences between them were statistically significant ($F(3, 13) = 6.782$, $P\text{-value} = 0.005 < 0.05$). Post-hoc analyses using SNK and Duncan tests were conducted and they revealed that plosives, fricatives, and nasals were homogeneous subsets but that the homogeneity did not apply to the affricates, to which subjects made a significantly higher percentage of correct responses than to the other three subsets. The difference between the affricates and the homogeneous subgroup made up of the plosives, the fricatives, and the nasals may simply be due to the position of the affricates in the experimental task rather than to any inherent characteristic of the stimuli themselves. Thus, it should be emphasised that, although the stimuli were semi-randomly ordered, the four examples of affricates appeared in the positions 24, 25, 53, and 55, after each type of plosive and fricative phoneme had appeared once (in the case of stimuli 24 and 25) and after each type of plosive, fricative, and nasal had been presented once again (see section 3.2.2.2., table 2). The fact that subjects tended to make more errors at the beginning of the task as their criterion was still being formed and that the first 12 positive stimuli were plosives and fricatives (6 instances each) may have contributed to subjects' better performance on the affricates, which appeared later in the task. However, the fact that nasals, which always appeared after the affricates, are homogeneous with the plosives and fricatives and not with the affricates seems somehow puzzling. Perhaps nasals are less typical examples of consonantal sounds. It could be argued that just as typical vowels are presumably oral, the typical consonant is also oral and the nasality of / m / and / n / may have somehow influenced subjects' overall performance on this subset. Future experiments controlling for positional effects of consonantal sounds in tasks could be conducted to look at

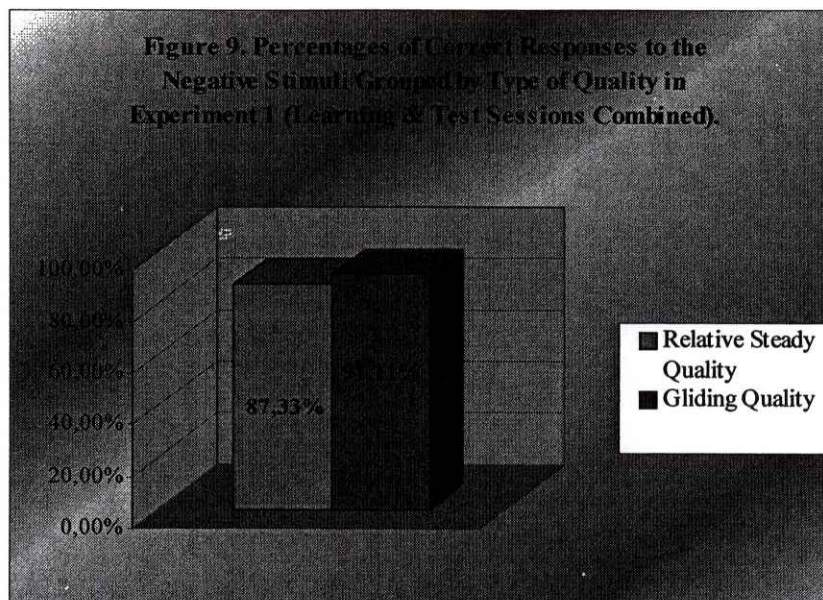
the typicality of different subsets of consonants or specific consonantal phonemes as members of the category.

In order to find out whether there were within-category differences between the different types of phonemes an analysis of variance was conducted on the percentages of correct responses to each consonantal phoneme. The results showed that the differences were significant ($F^*(16, 148) = 3.13$, $P\text{-value} = 0.0001 < 0.05$). Post-hoc analyses using SNK and Duncan tests were conducted and they revealed that all the plosives, the nasals, and the fricatives except / ð ʃ / were an homogeneous subset and that the fricatives / s z f v ð /, the plosives / p t k /, and the nasals / m n / made up another homogeneous subset (by the Duncan ,multiple range test) and all phonemes except / ʃ tʃ dʒ / on the one hand and all phonemes except / b k / on the other (by the SNK Multiple Range Test). These findings show that the affricates / tʃ dʒ / and the fricative / ʃ / were by far the phonemes that were significantly easier than any other. As / ʃ / represented as positive control in the test session, the significantly higher accuracy that it elicited indicates subjects' successful formation of the category because they generalised the classifying behaviour acquired in the learning session to that particular phoneme with the utmost accuracy.

While statistical analyses reveal significant differences between different types of consonantal sounds, the negative stimuli seemed to be more homogeneous than the positive ones. Table 7 shows the numbers of correct, incorrect, and null responses to the negative stimuli of both the learning and test sessions (combined). These stimuli were divided in this case into two categories or groups defined by the type of quality of the realisation of the phoneme in the experimental task, namely, with relatively steady quality or with changing quality. Like in table 6, the percentages of correct responses to each phoneme and per "vowel quality" group, the number of items per type of phoneme and per group, and the number of responses elicited are also indicated.

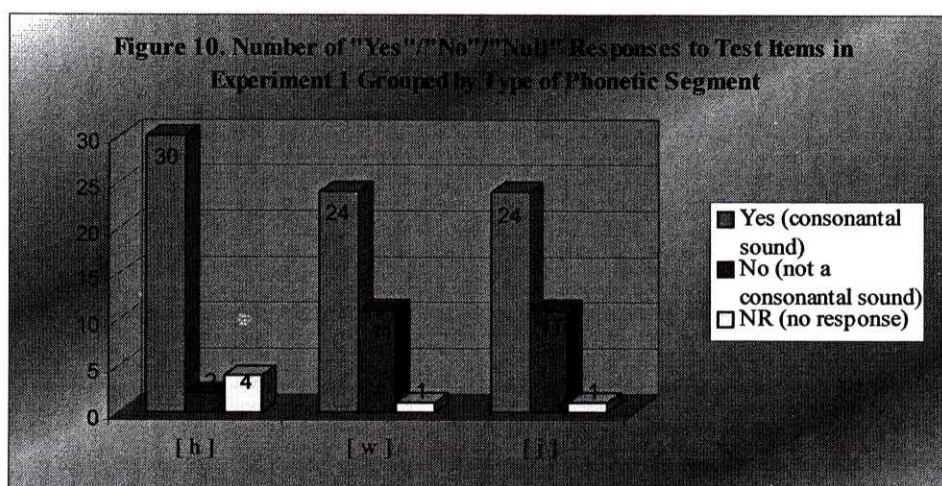
Table 7. Number of Correct/Incorrect/Null Responses and Percent Correct Responses to Negative Stimuli Grouped by Type of Quality and Type of Phoneme in Experiment 1 (Learning and Test Sessions Combined).							
Stimulus Type	Phoneme	Type of Response			% Correct Responses	Items	Number of Responses Elicited
		C	I	NR			
(Relative) Steady Quality	/ɪ/	33	3	0	91.67%	3	36
	/i:/	32	3	1	88.89%	3	36
	/u:/	11	1	0	91.67%	1	12
	/e/	30	4	2	83.33%	3	36
	/ɜ:/	33	2	1	91.67%	3	36
	/ɒ/	41	5	2	85.42%	4	48
	/æ/	19	3	2	79.17%	2	24
	/ʌ/	19	2	3	79.17%	2	24
	/ɑ:/	21	1	2	87.5%	2	24
	/ɔ:/	23	1	0	95.83%	2	24
SUBTOTAL		(262)	(25)	(13)	(87.33%)	(25)	300
Gliding Quality	/aɪ/	11	1	0	91.67%	1	12
	/ɔɪ/	11	1	0	91.67%	1	12
	/eɪ/	87	3	6	90.62%	8	96
	/əʊ/	24	0	0	100%	2	24
	/aʊ/	21	3	0	87.5%	2	24
	/aʊə/	10	2	0	83.33%	1	12
SUBTOTAL		(164)	(10)	(6)	(91.11%)	(15)	(180)
TOTAL		426	35	19	88.75%	40	480

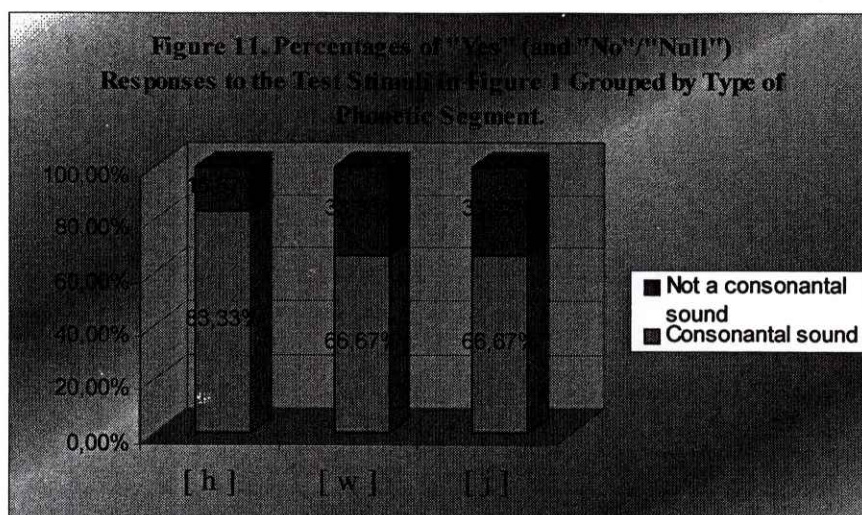
In order to find out whether there were within-category differences between the different types of phonemes an analysis of variance was conducted on the percentages of correct responses to each vocalic phoneme. The results showed this time that the differences were not significant ($F^*(15, 120) = 0.86$, $P\text{-value} = 0.6055 > 0.05$). This finding shows that the triphthong /aʊə/, spelled with “h”, was not more difficult than any other vowel and, consequently, that its potentially misleading spelling representation did not interfere in subjects’ formation of the category. Moreover, the results of a t test reveal that the differences between the percentages of correct responses to the two subsets of negative stimuli (i.e. with relative steady quality and with changing quality), which are also shown in figure 9, were not statistically significant ($t(14) = -1.298$, $P\text{-value}: 0.215 > 0.05$). This means that no group of negative stimuli as defined by type of quality (i.e. stable -monothphtongs- or changing -diphthongs and triphthongs-) was more difficult than the other.



Regarding next the test stimuli, the numbers of “yes” (target sound = consonantal sound), “no” (target sound = not a consonantal sound), and “null” responses to the test words grouped by phonetic type (i.e. [h], [j], and [w]) are shown in table 8 and figure 10. Table 8 shows, in addition, the percentages of “Yes” responses to each of the three different types of target sounds in the test stimuli, which can also be seen in figure 11.

Table 8. Number of Yes/No/Null Responses and Percent ‘Yes’ Responses to Different Test Items Grouped by Phonetic Type in Experiment 1.						
Stimulus Type	Type of Response			% Y Responses	Items	Number of Responses Elicited
	Y	N	NR			
/ h /	30	2	4	83.33%	3	36
/ j /	24	11	1	66.67%	3	36
/ w /	24	11	1	66.67%	3	36
TOTAL	78	24	6	77.22%	9	108



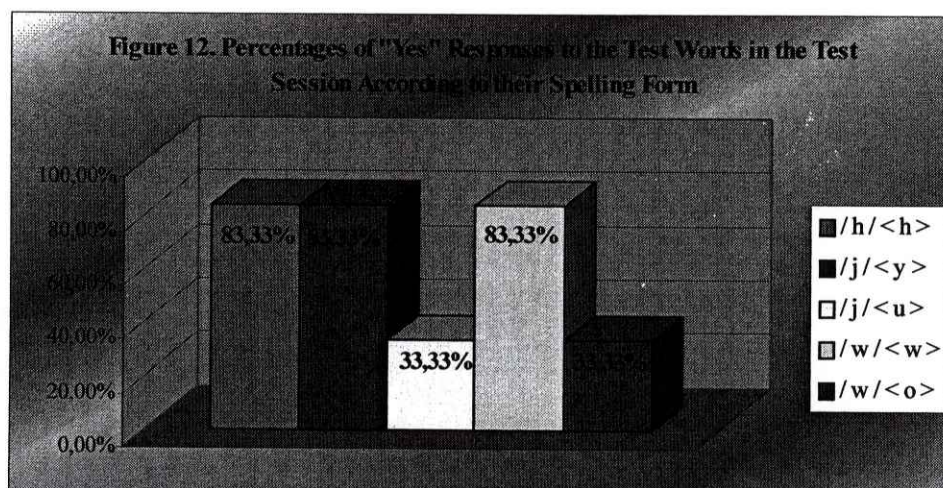


The data included in Table 8 and figures 10 and 11 reveal that out of the 36 responses elicited containing initial [h], subjects responded positively 30 times (83.33% of the time), negatively twice, and they provided no response on four occasions. Out of the 36 responses elicited for either [j] or [w], subjects responded positively 24 times (66.67% of the time), negatively eleven times, and provided no response once in each case.

Although the percentage of "Yes" responses to [j] and [w] (i.e. 66.67%) was lower than for [h] (i.e. 83.33%) but the differences were not statistically significant by a contrast of proportions ($P\text{-value} = 0.096 > 0.05$). These percentages indicate that subjects were including the test tokens in the category "consonantal sounds" most of the time. However, although these percentages leave [j] and [w] as members of the category "consonantal sounds" on pure percentage terms, a closer look at the test words beginning with either [j] or [w] reveals a curious phenomenon. When the conventional spelling of the target sounds is considered, that spelling seems to have had an impact on subjects' classifying behaviour. More specifically, as table 9 and figure 12 show, when [j] was spelled with the letter "y", typically referred to as a "consonant", out of 24 responses elicited, subjects responded positively 20 times (83.33% of the time), negatively 3 times, and, on one occasion, no answer was provided. Curiously, the same figures occurred for the words in which [w] is spelled with the *consonant letter* "w" (i.e. positively 20 times, negatively 3 times, 1 null response). However, when [j] was spelled with a *vowel letter* like "u", out of 12 responses elicited, subjects answered positively on 4 occasions (33.33% of the time) and negatively 8 times. The same pattern also applies to the test words beginning with [w] and spelled with a *vowel letter* (i.e. "o"). Comparing the percentage of "Yes" responses to /j/ spelled with "u" (i.e. 33.33%) with the percentage of "Yes" responses to /j/ spelled with "y" (i.e. 83.33%) the difference turns out to be statistically significant ($P\text{-value}$:

0.001 < 0.05 by a contrast of proportions). The same applies statistically significant difference applies to the contrast between the percentages of “Yes” responses to / w / spelled with “o” (i.e. 33.33%) and to / w / spelled with “u” (i.e. 83.33%).

Stimulus Type	Spelling	Type of Response			% Y Responses	Items	Number of Responses Elicited
		Y	N	NR			
/ h /	< h >	30	2	4	83.33%	3	36
/ j /	< y >	20	3	1	83.33%	2	24
	< u >	4	8	0	33.33%	1	12
/ w /	< w >	20	3	1	83.33%	2	24
	< o >	4	8	0	33.33%	1	12



Next, if we divide the test words containing [j] and [w] into two groups, namely, [j] and [w] spelled with “consonant letters” (i.e. “y” & “w”) and [j] and [w] spelled with “vowel letters” (i.e. “u” & “o”), the percentages of “Yes” responses for both groups are 83.33% and 33.33% respectively, as table 10 shows. A comparison of these percentages reveals that subjects significantly treated [j] and [w] as members of the category “consonantal sounds” more often when the [j] and [w] were spelled with letters typically referred to as “consonants” (i.e. “y” & “w”) than when the test sounds were spelled with “u” or “o”, the so-called “vowel” letters (83.33% versus 33.33%; P-value = 0.000 < 0.05 by a contrast of proportions). These findings seem to indicate that the orthography representation of [j] and [w] seems particularly important in subjects’ classification of those target sounds.

Table 10. Number of Yes/No/Null Responses and Percent 'Yes' Responses to Test Items [j] and [w] Grouped by Orthographic Spelling in Experiment 1.

Stimulus Type	Letters	Type of Response			% Y Responses	Items	Number of Responses Elicited
		Y	N	NR			
[j, w] Spelled with Consonant Letters	< y, w >	40	6	2	83.33%	4	48
[j, w] Spelled with Vowel Letters	< u, o >	8	16	0	33.33%	2	24

3.2.3.2. Post-Experimental Interview.

The individual interviews conducted after the test session of the CF task permitted the experimenter to obtain useful information about subjects' perception and conception of the category under investigation. This information allowed the experimenter to better understand the characteristics of the category "consonantal sounds" (and its opposing category, that is, "vocalic sounds") as it was conceived by the experimental subjects.

Table 11 contains the information provided by the experimental subjects in the post-test interviews. The table evinces that subjects provided very similar descriptions of the to-be-formed category. More specifically, subjects referred to the target sounds with synonymous expressions like "consonants" (2, AH; 3, CL; 6, RC), "consonant" (7, KR; 8, SH; 9, LH; 11, ES; 12, HM), "(more) consonantal" (1, KL; 4, KG), or "consonant sounds" (5, KP).³⁰ Subject number 10 (FS) was the only one who defined the category in a different way from the other subjects, namely, by specifying the negative category (i.e. "anything that's not a vowel"). In this respect, it should be noted that subjects' homogeneity in labelling the positive stimuli was paralleled by their similarity in labelling the negative stimuli, to which they referred using synonymous expressions like "vowel" (4, KG; 7, KR; 8, SH; 9, LH; 11, ES; 12, HM), "more vowel" (1, KL), "vowels" (2, AH; 3, CL; 6, RC), or "plain vowel sounds" (5, KP). This indicates, by the way, that the opposing category to the one under investigation was also a somehow salient conceptual category and not simply an ad-hoc category created circumstantially to place all non-positive stimuli within.

As table 11 also indicates, subjects also provided abundant information about the production of the members of the category "consonants". In general, subjects' reports indicate that the production of consonants was perceived as involving more movement of the articulators in the oral tract while the production of vocalic sounds required, according to them, less

³⁰ The numbers given in table 11 to the subjects do not necessarily indicate the order in which the latter were run in the experiment.

articulatory movement. For example, subject 1 (KL) said that the production of consonants would make subjects “move the mouth forcefully”; subject 7 (KR) claimed that “consonants involve more movement of the mouth”; subject 10 (FS) said that “vowels imply less movement of the lips”; subject 11 (ES) said that consonants required “more mouth” while vowels were “more open, more immediate”; finally, subject 12 (HM) mentioned that “the tongue moves...” in the production of consonants but for vowels “the tongue doesn’t move”. In addition, explicit references were made to important articulating organs (lips, tongue, teeth, and roof of the mouth) taking part in that “movement of the mouth” necessary for the production of consonantal sounds. For instance, subject 2 (AH) said that in the production of consonants “the lips or the tongue (are) involved”; subject 5 (KP) claimed that “the tongue, lips, (and) teeth (are) involved”; subject 9 (LH) said that “lips, tongue, teeth, alveolar ridge... change position, move” in order to produce different types of consonants. Finally, subject 10 (FS) referred to, as mentioned earlier, the “movement of the lips”. A few references were also made to the fact different types of consonantal sounds require diverse manners of articulation. For example, being a valid description for plosives, affricates, or nasals subject 2 (AH) said that, in order to produce (certain) consonantal sounds, one needs to “stop the air” and then “blow them out”; subject 12 (HM) said that for consonantal sounds to be produced “the mouth (is) more closed” while the mouth is “more open” in the production of vowels; subject 8 (SH) claimed that “the mouth is more narrow” for some consonants, which obviously applies to the fricatives; and subject 6 (RC) said that some consonants “have air in the nose”, an implicit reference to the nasals. The references to different types of consonants regarding their manner of articulation were more numerous when subjects described the perceived auditory impressions caused the positive stimuli. Probably referring to the (fortis) plosives, subjects said that some consonants are “explosive” (6, RC) or “come stronger” (12, HM), that some consonants like “*b*, *d* (are) blent” (11, EF). In addition, some consonantal sounds were identified as “nasal” (10, FS) and others as “hissing” (6, RC; 10, FS). This latter groups refers to the fricatives, something more evidence in subject 11 (ES)’s comment that “*s* and *z* are hissing”.

Further references to the auditory impression caused to the experimental subjects by both consonants and vowels were made. Putting all this information together, we can say that consonants were perceived, in general, as “shorter” (1, KL), “sharper” (1, KL), “harder” (3, CL; 6 RC), “more stunted” (2, AH), or “more powerful” (8, SH) than vowels, perceived to be “softer” (1, KL; 3, CL; 6, RC) and “rounder” (1, KL; 8, SH), with which they probably implied that vowels are produced with little resistance to the air passage. Subject’s 4 (KG) characterisation of consonantal sounds as “short and sharp” somehow summarises the two main characteristics of consonants in opposition to those of vowels: shorter duration and more forceful articulatory. In may also be the case that some of these descriptions were motivated by occasional reflections over specific subsets of consonants, like plosives, fricatives, etc.



It is interesting to note that subjects provided many different examples of the category “consonantal sounds” (i.e. 25) but very few (i.e. 5) of the category “vowels”. Out of all the positive instances, subjects provided 12 examples of plosives (/ p /, 1 instance; / b /, 5 instances; / d /, 2 instances; / k /, 1 instance; / g /, 3 instances), 12 of fricatives (/ f /, 2 instances; / s /, 1 instance; / z / 4 instances; / θ / and / ð /, 1 instance each; / ʃ /, 3 instances), and 5 of affricates (/ tʃ /, 1 instance; / dʒ /, 3 instances). No example of the nasals was provided. As far as the target sounds of the test stimuli are concerned, examples of [h], [j], [w] were mentioned 1, 3, and 4 times respectively and mainly as examples of the category “consonantal sounds”. However, it should be mentioned that most of the subjects who mentioned words beginning with any of those sounds reported feeling the potentially ambiguous nature of [h], [j], and [w]. Subject 4 (KG) claimed that words like “which” or “use” were “a bit ambiguous” although he consistently included them in the category “consonants”. Similarly, subject 5 (KP) claimed to have included / w / and / h / in the category “consonantal sounds” but acknowledged that these were “a bit confusing”. According to subject 7 (KR), [w] and [j] were “a bit misleading” but she was more prone to considering them as vowels. Subject 8 (SH) gave “young” and “wet” as positive instances of the category “consonantal sounds” but claimed that the initial sounds were “a bit confusing”. Finally, subject 12 (HM) treated the words “use” and “hours” as non-members of the category but said that these were “a bit confusing”. It is also interesting to note in connection with the reference to the word “hours”, that two more subjects (7, KR; 8, SH) said that “hours” was an example of “vowels”. The fact 3 out of the 6 examples of vowels were the word “hours” shows a special concern with pointing out that orthography was irrelevant in the formation of the category (at least for subjects 7, 8, & 12).

Finally, although 8 out of the 20 subjects who part in this experiment did not reach criterion, a few words could be said about them. The post-experimental interviews conducted with six out of those eight subjects evinced that they had actually *not* formed the category intended by the experimenter. These subjects reported that the difference between the green words and the red ones was one due to the “overall sound structure” (subj. 13), “soft sounds in the green ones” (subj. 14), “distinctive *sh*” (subj. 15), “longer vs. shorter words” (subj. 16), “explosive vs. harder sounds” (subj. 17), and “more stressed vs. less stressed” (subj. 18). As mentioned above, these subjects had not reached criterion, making 32, 28, 31, 28, and 28 correct responses respectively in the learning session. There were, however, two subjects who provided a description of the category under investigation (and its opposing category) that seemed to indicate that they had somehow realised what the categorical distinction was pre-established and intended by the experimenter. These two subjects described the stimuli as “sounds involving less opening of the mouth vs. more round sounds” (subj. 19), and “anything’s that’s not a vowel” (subj. 20). However, these two people did not reach criterion, making only 27 and 28



correct answers in the learning session respectively, the lowest scores in the experimental group. There seem to be two reasons why these two subjects provided such apparently relevant descriptions: first, the two subjects realised at the very end of the task what the actual difference was between the positive and the negative stimuli, or second, they struck upon the difference after the test session while thinking about some of the stimuli encountered during the task. For this reason, their motor responses during the CF task were considered of lesser value and they were not taken into account.

Table 11. Name Given to the Category, Articulatory and Auditory Characteristics of the Category Instances, and Examples of the Category Provided in the Post-Experimental Interview.					
	Subj.	Name Given to the Category	Actions to Produce the Sound	Auditory Impression	Examples
1	KL	More consonantal vs. more vowel	Move the mouth forcefully	Consonants: shorter, sharper Vowels: softer, rounder	+ < cut >, + < zip >, + < sharp >, - < owl >
2	AH	Consonants vs. Vowels	Lips or tongue involved, stop the air, then blow them out	Some are more stunted (e.g. <i>p, b</i>)	+ < ship >, + < fish >, + < pill >, + < bill >
3	CL	Consonants vs. Vowels		Consonants: harder Vowels: softer	+ < zoom >, - < orb >
4	KG	Consonantal vs. vowel		Short and sharp	+ < Judge >, + < fish >, +? < which >, +? < use >
5	KP	Consonant sounds vs. plain Vowel sounds	Tongue, lips, teeth involved		+ < b >, + < d >, + < j >, + < w >, + < h >
6	RC	Consonants vs. Vowels	Some have air in the nose	Consonants: harder, some explosive, some hissing Vowels: softer	+ < sheet >, + < go >, + < zoom >
7	KR	Consonant vs. Vowel	Consonants involve more movement of the mouth		< hours >, + < w >, -? + < j >
8	SH	Consonant vs. Vowel.	The mouth is more narrowed for some of the consonants	Consonants: + powerful, stress initially is considerable. Vs: rounder	-? < young >, -? < wet > - < hours >
9	LH	Consonant vs. Vowel	Lips, tongue, teeth, alveolar ridge, change position, move		+ < jump >, + < won >
10	FS	Anything that's not a vowel	Vowels imply less movement of the lips	Some are nasal, some are hissing	+ < goose >, + < this >, + < chin >
11	ES	Consonant vs. Vowel	More mouth vs. more open, more immediate	<i>S, z</i> : hissing <i>b, d</i> : blent	+ < boom >, + < judge >, + < bet >, + < dog >, + < set >, + < zip >
12	HM	Consonant vs. vowel	Mouth more closed, tongue moves vs. more open and the tongue doesn't move.	Some (e.g. <i>b</i>) come stronger	+ < bet > - < hours >, - < use >

2.4. Discussion.

The above experiment intended to obtain psycholinguistic evidence on the formation of the category "consonantal sounds" and the classification of the segments [h], [j], and [w] in word-initial position by adult native speakers of English.

The results of the both the learning and test sessions indicate that 12 out of the 20 subjects who took part in the experiment were able to form the category. First, those subjects

reached the pre-established criterion of providing at least 37 correct answers in the learning session, the mean number of correct responses being 48.67 (range 37-59, s.d.= 6.50). Second, subjects' performed increasingly better as the task proceeded, making significantly more correct responses in the test session (93.01%) than in the learning session (84.44%) and, consequently, fewer errors in the test session (6.99%) than in the learning session (15.56%) despite the fact that stimuli were unreinforced in the test session. Third, as can be seen in tables 6 and 7, subjects generalised accurately to both positive and negative controls. The extension of them acquired classifying behaviour to uncontroversial unreinforced members not previously encountered supports the claim that subjects' formed the category under investigation. The percentages of correct responses to the single positive control stimulus (i.e. / ʃ /) was 100%, the highest of all the different types of positive items (except for / tʃ / and / dʒ /, which had the same percentage). The same applies, *mutatis mutandis*, to the negative controls (/ ɔ: /: 95.83%; & / əʊ /: 100%). In addition, most subjects provided similar verbal descriptions of the category that were also similar to the traditional names and characteristics cited in the specialised literature. Thus, for at least those 12 speakers the category "consonantal sounds" had some psychological validity, in the sense of being a conceptual category or scheme according to which they were able to categorise phonetically different sounds together as members of the same class. In view of all these facts, it seems that the answer to the first research question, namely, whether subjects could form the category "consonantal sounds" is 'yes' for the criterial subjects and the hypothesis entertained (i.e. that they would be able to form the category) is confirmed.

The results of the test session also show that subjects tended to categorise word-initial, prevocalic [h], [j], and [w] as members of the category "consonantal sounds". In all cases the percentage of responses considering these test speech sounds as consonantal sounds was higher (i.e. / h /: 83.33%; / w /: 66.67%; / j /: 66.67%) than the percentage that did not (i.e. / h /: 16.66%; / w /: 33.33%; / j /: 33.33%). The answer to the second research question, that is, how subjects would classify [h], [j], and [w], is that they treat these speech sounds (at least in word-initial position) as members of the category consonantal sounds and the hypothesis formulated in 3.2.1., that subjects would behave as indicated, is again confirmed.

However, the actual answer to the second research question is a qualified yes in the sense that a close scrutiny of each test word reveals that the knowledge of the orthographic representation of the target sounds in the test stimuli probably influenced subjects' responses to some extent. More specifically, when the target sounds (i.e. [h], [j], and [w]) were written with letters of the alphabet that people usually refer to as "consonants" (i.e. "h", "y", "w" respectively), orthography did not seem to interfere in subjects' responses. If any influence at all, it could be argued, the conventional spelling of the test words in these cases may have



served to reinforce subjects' conviction that the target sounds were, for them, members of the category "consonantal sounds". However, when the target sounds were spelled with letters of the alphabet commonly referred to as "vowels" (e.g. "o" or "u"), as in "once" or "use", subjects seemed, in general, to be prevented from classifying the target sounds as members of the category "consonantal sounds". This seemed to happen that even when subjects were aware that they exclusively had to pay attention to sounds, not letters, and even though subjects did not see the written form of any stimulus type at any moment before, during, or after the experimental task.

This explanation seems plausible in view of the fact that, as Abercrombie says, the terms vowel and consonant "have by now become fairly ambiguous (for example, they are often applied to letters of the written alphabet, not to units of the spoken language at all)" (1967: 39). It is also likely that subjects' everyday conception of vowels and consonants as categories of sounds becomes intertwined with their knowledge about the spelling of sounds since the earliest days of their formal schooling and exposure to the written language.³¹ It is also very likely that many (if not all) subjects who took part in this experiment came to the experimental situation with an already pre-established category of "consonants" and "vowels" in which spelling and letter names are closely related (and somehow equated) and that when confronted with a speech segment somehow ambiguous for them, they consciously accessed the spelling of the word in which the sound appeared, which would tilt the balance on the *vowel letter=vowel sound* side or the *consonant letter=consonant sound* one depending on the spelling of the target sound.

If the experimenter wished to neutralise the apparently obvious role of orthography in the classification of word-initial [h], [j], and [w], it might be interesting to conduct CF studies similar to the one reported above with English-speaking illiterate adults. In fact, this recommendation has already been proposed in the experimental phonological literature in relation to other phonological problems (e.g. Jaeger 1980a; Nearey 1981). In addition, it might also be interesting to conduct CF studies with literate subjects but using non-words (i.e. accidental gaps),³² or even with preliterate children. Such studies might yield very different results from the ones obtained in this dissertation.

However, the possible influence of orthography in the results should not be regarded as an undesirable phenomenon. If one is interested in the categories "consonantal sounds" or "vocalic sounds" as they are conceptualised by the average literate subject that has received education at school, the fact that subjects' literacy is intimately linked to their phonological

³¹ Related to the specific phonological problem that we have been concerned with, it is interesting to note that orthography has already been shown to have some effect in the way subjects perceive and deal with [j] and [w]. Van Ooyen and his co-workers (1992) showed that in a phoneme monitoring task in which subjects had to detect the presence of [j], subjects missed considerably more responses for word-medial / j / when there was no corresponding grapheme "y" for it (e.g. "dune", "cubic", "fuse") than for word-initial / j /, where the orthographic representation of their stimuli was always "y". These researchers claimed that "since, in a language like English, such judgements can usually be supported by orthographic evidence, making these judgements is particularly difficult where no orthographic evidence is available" (p. 106).



awareness and the fact that the way conceptualise the sounds of the language derives, to a great extent, from their experience with the written language and educational background cannot be overlooked and should not be denied. On the contrary, orthography should be taken into account in a phonological description, account, or theory that tries to represent what is psychologically valid for speakers. As Jaeger puts it:

It is important that in making claims about the psychological status of linguistic phenomena, linguists try to sort out intuitively acquired linguistic behavior from education-derived behavior, but not be afraid to consider both as potentially psychologically real.

Jeri J. Jaeger (1980a: 262).

Acknowledging the influence or importance of orthography in the classification of [h], [j] or [w] does not mean that subjects were basing *all* their answers (not even the majority) to positive and negative stimuli on purely orthographic criteria during the experimental session. In the first place, the reports in the post-experimental interview by *all* the subjects who succeeded in the task reveal that they had formed, as told in the instructions, a category of *sounds*, not a letter-based concept. In fact, subjects, often referred to the articulatory gestures needed to produce the instances of the category under investigation and absent in the production of the members of the opposing category (i.e. “vocalic sounds”). In addition, they mentioned auditory impressions produced by the members of either the positive or negative categories. It seems then that the primary basis of classification or classificatory strategy in subjects’ performance was attention to similar articulatory and acoustic overall impression of the positive instances of the category “consonantal sounds” in opposition to the auditory impression and articulatory characteristics of the category “vocalic sounds”.

In the second place, the answers to the orthographically-interfering negative stimulus “hours” reveals that subjects were increasingly aware that the letters with which the stimuli are conventionally spelled were irrelevant for category membership. Subjects provided a 83.33% of correct responses to “hours” even though that word occupied position 23 in a 60-word sequence in the learning session. This indicates that even before the middle of the learning session, subjects were judging sounds, not letters. As subject 7 (KR) said: “*hours* made me think spelling had nothing to do”. In addition, it is also remarkable that 3 out of the 6 examples provided of “vowels” by the criterial subjects in the post-experimental interview were the word “hours”, which seems to indicate subjects’ emphasis on the fact that the category “consonant(s)” was not spelling-based.

In the third place, it should be mentioned that for a word like “use”, a phonetic-based classificatory criterion complementary to the orthographic one cannot be ruled out. Given that post-consonantal [j] is almost always followed by [u:] provided it does not undergo vowel reduction) (e.g. “puny”, “beauty”, “fuse”, “muse”, “view”, “cue”, “hue”, “tube”, “duty”, “new”, “suit”, “student”, “skew”) in the onset position of syllables in British English, [ju:] seems so

close-knit a sequence that some linguists prefer to consider the whole sequence as a single vowel unit (Kreidler 1989: 121). Although no experimental evidence is available confirming that subjects actually treat the sequence as a single vowel unit, this possibility cannot be ruled out.

Finally, it is remarkable that even in cases like “one” and “use”, in which spelling may have influenced subjects’ responses, the target sounds (i.e. [j] and [w]) were classified as members of the category “consonantal sounds” one third of the times. This implies that some of the subjects were not influenced by spelling and considered [j] and [w] as “consonants”.

Why did a third of the answers to [j] and [w] treat them as consonants?. This could be possibly be due to the double articulation of [w], which is a labio-velar sound that involves considerable “movement” of the mouth” (i.e. a double articulation with rounded lips and a raising of the back of the tongue towards the velum) and to the close position of the front of the tongue to the hard palate in the production of [j]. In other words, [j] and [w] involve a close-front and a close-back tongue position respectively. However, the close tongue positions of [j] and [w] do not seem to explain why [i:] (3 instances) and [u:] (2 instances) were not classified as consonants (88.89% and 91.67% of the time respectively –these percentages are higher than the mean percentage of correct responses to the monophthongal phonemes). Perhaps the fact that [w] and [j] imply a rapid vocalic glide from a close position to a more open one necessary for the following vowel, which is also usually more steady in quality, could explain why [w] and [j], spelled with “o” and “u” respectively, were considered as consonantal sounds 33.33% of the time. However, diphthongs also imply a gliding movement as their quality changes but they were clearly identified as non-members of the category consonantal sounds even more accurately (91.11% of the time) than the monothphongs (87.33%) so the gliding nature of [j] and [w] seems to be have had no effect on subjects’ responses to [j] and [w] as consonants 33.33% of the time when these test segments were spelled < u > and < o > (respectively). Perhaps [j] and [w] were sometimes considered as consonants because they occupy a non-nuclear position, like typical consonants, according to the specialised literature. Unfortunately, the fact that both consonantal sounds and vocalic sounds appeared in word-initial position does not allow us to assess directly the importance of distributional factors in either the difference between vowels and consonants or in the classification of the test words. However, we suspect that the fact that [h], [j], and [w] occupy the onset position of the syllable according to traditional interpretations, may have contributed to subjects’ classification of the test items as members of the category “consonantal sounds” as well as their spelling representations. Looking at [h], we may see why this could be so.

[h], which is a voiceless version of the following vowel, is usually classified as phonetic fricative consonant because it involves friction but the “friction” it involves is actually

cavity friction which should be distinguished from the *local friction* typical of clear examples of fricatives, although these two phenomena are not usually satisfactorily distinguished (Pike 1943: 71-72). To put it briefly, local friction results from stricture at a single local point while cavity friction is due to cavity friction that is, voiceless resonance of a chamber as a whole caused by air going through it. As both voiceless and voiced vowels also have cavity friction, there is no reason why they should not be classified as phonetic fricatives if [h] is because it involves cavity friction. However, although no author would agree on that, [h] is probably classified as a *consonant* probably because it functions as a consonant in the phonological sense (i.e. marginally rather than as a syllable nucleus) given that it is actually a phonetic vowel and there is other phonetic criterion to convincingly classify it as a consonantal sound, although some try to justify such classification on the inappropriate criterion of *cavity friction* (for a sound to be classified as a fricative consonant). A similar situation applies to [j] and [w] are also described as phonetic vowels but they are in the end erroneously described as *phonetic consonants* because of the influence of their marginal function in the phonological syllable, and this classification is justified in phonetic terms due to allophones that are voiceless and/or involve friction. The unconscious mixture of both phonetic and phonological criteria in the classification of sounds or units is widespread (Abercrombie 1967: 78; Pike 1943: 77-79).³³ The experimental subjects may have behaved similarly to the way phoneticians influenced by phonological criteria do except for the fact that phoneticians justify their treating [h], [j], and [w] as phonetic consonants due to often marginal allophones that are voiceless or involve friction but language users justify that treatment also on the basis of conventional spellings but their classifications may be strongly influenced by the onset position of [h], [j], and [w] in the syllable.

To conclude this discussion, we should like to draw the reader's attention to the fact that the present experiment has elicited as many questions as it has tried to answer and that the classification of word-initial [h], [j], and [w] obviously deserves further experimental attention. Although the criteria according to which phonetically naive language users classify word-initial [h], [j], and [w] are far from being completely understood at present given the results of experiment 1, it seems that we have started to understand why the issue is so formidable and the variables and phenomena where to look in order to produce a psychologically accurate account of the classification of these speech segments.

³³ As Pike says "although many articulatory and acoustic criteria may be mentioned by an author, somehow the resultant groupings always reflect the phonemic divisions as seen in permitted groupings of phonemes, sequences of syllables, and the like" (Pike 1943: 74)



3.3. Experiment 2. The Category “Oral Plosives” in English.

3.3.1. Design: Rationale, Purpose of the Experiment, Research Questions, and Hypotheses.

The complete articulation of an oral stop made with a pulmonic egressive air-stream mechanism is called (oral) plosive.³⁴ The production of a plosive consists of three brief stages: (1) a “closing” stage, (2) a “hold” stage, and (3) an “opening” stage (see Abercrombie 1967: 48, 140-141; Gimson 1980: 150-151; Gleason 1955: 22; Hockett 1955: 35-36, 245; Jones 1989: 138, 152; Ladefoged 1993: 9; Roach 1983: 28-31).³⁵

During the *closing* stage, the articulating organs move together so as to form a complete obstruction of the air passage. One articulator is moved against another, or two articulators are moved against each other. A stricture that allows no air to escape from the vocal tract is then formed while the closure is maintained.³⁶ This momentary but complete interruption of the air-stream is accompanied simultaneously by a velic closure which prevents the air from entering the nasal cavity.³⁷ Next, during the *hold* stage, which may be accompanied or not by voice (i.e. vibration of the vocal cords), the air is completely dammed up, and unable to get through the vocal tract at all while lung action momentarily compresses the air behind the stricture formed during the closing stage. Finally, during the *opening* stage, the organs forming the obstruction part rapidly. This allows the compressed air to be released and, if sufficient pressure built up behind the stop while there was occlusion, to escape abruptly with a small explosion. Ideally, the air behind the stricture is still under pressure when the release is made. In this case, the escape of air produces noise loud enough to be heard. This noise is called plosion.³⁸

In English there are six consonantal phonemes of a wide distribution that are usually realised as plosives. These consonants are usually grouped in three different pairs: / p /-/ b /, / t /-/ d /, and / k /-/ g /. The point of articulation is bilabial for the first pair, alveolar for the second pair, and velar for the third. The six consonantal phonemes also imply a velar closure, which also distinguishes them from the homorganic nasals (see Figure 8). Traditionally, these pairs have been distinguished by voicing. The first member of the pairs (i.e. / p t k /) is said to

³⁴ The pulmonic (egressive) air-stream mechanism is the basis of almost all human speech sounds (Abercrombie 1976: 24). For a description of other air-stream mechanisms see, for example, Abercrombie (1967: 28-33), Bloch & Trager (1942: 31-32) or Pike (1943: 89-106).

³⁵ The closing stage is occasionally referred to as “shutting” stage, and the hold stage as “closure”, “occlusion”, or “compression” stage. The “opening” stage is also called “burst”, “release”, or “(ex)plosion” stage.

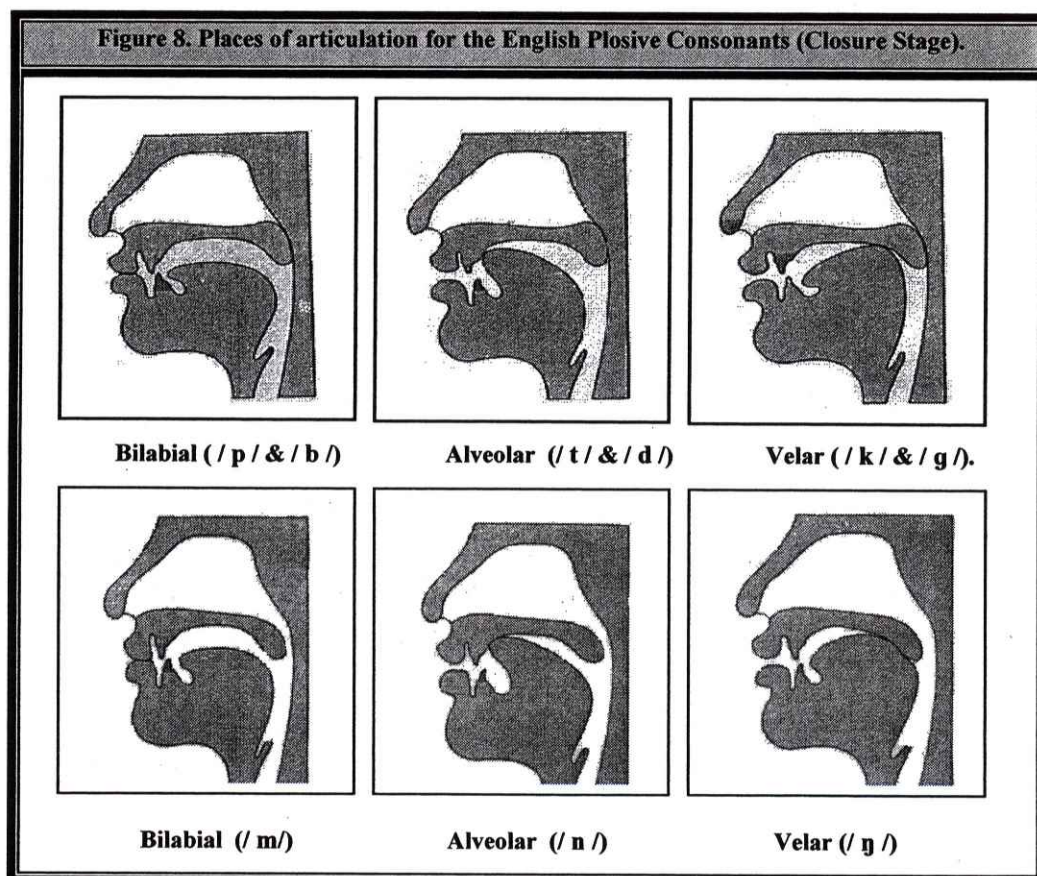
³⁶ Stricture is the technical name for the position taken up by an active articulator in relation to a passive one.

³⁷ A nasal is a type of consonant segment which, like a stop, is produced by a stricture of complete closure in the oral tract but with no simultaneous velic closure. The air-stream, therefore, though prevented from passing through the mouth, is not dammed up but is entirely diverted through the nose and is allowed to set up resonance there (see, for example, Abercrombie 1967: 30, 43, 48; Brown 1977: 23; Gimson 1980: 31, 193; Ladefoged 1993: 8-9).

³⁸ Some authors divide the third stage into two different ones: an “opening” or “release” stage (an articulatory feature which is the converse of the closing stage), and a “burst”, “explosion” or “post-release” stage (an auditory feature) (Arnold 1967; Roach 1983: 28-31; see also Gimson 1980: 150, footnote 2).



be voiceless and the second (i.e. / b d g /) voiced. However, although / b d g / may have full voice during their second stage when they occur between voiced sounds (e.g. “robber”, “lager”, “lager”), they are often partially devoiced or completely voiceless in word initial position as in “bill”, “dale”, “game”, in which case aspiration of / p t k /³⁹ is the main feature that distinguishes each pair as in [p^hrɪ] vs. [bɪrɪ], [t^herɪ] vs. [dɛrɪ], and [k^heɪm] vs. [ġeɪm] (e.g. O'Connor 1973: 132; Reeds & Wang 1961; Winitz *et al.* 1975). In final position, as in “lob”, “lad”, “lag”, / b d g / are also partially devoiced or completely voiceless and it is the length of the preceding vowels that helps discriminate which phoneme is instantiated (Gimson 1980: 151-152; Ladefoged 1993: 46, 48; Lisker & Abramson 1964: 420; O'Connor 1973: 130-131; Roach 1983: 30-31). However, the two sets are differentiated in this position, as in any other, because / p t k / tend to be pronounced with more muscular energy and a stronger breath effort than / b d g /; the former are known as relatively strong or *fortis*, the latter as relatively weak or *lenis* (Gimson 1980: 151; Lisker & Abramson 1964: 420; O'Connor 1983: 129; Roach 1983: 31).⁴⁰



³⁹ Aspiration is a strong and very brief burst of air followed by a short period of voicelessness, while the air escapes.

⁴⁰ It is probably true that / p t k / are produced with more force, but articulatory force has typically been a very difficult notion to define (either articulatory or acoustically) and measure (Lisker & Abramson 1964: 320; Roach 1983: 31). According to Lisker and Abramson (1967), the assessment of articulatory force appears ultimately to be a matter of proprioceptive judgement. This judgement, however, is often claimed to depend directly on the audible features of closure duration and the loudness of the stop explosion. The duration of stops is greater for the fortis than for the lenis ones. Acoustically the intensity of the burst of the fortis plosives / p t k / is greater than that of / b d g /.

Although / p /- / b /, / t /- / d /, and / k /- / g / are normally realised as plosive consonants with a rapid release of compressed air leading to a short, sharp explosion, these stops are sometimes *affricated*. This means that when the release of the closure is not made rapidly, a fricative sound, articulated in the same area of articulation as the plosive, is heard (Bloch & Trager 1942: 33; Gimson 1978: 158-159; O'Connor 1973: 132). In these cases, the separation of the articulators is slowed down so that they take long enough to pass through a stricture of close approximation for the resulting homorganic friction to be audible, even though it is brief. This is known as affrication, and the plosion is then said to be *affricated*, one of the ways in which the release stage of a stop may be modified. In English, only / t / and / d / are commonly affricated, in particular, in popular London speech. In this variety of English, plosives made with this slow, fricative release (i.e. [t^s] and [d^z]) may be heard in strongly accented positions (e.g. "time" [t^saɪm], "door" [d^zɔː]), in relatively weakly accented positions (e.g. "waiting" ['weɪt^sɪŋ], "riding" ['aɪd^zɪŋ], and in final positions (e.g. "cat" [k^hæt^s], "bed" [bɛd^z]).

Affricated stops like [t^s] and [d^z] differ from the realisation of the plural terminations [t] + [s] and [d] + [z] mainly in the brevity of the friction associated with the former. When the friction following the stop is sufficiently marked to be considered a separate segment, as is the case with [t] + [s] and [d] + [z], the cluster of stop and homorganic fricative is usually known as *affricate*. In other words, an affricate is merely a close-knit sequence of a stop produced with a delayed release and followed by a homorganic fricative, that is, a fricative articulated at the same point of articulation (Abercrombie 1967: 148; Chomsky & Halle 1968: 318-322; Gleason 1955: 22-23; Hockett 1955: 40, 164; Ladefoged 1993: 60; Ladefoged & Maddieson 1996: 90; O'Connor 1973: 130, 138).⁴¹ In English, affricates are exemplified by the sequences [tθ] (e.g. "eighth"), [dð] (e.g. "width"), [tʃ] (e.g. "tray"), [dʒ] (e.g. "dread"), [tʃ] (e.g. "chain"), [dʒ] (e.g. "jet"), and, of course, [ts], as in "cats", and [dz], as in "seeds".⁴²

⁴¹ Plosives are different from fricatives. In the production of fricatives air escapes through a small passage while there is never a complete obstruction as there is for a plosive. Instead, the air expelled by pressure from the lungs escapes with a kind of hissing, hushing, or buzzing sound produced because one articulator is placed so close to another so as to cause turbulence of the air-stream (see Abercrombie 1967: 48-49; Brown 1977: 22-23; Gimson 1980: 31, 178; Gleason 1955: 21-22; Jones 1989: 179; Ladefoged 1993: 9-10; Roach 1983: 35).

⁴² There has been a long-standing debate on how to represent affricates in either phonetics or phonology depending mainly on whether affricates are seen as unitary or complex sounds and/or units (Ladefoged 1993: 26). Although the IPA has never supplied separate symbols for the affricates (see Hockett 1955: 40 for a plausible explanation), affricates may be symbolised in many different ways: (1) modifying a some phonetic symbol with the diacritics [˥] or [˦] (e.g. [t̪˥] and [d̪˦]), (2) with groups of two consonants, like [ts] or [dz] (ligatured or not), or (3) with the marks [͡] or [͜] (e.g. [t͡s] or [d͡z]). Gleason (1955: 314) provides a summary of the conventions employed by different authors.



However, of all the possible affricates (i.e. [tʃ], [dʒ], [ts], [dz], [tʃ̺], [dʒ̺], [tʃ̺̺], and [dʒ̺̺]) and affricative combinations of plosives + non-homorganic fricatives,⁴³ only the palato-alveolar pair /tʃ/ and /dʒ/, the majority of phonologists and/or phoneticians claim, have a special status in English phonology (see Bloch & Trager 1942: 49; Bloomfield 1979: 20; Gimson 1980: 171-173; Gleason 1955: 22-23; Ladefoged 1993: 60; Roach 1983: 94-97; Trager & Bloch 1941: 229; Trager & Smith 1951; Wells 1982a: 48-49 for discussions). This “special” status is simply that they are not considered consonant clusters (–i.e. /t + ʃ/ and /d + ʒ /), what we might call the “two-phoneme” or bisegmental analysis. Instead [tʃ] and [dʒ], unlike other affricates or affricative clusters, are considered as *single* units (the one-phoneme or monosegmental analysis) for different distributional, combinatorial, and psycholinguistic reasons. First, [tʃ] and [dʒ] are the only affricates in English that can occur in syllable initial position, in the middle of the word or at the end, not necessarily as the result of the formation of a regular plural morpheme or the addition of some other suffix (Gimson 1980: 171-172; Ladefoged 1993: 60; Lass 1984: 26-27; Roach 1983: 37-38). This makes the distributional patterning (i.e. pattern congruity) of /tʃ/ and /dʒ/ similar to that of most single consonantal phonemes. Second, the possibilities of commutation of the supposedly two members of /tʃ/ and /dʒ/ are very limited. For example, any of the elements of /d + ʒ/ cannot be easily substituted for another consonantal phoneme (the other element remaining the same) producing an acceptable consonantal cluster in the same positions where /dʒ/ occurs. Finally, experimental evidence using the CF experimental paradigm (e.g. Jaeger 1980a: experiment 4) has shown that phonetically naive native speakers of English consider that /tʃ/ and /dʒ/ in words like “chip” or “gyp” are each “one sound” and not clusters of two consonants.⁴⁴ Also, beginning spellers, who usually represent affricates like [tʃ] with “t” (e.g. “tese” = “cheese”) or “s” (“teasur” = “teacher”) and [dʒ] with “d” (e.g. “dost” = “just”) or “s” (“losuz” = “largest”) never produce spelling clusters like “tsh” for /tʃ/ or “ds” for /dʒ/ because, despite appreciating the two-part nature of affricates, the two parts form a tightly-closed unit as the spelling behaviour suggests (Treiman 1993: 241, 284).

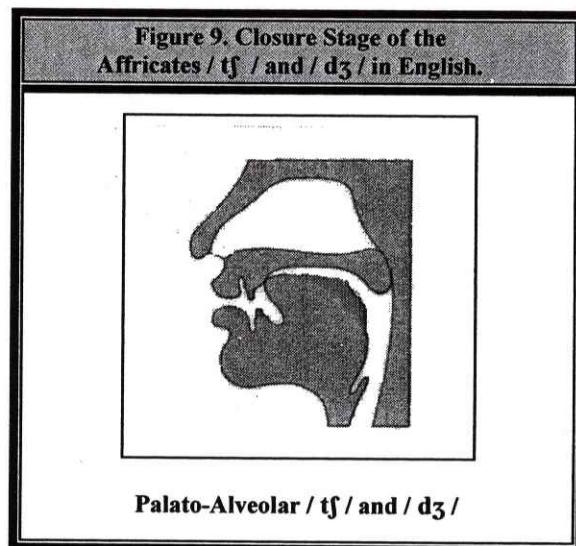
Although the “unisegmental (i.e. one-phoneme) vs. bisegmental (two-phoneme) analysis” of [tʃ] and [dʒ] has been a traditional phonological problem of analysis, there is a second problem of assignment or categorisation concerning those affricates that has been less often dealt with. Although /p t k/ and /b d g/ are, for almost every phonetician or phonologists, clear and uncontroversial examples of the phonological category “oral plosives”, /tʃ/ and /dʒ/

⁴³ Affricated stops and affricates should be distinguished from simple sequences of plosives plus fricatives that are not homorganic (e.g. [ps], as in “lapse”, [kf] as in “breakfast”, etc.).

⁴⁴ From now on, and following Gimson (1980: 173), the symbols /tʃ/ and /dʒ/ will be used for convenience but with the implication that they are single units, despite the apparently counterintuitive fact that two symbols are used and not just one.



could qualify as “plosives” because they imply a closing followed by a hold stage like / p /, / t /, etc. They would be then rather atypical plosives because of their affricative stage but they imply a complete obstruction of the air passage like any plosive. Figure 9 shows the closure stage of the plosives / tʃ / and / dʒ /. Brown (1977: 22), for example, seems to be of this opinion. However, / tʃ / and / dʒ / could also be considered as some sort of atypical fricatives because of their final stage in which friction is heard. However, the most common position is that / tʃ / and / dʒ / are phonetic affricates that are the (sole) members of a third phonological category called “affricates” distinct from the categories “plosives” and “fricatives”.



Depending on which features and how important they are taken to be, it seems that, on phonetic grounds, there might be reasons for each of the three views.⁴⁵ First, like the fortis plosives, / tʃ /, for example, may have a little aspiration, in the positions where / p t k / are aspirated. Also, after the friction segment / tʃ / is frequently glottalised, like the plosives, before pause and before consonants. Second, like the pair of fricatives / ʃ / and / ʒ /, / tʃ / and / dʒ / are palato-alveolar segments and/or units and imply lip-rounding. In addition, they end up like fricatives with a hissing sound. Finally, despite sharing a few features with both plosives and fricatives,⁴⁶ [tʃ] and [dʒ] are also somehow different group from both plosives and fricatives. For example, unlike the plosives, [tʃ] is frequently glottalised before vowels, the friction present in them is of shorter duration than that which characterises the fricatives proper, the stop period is shorter than it is for the other stops, and the release or the closure is very gradual giving rise to a strongly fricative sound.

⁴⁵ For the phonetic details mentioned see, amongst others, Brown (1977: 22, 30-31), Gimson (1980: 169, 171, 174-175), Hockett (1955: 22-23), Ladefoged (1993: 11, 48-49), Lass (1984: 27), O'Connor (1973: 129, 138), Roach (1983: 42-44), and Wells (1982a: 48-49).

⁴⁶ [tʃ] and [dʒ] represent a fortis-lenis pair like pairs of plosives - / p /- / b /, / t /- / d /, etc- or fricatives / s /- / z /, / θ /- / ð /, with the voicing distributional characteristics and effects of them -devoicing of [tʃ] in initial position, shortening of preceding vowels before [tʃ] in final position.

In view of the phonetic similarity of affricates to both stops and fricatives, a new CF experiment was designed to investigate the status of affricates from the point of view of the phonetically naive native speaker of English. The purpose of the present experiment was to provide evidence on the ability of subjects to form the categories “oral plosives”, and to find out how subjects classify / tʃ / and / dʒ /. The specific research questions that this study investigated were:

- 1) Are native speakers of English able to group different oral plosives together?. In other words, can speakers of English recognise plosives as a group?
- 2) Do people consider / tʃ / and / dʒ / as members of the oral plosive category?.

By the label “oral plosives” we mean the group of speech sounds that, according to traditional phonetic criteria, qualify as clear, uncontroversial examples of a pulmonic egressive stop realisation (i.e. the phonemes / p t k b d g / as they are realised naturally in word-initial position).

The hypotheses formulated before conducting the CF experiment designed to answer those research questions were that subjects would learn the category “oral plosives” and that they would not consider affricates as members of the category “oral plosives”. This second hypothesis is based on the fact that the fricative part of the affricates may be strong enough for subjects to exclude them from the category.

3.3.2. Method.

3.3.2.1. Subjects.

Twenty subjects between the ages of 20 and 39 (mean age 23) were selected randomly from the 80-subject group recruited for the CF experiments reported in this work in order to take part in this experiment. There were 9 men and 11 women. Subjects were paid for their participation. The subjects had the characteristics described in 3.1.1.2.

3.3.2.2. Stimuli.

The stimuli for the present experiment consisted, like in experiment 1, of 100 monosyllabic English words. All the stimuli (positive, negative, and test) had a consonant+vowel+consonant (CVC) syllable structure. The type of sound that subjects had to detect was again at the beginning of each word. The stimuli had a mean duration of 601 msc.

In the learning session, positive instances (32 items in total) included the plosive phonemes / p t / (7 instances each) and / b k g / (6 instances each) in word-initial position. Negative instances (28 items in total) contained the non-interfering fricative phonemes / s z f v ʃ / (3 items each).⁴⁷ There were a few negative stimuli potentially interfering because of certain orthographic and/or phonetic characteristics. Orthographic interference might occur, it was argued, whenever a phoneme is spelled with a letter that is a typical orthographic representation of one of the members of the oral plosive category or such letter has no phonological value. In the present study, the orthographic interfering words were word-initial instances of the phoneme / s / spelled with the digraph “ps-” (i.e. “psalm”) and the phonemes / θ / and / ð / (3 instances each), spelled with the digraph “th” (e.g. “thief”, “this”, etc.). In addition, there were 4 phonetically interfering negative stimuli starting with / m / (3 items) and / n / (1 item), spelled “m” and “n” respectively. Phonetic interference derives from the fact that, except for the lack of velic closure, nasal consonants are very similar to oral stops in that they involve a complete closure in the oral tract in their production and share the point of articulation with some of the oral stops (e.g. / m / is bilabial like / p / and / b /, and / n / is alveolar like / t / and / d /). Finally, there were 2 examples of word-initial / n / with both phonetic and orthographic interference because the nasal was spelled with “kn-” (“knife” and “kneel”), where “k” is a typical spelling form of / k /, a clear example of the category “plosives”.

In the test session, positive instances (19 items in total) included all the consonantal phonemes already encountered in the learning session (3 instances per phoneme). In addition, 4 positive instances included the lenis alveolar stop / d / (i.e. “dish”, “deaf”, etc.) as their initial consonantal sound. / d / was a plosive phoneme not previously encountered by the experimental subjects in the learning session and that functioned as a positive control. Negative instances (12 in total) included examples of all the non-interfering fricative phonemes already encountered in the learning session, namely / s z v ʃ / (1 instance each) and / f / (2 instances). There were also 3 orthographically interfering words (“th”= / θ /, 1 instance; “th”= / ð /, 2 instances), 2 phonetically interfering (/ m /, 1 instance; / n /, 1 instance), and 1 orthographically and phonetically interfering item (“gn-”= / n /, in “gnash”), which was a (negative) control as its spelling form for the nasal had not been previously encountered. Test tokens (9 in total) included 5 / tʃ /-initial monosyllables (i.e. “chill”, “chief”, etc.) and 4 / dʒ /-initial ones (e.g. “jazz”, “juice”, etc.).

No items in either the learning or the test sessions included an oral stop in final position. In other words, all positive items contained an oral plosive in initial position (e.g. “pill”, “deaf”,

⁴⁷ This study did not include either as negative, positive or test tokens the phonemes / h r l w j / because of their somehow atypical nature as consonants. In addition, as in the case of experiment 1, no monosyllabic words including / ʒ / and / ŋ / were included as these consonants constitute phonological segmental constraints in English at the beginning of words.

“case”) but never a plosive in final position (e.g. “lap”, “sear”, “seek”). The same applies to the negative items (except for the fact that they did not begin with a plosive). The avoidance of word-final plosives in the negative stimuli was made in order not to give the experimental subjects any opportunity to treat them as positive instances of the category after focusing perhaps unconsciously on the overall sound structure of the stimuli.

The consonantal sounds at the beginning of all stimulus types were spelled with one letter except for: 1 positive (“gu”= / g /, in “guess”), 4 non-interfering negatives (/ ʃ /= “sh”), 13 interfering negatives (“th”= / ð /, 5 instances; “th”= / θ /, 4 instances; “kn”= / n /, 2 instances; “gn”= / n /, 1 instance; “ps”= / s /, 1 instance), and 5 test items (“ch”= / tʃ /).

Table 12 offers a summary of the types of stimuli used and table 13 the actual list of words employed as well as their category status.

Table 12. Types of Stimuli and Number of Items per Subtype of Stimuli Used in Experiment 2.				
Type of Stimulus	LEARNING SESSION		TEST SESSION	
	Subtype of Stimulus	Number of Items	Subtype of Stimulus	Number of Items
POSITIVE ITEMS	/ p b t k g /	32	/ p b t <u>d</u> k g / ⁴⁸	19
	TOTAL	(32)	TOTAL	(19)
NEGATIVE ITEMS	<i>Non-interfering:</i>		<i>Non-interfering:</i>	
	/ s z f v ʃ /	15	/ s z f v ʃ /	6
	<i>Interfering:</i>		<i>Interfering:</i>	
	Orthographically < ps > (= / s /) & < th > (= / θ, ð /)	1 6	Orthographically < th > (= / θ, ð /)	3
	Phonetically < m n >	4	Phonetically < m n >	2
	Orthographically & phonetically < kn > (= / n /)	2	Orthographically & phonetically < gn > (= / n /)	1
	TOTAL	(28)	TOTAL	(12)
TEST ITEMS	----	---	/ tʃ /	5
			/ dʒ /	4
			TOTAL	(9)

⁴⁸ Underlined phonemes in bold letters represent control stimuli.

Table 13. Stimulus List for Experiment 2:
Category "Oral Plosive(s)".

Order	Stimulus	Positive (+) Negative (-)	Order	Stimulus	Positive (+) Negative (-)
LEARNING SESSION					
1.	push	+	53.	cave	+
2.	fall	-	54.	zone	-
3.	bus	+	55.	gull	+
4.	verse	-	56.	shell	-
5.	tall	+	57.	thighs	- int
6.	cash	+	58.	path	+
7.	gas	+	59.	thus	- int
8.	safe	-	60.	ton	+
9.	pace	+	TEST SESSION		
10.	zeal	-	1.	pave	+
11.	beef	+	2.	bath	+
12.	tough	+	3.	tool	+
13.	shove	-	4.	file	-
14.	kill	+	5.	coal	+
15.	girl	+	6.	chill	test
16.	thief	- int	7.	seethe	-
17.	these	- int	8.	veil	-
18.	fish	-	9.	goose	+
19.	pill	+	10.	pause	+
20.	verve	-	11.	zoom	-
21.	bill	+	12.	ball	+
22.	sell	-	13.	jazz	test
23.	zoos	-	14.	shoal	-
24.	toes	+	15.	dish	+ control
25.	case	+	16.	thieve	- int
26.	gaze	+	17.	tale	+
27.	knife	- int	18.	chief	test
28.	pile	+	19.	deaf	+ control
29.	psalm	- int	20.	juice	test
30.	kneel	- int	21.	call	+
31.	beige	+	22.	those	- int
32.	shave	-	23.	give	+
33.	tail	+	24.	pull	+
34.	cough	+	25.	moth	- int
35.	thaws	- int	26.	dull	+ control
36.	miss	- int	27.	nose	- int
37.	goal	+	28.	choose	test
38.	nerve	- int	29.	buzz	+
39.	puff	+	30.	cheese	test
40.	booze	+	31.	fill	-
41.	this	- int	32.	time	+
42.	tease	+	33.	jail	test
43.	mill	- int	34.	gnash	- control
44.	cool	+	35.	choice	test
45.	gash	+	36.	deal	+ control
46.	five	-	37.	cause	+
47.	mass	- int	38.	guess	+
48.	veal	-	39.	jaws	test
49.	pass	+	40.	these	- int
50.	bush	+			
51.	sauce	-			
52.	toll	+			

3.3.2.3. Procedure.

3.3.2.3.1. Instructions.

The instructions were exactly the same as the ones that subjects had read in the first experiment. All subjects reported having understood the instructions fairly well. The training session began only when the experimenter had checked that this was actually so after discussing the contents and sequence of events of the experimental task with the subjects.

3.3.2.3.2. Training Session, Test Session, and Post-Experimental Interview.

Subjects were run individually in a sound-treated room as in experiment 1. The training and test sessions proceeded as described in 3.1.1.5. After the computer-based task finished, a personal post-experimental interview was conducted. No incidences occurred in any of these sessions.

3.3.3. Results.

3.3.3.1. Training and Test Sessions.

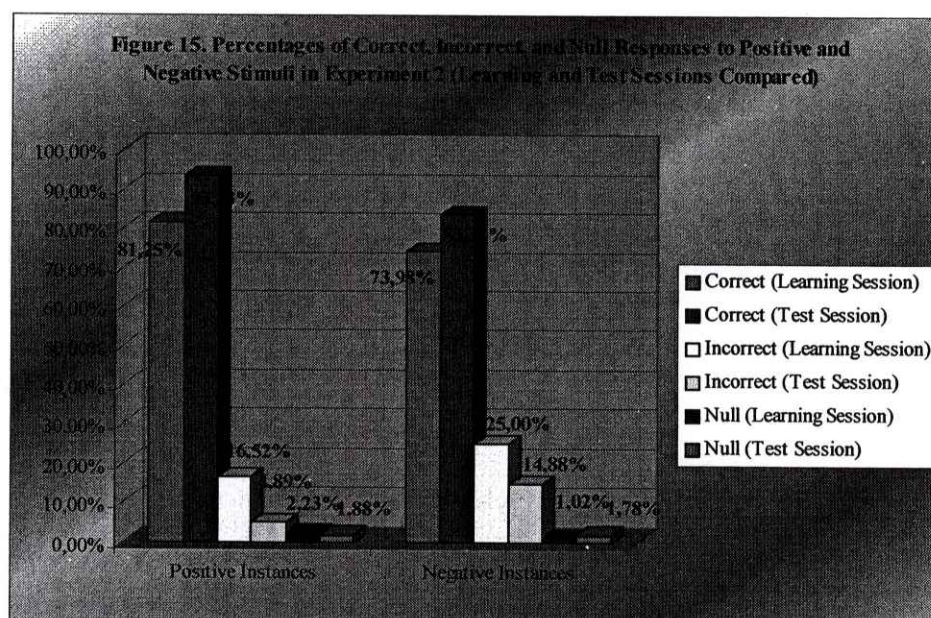
14 out of the 20 subjects who participated in experiment 2 (9 women and 5 men) reached the 37-correct-response criterion that guaranteed that they had not responded randomly (P-value: $0.03 < 0.05$) and, consequently, they were considered to have formed the category. The mean number of correct responses in the learning session was 47 (range 37-59, s.d.= 7.06).

Subjects' correct (C), incorrect (I), and null responses (NR) to both positive and negative stimuli and the percentages of correct responses in the learning session are shown in Table 14. The table also shows the number of items per type of stimulus, and the number of responses elicited in each case. The same information obtained from subjects' performance in the test session is shown in Table 15.

Table 14. Number of Correct/Incorrect/Null Responses and Percent Correct Responses to Positive and Negative Stimuli in Experiment 2 (Learning Session).						
Stimulus Type	Type of Response			% Correct Responses	Items	Number of Responses Elicited
	C	I	NR			
Positive Instances	364	74	10	81.25%	32	448
Negative Instances	290	98	4	73.98%	28	392
TOTAL	654	172	14	77.86%	60	840

Table 15. Number of Correct/Incorrect/Null Responses and Percent Correct Responses to Positive and Negative Stimuli in Experiment 2 (Test Session).						
Stimulus Type	Type of Response			% Correct Responses	Items	Number of Responses Elicited
	C	I	NR			
Positive Instances	248	13	5	93.23%	19	266
Negative Instances	140	25	3	83.33%	12	168
TOTAL	388	38	8	89.40%	31	434

Looking closely at the overall percentage of correct responses in the test session compared with that in the learning session, it can be seen that many more correct responses occurred in the test session. In fact, subjects were responding significantly more accurately in the test session than in the learning session (77.86% versus 89.40%; $P\text{-value} = 0.000 < 0.05$ by a contrast of proportions). In fact, subjects' significantly better performance in the test session is also reflected in figure 15, which shows visually the increase of correct responses to positive stimuli as well as the general decrease of incorrect and null responses to positive and negative stimuli from the learning session to the test session.



Looking at possible relationships between type of stimulus (i.e. negative, positive) and type of response (i.e. correct, incorrect, null), it was actually found out that in the learning session there was a significant relationship between type of stimulus and type of response (P -

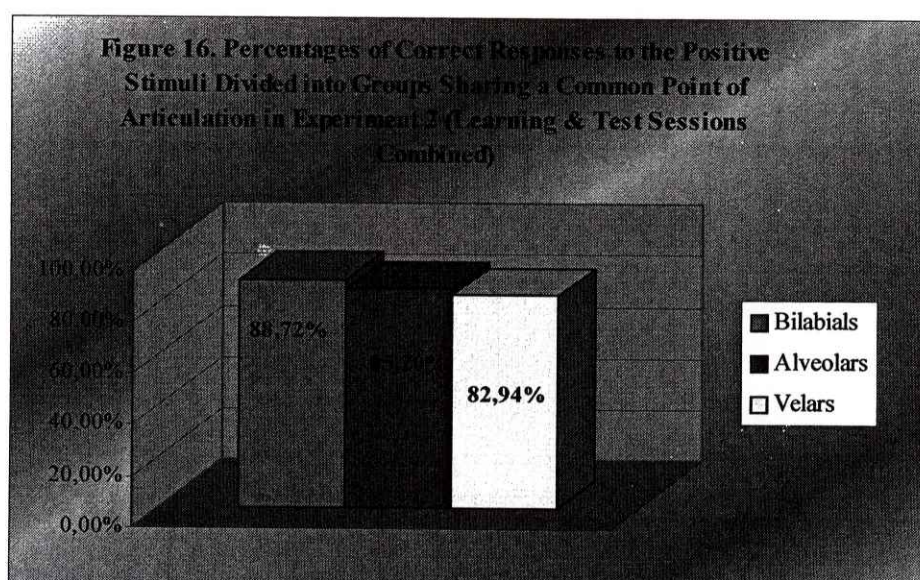
value = $0.005 < 0.05$ by a test of independence or chi-square test). More specifically, focusing on the corrected residuals of the test, positive stimuli were found to be in inverse proportion to the number of incorrect responses while negative stimuli were in direct proportion to the number of incorrect responses. This indicates that it was harder to be right with negative stimuli than with positive stimuli. In the test session, the expected minimum frequency was below 5 and so the test of independence was not reliable.

Collapsing the values of the variable “type of response” (i.e. correct, incorrect, and null response) into two categories (i.e. correct vs. incorrect/null), a significant relationship between type of stimulus and type of response was also found in the learning session (P-value: $0.014 < 0.05$ by a test of independence using the statistic Yate’s corrected chi-square test). Focusing on the corrected residuals, it became clear that the number of positive stimuli were again in inverse proportion to the number of incorrect/null responses and in direct proportion to the number of correct responses. On the contrary, negative stimuli were in inverse proportion to the number of correct responses and in direct proportion to the number of incorrect/null responses. In the test session, and focusing again on the corrected residuals of the test, a significant relationship between type of stimuli and type of response was also obtained (P-value: $0.02 < 0.05$ by a test of independence using the statistic Yate’s corrected chi-square test). More specifically, the number of positive instances was in direct proportion to the number of correct responses and in inverse proportion to the number of incorrect/null responses. On the contrary, the number of negative stimuli was in inverse proportion to the number of correct responses and in direct proportion to the number of incorrect/null responses.

The fact that the number of negative stimuli is in direct proportion to the number of incorrect responses in the learning session and to the number of incorrect/null responses in both the learning and the test session may be related, as mentioned in section 2.3.1.2. (see also Hovland & Weiss 1953) to the fact that it is easier to be right with positive stimuli than with negative ones, which tend to hinder concept learning.

Looking next at subjects’ responses to the positive stimuli in more detail, it seemed interesting to find out whether there were any significant differences between different types of positive stimuli. In this respect, Table 16 shows then the numbers of correct, incorrect, and null responses to the positive stimuli in both the learning and test sessions (combined) grouped by common point of articulation (bilabial, alveolar, or velar) as well as by type of phoneme (e.g. / p /, / b /, / t /, etc.). The percentages of correct responses to each type of phoneme and point-of-articulation group, the number of items per type of phoneme and per group, and the number of responses elicited in each case are also indicated. The percentages of correct responses to the four different types of positive stimuli in both the learning and the test sessions (combined) can also be seen in figure 16.

Table 16. Number of Correct/Incorrect/Null Responses and Percent Correct Responses to Positive Stimuli Grouped by Place of Articulation and Type of Phoneme in Experiment 2 (Training and Test Sessions Combined).							
Type of Plosive	Phoneme	Type of Response			% Correct Responses	Items	Number of Responses Elicited
		C	I	NR			
Bilabials SUBTOTAL	/p/	123	9	8	87.85%	10	140
	/b/	113	11	2	89.68%	9	126
		(236)	(20)	(10)	(88.72%)	(19)	(266)
Alveolars SUBTOTAL	/t/	114	24	2	81.43%	10	140
	/d/	53	2	1	94.64%	4	56
		(167)	(26)	(3)	(85.20%)	(14)	(196)
Velars SUBTOTAL	/k/	99	25	2	78.57%	9	126
	/g/	110	16	0	87.30%	9	126
		(209)	(41)	(2)	(82.94%)	(18)	(252)
TOTAL		612	87	15	85.71%	51	714

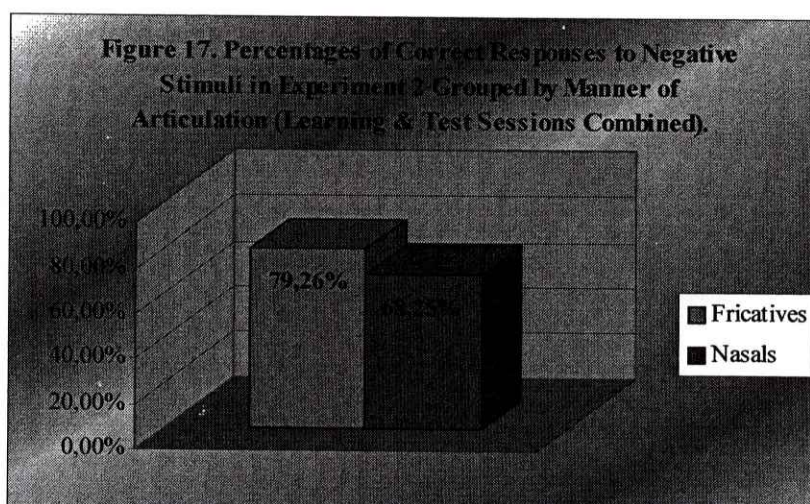


As there are very few types of phonemes for each type of plosive (two phonemes per category of point of articulation), a non-parametric Kruskal-Wallis test was used. The results of this test showed that there were not significant differences between the mean percentages of correct responses to each of the three subtypes of plosives, that is, *bilabials*, *alveolars*, and *velars* ($\chi^2_2 = 2$, P-value: $0.533 > 0.05$). In other words, the bilabials, alveolars, and velars were relatively homogeneous subsets of plosives as far as subjects' accuracy in their correct responses is concerned. In addition, when an analysis of variance was carried out on the percentages of correct responses to each type of phoneme, it was found out that the differences were not significant ($F(5, 78) = 6.14$, P-value = $0.0001 < 0.05$). However, post-hoc analyses using SNK and Duncan tests were conducted and they revealed that the phonemes /p/, /t/, /b/, /d/, and /g/ made up a homogeneous subset and the phonemes /t/, /k/, and /g/

another. However, the phoneme / d / was not homogeneous with any of those subsets. The reason for the difference between / d / and the rest of the plosives seems to indicate that subjects made more responses to this particular plosive simply because it was a positive control found in the test session in which subjects had a well-established classifying behaviour that made them err to a lesser degree.

Looking next at the negative stimuli in more detail, these were divided into two groups according to the manner of articulation with which they were produced, namely, either fricative or nasal. In this respect, table 17 shows the patterns of response to the negative stimuli of both the learning and test sessions (combined) divided by type of phoneme and type of manner of articulation. Figure 17 also represents the percentages of correct responses to the two types of negative stimuli grouped by manner of production (i.e. fricatives and nasals).

Table 17. Number of Correct/Incorrect/Null Responses and Percent Correct Responses to Negative Stimuli Grouped by Manner of Articulation and Type of Phoneme Category in Experiment 2 (Training and Test Sessions Combined).							
Manner of Articulation	Phoneme	Type of Response			% Correct Responses	Items	Number of Responses Elicited
		C	I	NR			
Fricatives	/ f /	46	22	2	65.71%	5	70
	/ v /	47	9	0	83.93%	4	56
	/ s /	55	14	1	78.57%	5	70
	/ z /	48	8	0	85.71%	4	56
	/ θ /	52	4	0	96.27%	4	56
	/ ð /	57	13	0	81.43%	5	70
	/ ʃ /	39	17	0	69.64%	4	56
SUBTOTAL		(344)	(87)	(3)	(79.26%)	(31)	(434)
Nasals	/ m /	36	17	3	64.28%	4	56
	/ n /	50	19	1	71.43%	5	70
SUBTOTAL		(86)	(36)	(4)	(68.25%)	(9)	(126)
TOTAL		430	123	7	76.78%	40	560



In order to find out whether nasals were more difficult than fricatives in view of the predicted phonetic interference that could be produced by the nasals in the experimental task, a *t* test was carried out. The results of the *t* test reveal that the differences between the percentages of correct responses to the two subset of negative stimuli (i.e. fricatives and nasals) were not statistically significant ($t(7) = 1.589$, P-value: $0.156 > 0.05$). This means that the mean percentage of responses to fricatives and nasals was relatively homogeneous or, in other words, that neither group of negative stimuli as defined by category of manner of articulation was more difficult than the other. In principle this finding seems to suggest that the predicted phonetic interference of the nasals (if any) was less important than expected. The possibility remained that there could be within-category differences regarding the types of phonemes used. So, an analysis of variance was conducted on the percentages of correct responses to the 9 types of phonemes in table 17. The results of the ANOVA showed that the differences were significant ($F^*(8, 92) = 7.94$, P-value = $0.0000 < 0.05$). A post-hoc analysis using the Duncan Multiple Range Test showed that the homogeneous subsets were: 1) the nasals and the fricatives / f ʃ s /; 2) / n / and the fricatives / f ʃ s ð /; 3) the fricatives / θ ð s z v /. Another analysis using the SNK test showed that the homogeneous subsets were 1) the nasals and fricatives / f ʃ s ð /; 2) the nasal / n / and the fricatives / ʃ s ð /; 3) the fricatives / s ð v z /; 4) the fricatives / v z θ /. These findings suggest that individual nasal phonemes are sometimes homogeneous with some of the fricatives and sometimes not and that, if nasals are considered to be interfering, so must be fricatives like / f / or / s /, which seemed to be particularly difficult in comparison with the rest of the fricative phonemes.

In an attempt to obtain further evidence on the possible impact of the nasals and, most importantly, of orthographic sources of potential interference on subjects' formation of the category, several contrasts of proportions were carried out with the data shown in table 18, namely the number of correct, incorrect, and null responses to the negative stimuli with potentially interfering phonetic features (i.e. / m / and / n /), with potentially interfering spellings (< th- > = / ð, θ /, < ps- > = / s /), or with both phonetic and orthographic potential difficulty (i.e. < kn > = / n /, < gn- > = / n /).

Table 18. Number of Correct/Incorrect/Null Responses and Percent Correct Responses to Negative Interfering Items in Experiment 2 (Learning and Test Sessions Combined).

Type of Interference	Type of Stimuli	Type of Response			% Correct Responses	Items	Number of Responses Elicited
		C	I	NR			
Phonetic	/ m /	36	17	3	64.28%	4	56
	/ n /	20	8	0	71.43%	2	28
<i>SUBTOTAL</i>		(56)	(25)	(3)	(66.67%)	(6)	(84)
Phonetic & Orthographic	< kn- > (= / n /)	19	8	1	67.86%	2	28
	< gn- > (= / n /)	11	3	0	78.57%	1	14
<i>SUBTOTAL</i>		(30)	(11)	(1)	(71.43%)	(3)	(42)
Orthographic	< th- > (= / ð, θ /)	109	17	0	86.50%	9	126
	< ps- > (= / s /)	11	3	0	78.57%	1	14
<i>SUBTOTAL</i>		(120)	(20)	(0)	(85.71%)	(10)	(140)

Several remarks should be made in relation to subjects' performance on these negative stimuli. In the first place, the percentages of correct responses to the two nasal phonemes (i.e. / m / and / n /) were not statistically significant ($P\text{-value} = 0.743 > 0.05$). The percentages between the two potentially interfering types of negative stimuli (i.e. < th- > = / ð, θ /, < ps- > = / s /) were also statistically nonsignificant ($P\text{-value} = 0.447 > 0.05$), and the difference between the two types of stimuli with potential orthographic and phonetic interference was not significant either ($P\text{-value} = 0.486 > 0.05$). The nonsignificant differences obtained show that we can actually treat the three groups as homogeneous sets. In the second place the percentages of correct responses to all the instances of each of the three subsets (i.e. "phonetic interference" group, "phonetic & orthographic" interference group, "orthographic interference" group) were compared by three different contrasts of proportions. The results of these comparisons show that the differences were not statistically significant for the "phonetic interference" and "phonetic & orthographic interference" comparison ($P\text{-value} = 0.493 > 0.05$), almost, but not significant between the "phonetic & orthographic interference group" and "orthographic interference group" comparison ($P\text{-value} = 0.059 > 0.05$), and significant between the "orthographic" and "phonetic" comparison ($P\text{-value} = 0.001 < 0.05$).

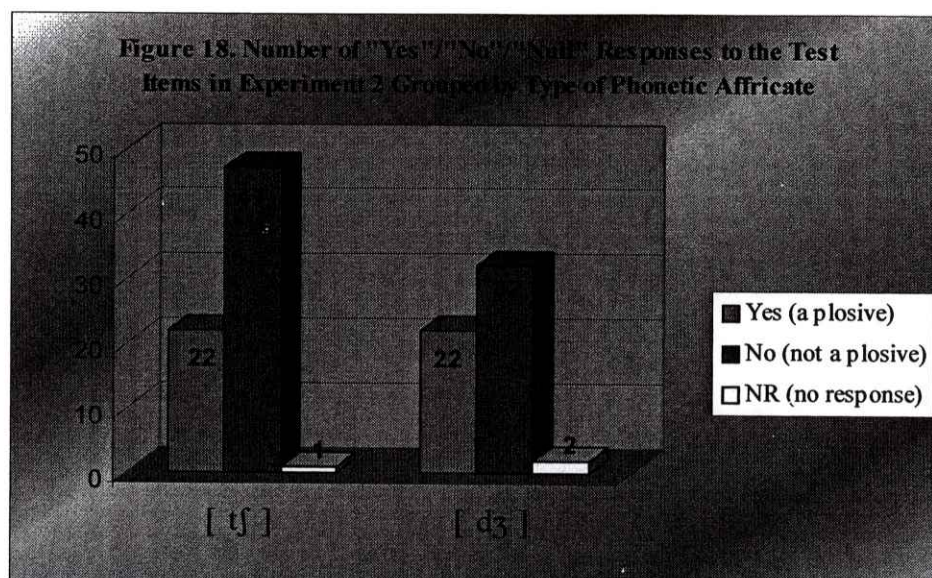
The significant difference between the percentage of correct responses to the "phonetic interference group" and the "orthographic interference group" can be interpreted as showing that, although the nasals do not seem to be excessively interfering, the orthographic interference group was decidedly *non-interfering*. This can also be inferred from an inspection of the percentages of correct responses to the phonemes / n / and / s /. As table 19 shows, / n / was

sometimes spelled “n” and at some other times “kn-” or “gn-”, in which case potential orthographic interference adds to the presumably phonetic one while / s / was spelled with “s” on some occasions and with “ps-” at other times, in which case the spelling may potentially interfere in subjects’ accuracy. However, as Table 19 shows, the percentages of correct responses to / n / was the same when it was spelled with < kn > and < gn- > (combined) than when it was spelled with < n >. This means that the presence of < kn- > and < gn- > did not cause a higher percentage of incorrect responses and that the easy or difficult status of / n / is simply a phonetic question. In addition, / s / was equally difficult when it was spelled with < s > than when it was spelled with < ps- >. This shows that < ps- > did not affect subjects’ percentage of correct responses to / s / and, consequently, that no orthographic interference was obvious.

Table 19. Number of Correct/Incorrect/Null Responses and Percent Correct Responses to Negative Stimuli Spelled with Typical Plosive Spellings or Spelled with Typical Fricative or Nasal Consonant Letters (Learning & Test Sessions Combined).						
Type of Stimulus	Type of Response			% Correct Responses	Items	No of Responses Elicited
	C	I	NR			
/ n / Spelled with < kn- >, < gn- >	30	11	1	71.43%	3	42
/ n / Spelled with < n >	20	8	0	71.43%	2	28
/ s / Spelled with < s >	44	11	1	78.57%	4	56
/ s / Spelled with < ps- >	11	3	0	78.57%	1	14

Finally, subjects’ answers to the test words will be analysed. Table 20 and figure 18 show the numbers of “Yes” (= the sound is an oral plosive), “No” (=the sound is not an oral plosive), and “Null” responses to the test stimuli in experiment 2. In addition, Table 20 shows the percentages of responses treating / tʃ / and / dʒ / as oral plosives on the one hand, and not as plosives on the other (with negative and null responses combined) as well as the number of items per type of test word, and the number of responses elicited in each case.

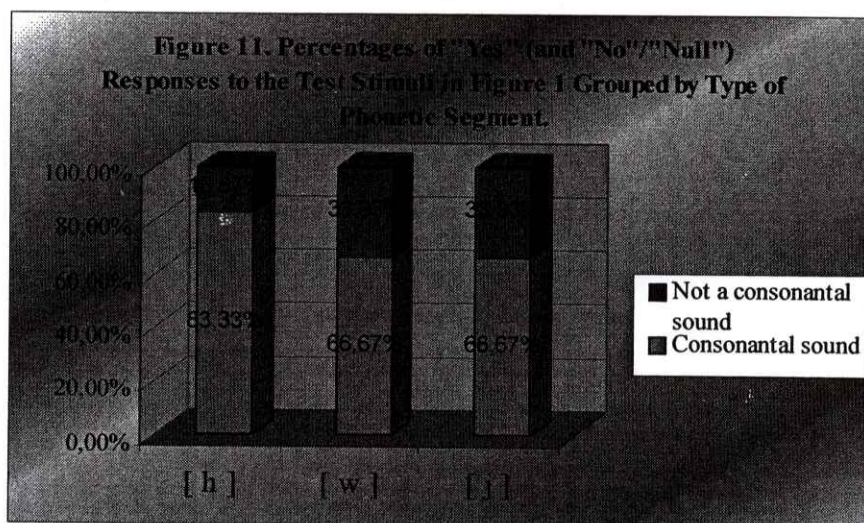
Table 20. Number of Yes/No/No Response and Percentages of Responses to Test Stimuli Grouped by Phonetic Type in Experiment 2 (Test Session).							
Type of Phonetic Affricate	Type of Response			% of Responses		Items	No of Responses Elicited
	Y	N	NR	As a Plosive	Not a Plosive/Null		
/ tʃ /	22	47	1	31.43%	68.57%	5	70
/ dʒ /	22	32	2	39.28%	60.71%	4	56
TOTAL	44	79	3	34.92%	65.08%	9	126



These data show that out of the 70 responses to stimuli containing word-initial /tʃ/, subjects responded positively 22 times, and negatively on 47 occasions while they provided no answers once. As far as test words starting with /dʒ/ are concerned, out of 56 responses elicited, subjects responded positively on 22 occasions, negatively 32 times, and provided no answers on two occasions. These data indicate that subjects were treating /tʃ/ and /dʒ/ less often as members of the category "oral plosives" than as members.

Looking now at the percentages of "Yes" and "No/Null" responses shown in table 20, which are also visually represented in figure 19, it also becomes clear that the percentages of responses considering /tʃ/ and /dʒ/ as oral plosives (31.43% and 39.28% respectively) are lower than the percentages not treating them as such (68.57% and 60.71% respectively), with negative and null responses combined. Looking at the question of whether the percentage differences between the inclusion of /tʃ/ and /dʒ/ in the category "oral plosives" and its non-inclusion (exclusion = "no" answer; undecided = null response) are significant, it was found out that the differences were statistically significant in the case of /tʃ/ (31.43% versus 68.57%, $P\text{-value} = 0.00 < 0.05$ by a contrast of proportions) and /dʒ/ (39.28% versus 60.71%, $P\text{-value} = 0.020 < 0.05$ by a contrast of proportions). The difference was also significant for the difference between the overall number of "yes" vs. "no/null" responses to the affricates (34.92% versus 65.08%, $P\text{-value}: 0.000 < 0.05$ by a contrast of proportions). In other words, subjects excluded /tʃ/ and /dʒ/ significantly more often from the category "oral plosives" than not. In addition, it should be emphasised that an analysis of the percentages of responses that treated /tʃ/ and /dʒ/ as members of the category "oral plosives" (i.e. 31.43% and 39.28% respectively) shows that there were not significant differences between the two phonemes ($P\text{-value} = 0.262 > 0.05$ by a

contrast of proportions). In short, not only were / tʃ / and / dʒ / *not* treated as members of the category “oral plosives” significantly more often than treated as such but there were not significant differences between / tʃ / and / dʒ /, which means that the pair of affricates behaved in a similar manner.



3.3.3.2. Post-Experimental Interview.

The interviews conducted after the test session of the CF task reported above permitted the experimenter to gather useful information about subjects’ ideas, thoughts, strategies, etc., while they were performing the task as well as the characteristics of the category under investigation as it was conceived by those experimental subjects. The interviews had the characteristics mentioned in 3.1.1.5., that is, subjects were asked different questions about their performance in the task including the name with which they would describe the type of sound they had to figure out, the examples they remembered, the articulatory gestures used to pronounce the sound in question, their subjective auditory impression of how the stimuli sounded, etc. Subjects had to write this information on a piece of paper, which was collected by the experimenter at the end of the post-test interview.

Table 20 shows that the positive stimuli were most commonly referred to with synonymous expressions that included the adjective “hard”: “a hard consonant” (1, RB; 2, LB; 3, PW; 5 KW; 8 RIO), “a hard sound” (9 JB; 11, KC; 12 SP), “(a) harder sound” (7, MZ), “harder consonants” (4, JR), By “hard”, “harder”, or “hardened” subjects seemed to mean that the target sounds were somehow forceful or powerful.⁴⁹ The perceived hardness of the sounds under investigation was additionally indicated by subjects 6 (NH), 12 (SP), 13 (NG), and 14

(AB), who called the target sounds “stressed sounds, hardened version of the first individual letter”, “(a) tough sound”, “stronger sounds”, and “more accented consonants” respectively. Far from their technical meanings, *stressed* and *accented* simply seem to convey the special emphasis attached to the pronunciation of plosive stops (as perceived by the experimental subjects). In general, subjects’ labelling behaviour seemed to refer to the concept of oral plosive in an abstract manner (i.e. providing a general name) but not with an exemplar-based one (i.e. mentioning specific instances of the category), except for subject 10 (EA), who called the category “words pronounced with *p, t, b, d, c,* and *k*”.

It is also interesting to note that, while subjects notoriously agreed with one another on the name given to the target sounds, they hardly ever provided a specific name for the negative stimuli. Although we might tend to think that the opposite of a “hard consonant” would be popularly be termed a “soft consonant”, this latter description was only mentioned by subject 8 (RI) and only two other subjects explicitly referred to the negative subjects calling them “words pronounced with *th, g, j,* and *s*” (10, EA), and “weaker sounds” (13, NG). So, no special emphasis was given to the opposing category when defining the positive one. It is also true, however, that the post-experimental questions did not explicitly ask for the name and characteristics of the negative category.

Apart from naming the positive stimuli, subjects provided useful information regarding the necessary articulatory actions for the production of “hard consonants”. In general, each of the 3 stages involved in the production of a plosive (or 4 if the third stage is divided into a release stage + a post-release stage) were explicitly referred to by at least a couple of subjects. First, as far as the *closing* stage is concerned, subjects said that for the production of hard consonants it was necessary to “stop the air with the teeth, tongue, or lips...” (2, LB) or to “close mouth (prior to expelling the word)” (3, PW). It is also interesting to remark that, in relation again to the closing state, subjects made a few references to the different points of articulation (i.e. bilabial, alveolar, and velar) of different types of plosives: “*p* and *b* come from the front of the mouth, *g* comes from the back” (9, JB), “pronounced with the lips, tongue, or throat” (8, RI), “lips come together for some of them (e.g. *p*)” (5, KW), or “dental, labial” (3, PW). Second, the *hold* stage was referred while claiming that in order to produce oral plosives it was necessary to “build up air before the sound is created” (3, PW) or to “keep air trapped behind the lips before saying the sound” (7, MZ); third, references to the *release* stage were made in the following expressions: “... the sound is released” (2, LB), “... (prior to) expelling the sound” (3, PW); finally, the *post-release* stage and, more specifically, the aspiration present in some of the positive items was referred to as a “short blow of air from chest..... an *h* sound after the first consonant, aspirated” (4, JR). Explicit references to the post-release were also made when

⁴⁹ The numbers given to the subjects in Table 21 do not necessarily indicate the order in which subjects were run in the experiment.



subjects described the auditory impression that the positive stimuli caused to them, which was one of “puffing” (7, MZ), of “air coming out in a rush” (10, EA), or of speech that sounded “... like a bang” (5, KW) or “like a short sigh” (4, JR). Perhaps noticing the fact that some oral plosives are aspirated while other are not, subject 7 (MZ) mentioned that some positive instances “were harder than others”. In general, from subjects’ reports of their auditory impression, we can say that the positive stimuli were perceived as “hard” (1, RB; 9, JB) or “harder” (7, MZ), adjectives also used to define the category, but also as “strong” (2, LB) or “stronger” (13 NG), “low and bass” (3, PW), “quicker” (8, RI), “harsh” (6, NH) or “harsher” (8, RI), and “louder” (12, SP).

As mentioned earlier, the release and post-release stages of a plosive seem to be related to subjects’ labelling of the target sounds as “hard” in the everyday sense of the word. A cue as to why “hardness” was important in both the definition and the auditory impressions provided comes from subject 7 (MZ), who indicated that positive instances were harder sounds “causing a puff of air to leave from the mouth”. In addition, an otherwise phonetically naive subject (4, JR) claimed, as mentioned earlier, that the positive instances implied “an *h* sound after the first consonant, present mainly after harder consonants”. It is then likely that the release and post-release stages of the plosives, sometimes accompanied by aspiration, may have played a significant role in subjects’ labelling plosives as “hard consonants/sounds”.

Looking at the examples provided by the experimental subjects it is clear that instances of the positive category (i.e. 38) outnumber the negative category (3: / *v* /, 2 instances; / *ʃ* /, 1 instance) of the opposing category (i.e. not plosives). Moreover, words containing bilabial plosives were the most frequently mentioned ones (/ *p* /: 15 examples; / *b* / 9 examples). The alveolars and the velars were mentioned almost as often (/ *t* /, 3 examples; / *d* /, 3 examples; / *k* /, 5 examples; / *g* /, 3 examples). On purely frequency grounds we could say that if a prototypical example of an oral plosive should be cited, / *p* / would be a likely candidate. The fact that negative stimuli were not very often mentioned follows subjects’ general tendency in the post-experimental interviews of providing little information about the negative stimuli, which probably amounted, for the subjects, to an ad hoc category with little psychological significance in contrast with that of “hard consonants”, which were “easier to detect” (5, KW).

In addition, it should be mentioned that subjects did not mention anything about negative interfering items with either orthographic or phonetic interference and the only example given, “kneel” by subject 9 (JB), was given as a non-example of the category under investigation. As far as the target sounds contained in the test items are concerned, only subject 5 (KW) mentioned that words beginning with “*ch* and *j* were confusing and, when defining the category, subject 10 (EA) said that the category was made up of “words pronounced with *p*, *t*, *b*, *d*, *c*, and *k* vs. words pronounced with *th*, *g*, *j*, and *s*”. The reference to *g* and *j* should be taken as an

example of the exclusion of the lenis palato-alveolar affricate (i.e. / dʒ /) from the category “oral plosives”, the letters *g* and *j* clearly referring to / dʒ /, with which those letters are pronounced.

Finally, a few remarks can be made about the six subjects who did not reach criterion in the learning session and whose motor responses and verbal reports were not taken into accounts. Three of these subjects reported that the difference between the green words and the red ones was one of “vowels, long sounds” (15), “nasal sounds at the back of the throat” (16), “length of vowels” (17). These descriptions as well as the fact that they had not reached criterion, making only 24, 34, and 30 correct responses respectively, automatically excluded them from the criterial group. There were, however, three subjects who provided a description that seemed to indicate that they had somehow realised what the differences between the positive and the negative stimuli were according to the distinction made by the experimenter. These three subjects described the positive stimuli as “harsh” (18), “sharp” (19), and “stronger” (20). However, additional comments made during the post-experimental interview failed to clearly reveal these subjects’ mastery of the category under investigation. It is then simply possible that they were likely to have struck upon the differences between the positive and the negative items after the task was over. Consequently, their answers were of lesser value and were not taken into account. In addition, these subjects did not reach the pre-established 37-correct-response criterion, making 35, 28, and 28 correct responses respectively.

Table 21. Name Given to the Category, Articulatory and Auditory Characteristics of the Category Instances, and Examples of the Category Provided in the Post-Experimental Interview.					
	Subj.	Name Given to the Category	Actions to Produce the Sound	Auditory Impression	Examples
1	RB	Hard consonant		"hard", i.e., there is a definite beginning to the word	+ < push >, + < dash >, + < put >, + < but >, + < burst >, + < puck >, + < tough >
2	LB	Hard consonant	Stop the air with the teeth, tongue, or lips before the sound is released	It sounds "strong" -useful for poetry, assonance and alliteration	+ < p >, + < k >, + < t >, + < b >, + < d >
3	PW	Hard consonant	Build up of air before the sound is created. Dental, labial. Close mouth prior to expelling the sound	It sounds "low", "bass"	+ < ball >
4	JR	harder consonants	Short blow of air from chest... an <i>h</i> sound after 1 st consonant, aspirated, present mainly after harder ones	It sounds like a short sigh	+ < call >, + < tall >, + < poor >
5	KW	Hard consonant	Lips come together for some of the green ones (e.g. <i>p</i>)	Easier to detect when it was in the category than when not It sounds like a "bang".	+ < c >, + < b >, + < p > ? < ch >, ? < j >
6	NH	Stressed sounds, hardened version of first individual letters	Round mouth/lips for some of the green ones	It sounds "harsh", like when you say the letters of the alphabet	+ < cough >, + < cell >
7	MZ	Harder sound causing a puff of air to leave from the mouth	Keep air trapped behind the lips before saying the sound	Harder, but some were harder than others Puffing	+ < pull >, + < bus >, + < goal >
8	RI	Hard consonant vs. soft consonant	Harsher attack of consonant at the beginning of word. Pronounced with lips, tongue, or throat	Hard C's: quicker, harsher. Soft C's, that allow you to "ease" into the word, more time to pronounce	+ < pull >, + < dish >, - < verse >, - < shawl >
9	JB	A hard sound	<i>p</i> , <i>b</i> come from the front of the mouth, <i>g</i> from the back	Hard	+ < push >, + < guess >, + < girl >, + < bus >, - < th >, - < kneel >
10	EA	Words pron. with <i>p</i> , <i>t</i> , <i>b</i> , <i>d</i> , <i>c</i> , and <i>k</i> vs. words pron with <i>th</i> , <i>g</i> , <i>j</i> , and <i>s</i> .	Press lips more	Air coming out in a rush	+ < push >, + < buzz >
11	KC	A hard sound			
12	SP	A hard, tough sound	Lots of movement of the mouth	Louder	+ < puff >, + < push >
13	NG	Stronger sounds vs. weaker sounds		Stronger	+ < puff >, + < bet >
14	AB	More accented consonants	Clear actions		+ < pull >, + < push >, - < verve >

3.3.4. Discussion.

The experiment reported above tried to obtain psycholinguistic evidence on the formation of the category "oral plosives" as it is understood by phonetically naive native speakers of English. The experiment also examined subjects' potential classification of the English palato-alveolar affricates as members of the category "oral plosives".

The results of both the learning and test sessions indicate that 14 out of the subjects who took part in the experiment formed the category correctly. First, these 14 subjects reached criterion in the learning session, making a mean number of 47 correct responses (range 37-59, s.d.= 7.06), a much higher number than the pre-established criterion (i.e. 37 correct responses).

Second, subjects' successful formation of the category under investigation is shown in their successful performance in the test session even without the provision of feedback. More specifically, subjects made significantly more correct responses in the test session (89.40%) than in the learning session (77.86%) and, consequently, fewer errors in the test session (10.60%) than in the learning session (22.14%). Third, subjects responded correctly to control stimuli with a high degree of accuracy, which indicates a highly acceptable level of generalisation of their achievement in the training session to unreinforced subtypes of stimuli not previously encountered. More precisely, subjects responded correctly to positive controls (i.e. those beginning with / d /) 94.65% of the time, which is the highest percentage of correct responses to any type of plosive, and they responded correctly to the single control negative stimuli (i.e. "gnash") 78.57% of the time, which is a considerable degree of accuracy if we take into account that in general it was more difficult to respond to negative stimuli correctly than to positive stimuli and that the nasals seemed to produce some sort of phonetic interference (see below). Finally, subjects' labelling of the target sounds was relatively homogeneous and their reports match the characteristics of the members of the category as described in the specialised literature. All these data indicate that the answers to the first research question (i.e. whether subjects could form the category "oral plosives") is "yes" and that the hypothesis formulated in 3.3.1., namely, that subjects could do so, is confirmed.

The results of both the learning and test sessions also suggest that there was a slight degree of phonetic interference in subjects' formation of the category and subsequent generalisation to unreinforced stimuli. More specifically, the presence of nasal consonants, in particular / m /, caused trouble to the experimental subjects: the percentage of correct responses to negative stimuli beginning with / m / was the lowest of all negative stimuli (i.e. 64.28%). Stimuli beginning with / n / seemed to be less interfering, the mean number of correct responses provided for them being 71.43% when / n / was spelled with < n >, 67.86% when it was spelled < kn >, 78.57% when it was spelled with < gn- >, and 71.43% when / n / was spelled < kn-, gn- >. The fact that the percentage of correct responses for / n / spelled < gn- > is higher than that of / n / spelled with < kn > probably has to do with the position of the sole word containing "gn-" (i.e. "gnash") in the experimental session. "Gnash" appeared very late in the test session (position 34 in the test session or 94 in the 100-stimulus list), when subjects had already had a considerable amount of information to more definitely rule out nasals as members of the category "oral plosives". Why were the nasals interfering? As subjects' report seems to indicate, one apparent reason for the interference produced by the nasal consonants is that the closure stage is an important cue for category membership and, since nasals, like stops, involve a complete obstruction of the air passage in the oral tract, this common articulatory fact between both types of sounds may have led subjects to initially treat nasals as members of the category

“oral stops”. However, subjects seemed to have eventually started to exclude nasals from the category. In fact, no subject made any reference to any degree of nasality in some of the positive stimuli in the post-experimental interview.

The results also indicate that orthography had a minor impact, if any, on the formation of the category: words beginning with / ð / and / θ /, spelled with “th”, for which some degree of orthographic interference was predicted, received very high percentages of correct responses (81.43% and 96.27% respectively). In addition, the single word spelled with < ps- > (i.e. “psalm”), with the phonological value / s /, had a percentage of correct responses (78.57%) similar to and even a little higher than the mean percentage of correct responses to negative stimuli with neither orthographic nor phonetic interference (i.e. 76.79%). This is more remarkable considering that “psalm” appeared relatively early in the CF task (position 29 in the learning session). Subjects were clearly *not* influenced by the fact that some of the negative stimuli were spelled with consonant letters that are canonical spelling forms of some of the plosive phonemes.

Finally, the results of the test session also show that subjects did not include the affricates / tʃ / and / dʒ / in the category “oral plosives” more frequently (i.e. / tʃ / = 68.57%; / dʒ / = 60.71%) than they treated them as members of that category (i.e. / tʃ / = 31.43%; / dʒ / = 39.28%) and that the differences were statistically significant. In addition, no differences were found between / tʃ / and / dʒ /, which means that the pair of affricates behaved in a similar manner (as far as their inclusion or non-inclusion in the category “oral plosives” is concerned). Thus, the answer to the second research question (i.e. whether subjects consider / tʃ / and / dʒ / as members of the oral plosive category) is that the majority of the time they exclude these items from the category “oral plosives” and so, the hypothesis entertained (that they would do so) is confirmed.

Why did subjects exclude the affricates / tʃ / and / dʒ / from the category “oral plosives” more often than not? To start with, it seems that orthography, as indicated above, had a very weak influence (if any) on subjects’ categorising behaviour. However, its possible influence in subjects’ classification of the affricates was considered. Regarding / tʃ /, there might be a reason to suspect that the dominant spelling of / tʃ / (i.e. “ch”) could have biased the experimental subjects towards considering the test items beginning with / tʃ / as members of the category “oral plosive”. After all, the digraph “ch” is a common spelling of one of plosive phonemes (i.e. / k /) in word initial position (e.g. “chemistry”, “chemical”, “chaos”, “character”, “chameleon”), word-medially (e.g. “archaic”, “archaeology”, “alchemy”), and word-finally (e.g. “monarch”, “patriarch”). In addition, / tʃ / is sometimes spelled with “t”, the spelling of another plosive

phoneme, namely / t / (e.g. “fortune”, “nature”, etc.), although never in word-initial position.⁵⁰ However, if subjects ever thought of orthography as a potential classificatory criterion, they may have probably come to the conclusion that despite sharing the same orthographic representation / k / and / tʃ / were two different sounds and that orthography by itself would not justify including / tʃ / as an instance of the category “hard consonants” in view of the small percentage of responses treating it as an oral plosive (i.e. 31.43%). Little can be said here concerning / dʒ / because, although it can also be spelled with “g” (e.g. “gin”, “gene”, “George”, “germ”, etc.), the typical spelling form of another plosive phoneme (i.e. / g /), the test items including / dʒ / in this study were all spelled with the letter “j”.

In order to more closely examine the potential role of orthography in the classification of the palato-alveolar affricates by native speakers of English, it might be interesting to conduct a similar CF study with either non-words and literate subjects or with illiterate subjects on whom formal schooling had no influence.⁵¹ Such studies might yield very different results from the ones obtained in this dissertation. For the time being and assuming that orthography, if of any relevance at all in subjects’ categorising behaviour, cannot provide the entire explanation for the classification patterns observed regarding / tʃ / and / dʒ /, other classificatory criteria should be considered. For instance, it might be the case that subjects paid attention to the phonetic makeup of the affricates themselves. This phonetic-based explanation would imply that, although phonetically naive language users have been shown to consider affricates as single units (e.g. Jaeger 1980a: experiment 4; Treiman 1985c; 1993: 241, 284), language users are not insensitive to the two-part nature of affricates like [tʃ] and [dʒ]; as Treiman suggests “the two-part nature of.... affricates... is not too subtle or too low-level to be available to ordinary people” (1993: 284). In fact, there is some evidence that seems to support this view. Under both naturalistic and experimental conditions, Treiman (1985c, 1993: 137-138, 147-149) showed that children spelled / tʃ / with the digraph “ch” but also with “t” (e.g. “cheese” = “tése”) and words reaching even a quarter of all spellings in her (1985c) experimental study and that children also spelled / dʒ / with both “g” and, most importantly, “d” (e.g. “dost” = “just”). However Treiman found out that children also spelled / tʃ / with “s” (e.g. “teasur” = “teacher”) and the same applied to / dʒ / (e.g. “losuz” = “largest”). In view of these data, Treiman suggested that children usually group affricates with fricatives that have similar places of articulation and with stop consonants that also have similar places of articulation as well as the same value of voicing. In other words,

⁵⁰ However, the digraph “ch” occasionally has the phonological value / ʃ / (e.g. “champagne”, “chamois”) although the lexical items where this is so are probably fewer than those in which “ch” stands for / k /. In addition, / ʃ / is also spelled “t” (e.g. “action”, “motion”).

⁵¹ It is important to emphasise, however, that orthographic knowledge might still influence subjects. As MacKintosh puts it (1995: 193), “when people are set the task of deliberately learning or memorizing some new information, they will not forget everything they already know just because they are acting as subjects in a psychological experiment and the experimenter tells them to. They will approach the task with suspicion and mistrust, rather than as a *tabula rasa*; they will wonder what the point of the experiment really is; they will look for *meaning*.”



beginning spellers are sensitive to the 2-part nature of affricates. Spellings point to a greater salience of the first part of the affricate but the difference is not significant. As Treiman suggests:

Children's groupings of affricates with fricatives and stops.... reflects the hybrid nature of affricates –the fact that they begin with a stop consonant gesture and end with a fricative gesture.

Rebecca Treiman (1993: 138)

To conclude this discussion, we should like to say that the present study does not support the notion that affricates are “full-time”, clear instances of the category “oral plosives” from the point of view of the phonetically naive speaker of the language. Affricates look very much like the unclear cases that experimental research on categorisation has found for other semantic categories (see section 1.2.1.3.4.). in the sense that they are not classified as members of a category like “oral plosives” far from 100% of the time. In view of these results, the door is still open for the possible belonging of affricates in still another category. Further research should clarify whether affricates are considered as fricatives or whether they deserve, according to language users' judgement, to be grouped into a different category with other affricated clusters like [tʃ] or [dʒ]. Clearly, the issue of the classification of the palato-alveolar affricates in English seems to be one of those areas in which, as mentioned in section 2.2.2.2. the use of experiments may contribute to the gathering of increasingly larger amounts of evidence that may ultimately make phonologists have more confidently held views on a given phonological problem. We shall then have to wait for further experimental evidence on the classification on the affricates.

3.4. Experiment 3. The Category “Phoneme / p /” in English.

3.4.1. Design: Rationale, Purpose of the Experiment, Research Questions, and Hypotheses.

The phoneme has been one of the most basic notions in linguistic theory (Jaeger 1980a: 26). Although sometimes rejected or considered of little value (e.g. Chomsky 1964; Chomsky & Halle 1968; Kaye 1989: 149-154),⁵² the phoneme has traditionally represented one of the building blocks of segmental phonology. In fact, almost every book on segmental phonology assumes that a given language has a given set of phonemes. However, despite its long history, there has never been general agreement on the nature of this unit. In general, the phoneme has been considered in four different ways: first, physically, as a family of phonetically similar sounds (e.g. Bloch & Trager 1942: 40; Jones 1984: 171-172, 1989: 49; Gleason 1955: 258, 261, 263) or as a physical entity made up of a bundle of invariant acoustic features (Bloomfield 1979: 79); second, phonologically, as structural units defined in terms of some (distinctive) function in a broader linguistic system (e.g. Abercrombie 1967: 86; Gimson 1980: 43-44; Hockett 1955: 74; O'Connor 1973: 65-66; Swadesh 1934: 118; Trubetzkoy 1939: 36, 41); third, in psychological terms, as mental or conceptual entities (Baudouin de Courtenay 1972; Sapir 1933);⁵³ and, finally, as mere constructs within the formal apparatus of the descriptive theory (Twaddell 1935).

The psychological view of the phoneme has persisted as a threat throughout the history of phonology and related disciplines. Not long ago, Savin and Bever (1970: 301) characterised phonemes as “primarily neither perceptual nor articulatory entities”, but rather as “psychological entities of a nonsensory, nonmotor kind”, and quite recently, Nathan claimed that phonemes are “mental images of real physical sounds” (Nathan 1999: 317). This latter phonologist is also responsible for the also recent view of the phoneme as a conceptual category instantiated by phonetically different sounds (or allophones), which speakers classify as members of the same category (e.g. Mompeán-González 1999, 2001a, in press; Nathan 1986, 1994, 1996, 1999; Taylor 1995). This mentalistic view owes a great deal to the experimental work of Jeri Jaeger (e.g. 1980a, 1980b) and John Ohala (1983, 1986). Their work, which showed that speakers actually classify various allophones of a given phoneme as members of the same category, was the first experimental verification of the traditional phonemic claims about grouping allophones together in phonological categories. Jaeger and Ohala showed that

⁵² Although generative phonology is the framework that has most conspicuously rejected the phoneme as a unit, there are generative phonologists who regard phonemes as viable phonological units for capturing relevant surface contrasts (e.g. Schane 1971).

the phoneme was a psychologically real linguistic category, rather than merely a formal theoretical device.⁵⁴

Despite the interesting experimental results obtained by Jaeger and Ohala, few laboratory experiments have continued to explicitly test whether or not language users categorise different allophones of a phoneme as being “the same” and, consequently, replicate the findings obtained by those researchers. In fact, except for these studies, since Sapir (1933), most of the evidence on this issue has been anecdotal in nature and taken from field experiences, particularly incidents occurring when teaching orthographic systems to speakers of unwritten languages (Jaeger 1980a: 19; Ohala 1983: 236).

Conducting more experiments on the ability of subjects to group allophones into phoneme categories may shed light on the assignment of allophones to phoneme categories, precisely one of the areas in phonology where a great deal of controversy has traditionally existed. For example, one of the most remarkable and problematic issues in phonological theory over which there has been long dispute has been the assignment of the voiceless unaspirated stops after tautosyllabic [s]-clusters in English (e.g. “speak”, “despise”) to phonological categories (Davidsen-Nielsen 1975: 3; Jaeger 1980a: 38-39; Lass 1984: 52). Phonologists have typically disagreed with one another on the classification of such oral stops. The problem arises because while the phoneme pairs /p/-/b/, /t/-/d/, and /k/-/g/ are clearly opposed to each other in initial position (e.g. /pɪn/ vs. /bɪn/, /tʌn/ vs. /dʌn/, /keɪm/ vs. /geɪm/), medially (e.g. /'ræpɪd/ vs. /'ræbɪd/, /'ʌtə/ vs. /'ʌdə/, /'lɒkɪŋ/ vs. /'lɒɡɪŋ/), or finally (e.g. /kæp/ vs. /kab/, /lɪt/ vs. /lɪd/, /li:k/ vs. /li:g/), these pairs are not distinguished after tautosyllabic [s]+oral stop clusters. In other words, English does not maintain the so-called fortis/lenis contrast of stops in the environment of a preceding /s/ within the same syllable and the contrast between them in that position is cancelled or neutralised (Ladefoged 1993: 48, 81; Roach 1983: 100; Wells 1982a: 53). For instance, [p^hɪt] contrasts with [bɪt] but [sprɪt] does not contrast with *[sbɪt]. Consequently, stops after tautosyllabic [s] appear to be in complementary distribution with both /p t k/ and /b d g/ and, from a structural point of view, they could be classified as instances of either of those sets with apparently equal justification. This leads to the classic question of whether the voiceless unaspirated stops after [s] should be assigned to the phonemes /p t k/ or to the phonemes /b d g/. However, there are actually four proposals as to how to classify such oral stops.

⁵³ For reviews of the history and different views on the phoneme see Twaddell (1935), Jones (1950), and more recently, Krámsky (1974). Lass (1984: 36) provides several references that are worth consulting.

⁵⁴ Another very recent and increasingly popular conceptual view of the phoneme regards it as a complex category of sounds from which people abstract a “schema”, which emerges through people’s ability to recognise what is common to the different members of a category (e.g. Bybee 1994, 1999; Langacker 1988; Taylor 1990; see also Mompeán-González, in press for a discussion of this view).

In the first place, many researchers claim that oral stops after tautosyllabic [s] are members of the phonemes / p t k /. Different criteria have been proposed to support such a view. One argument is that, as only / p t k / occur after / s / in syllable-final (coda) position (e.g. “wasp” / wɒsp /; “cast”, / kɑːst /; “bask”, / bɑːsk /) while / b d g / go with / z / (e.g. “sobs” / sɒbz /; “razzed”, / ræzd /; “bags”, / bægz /), the phonological system will therefore be simpler if the generalisation (i.e. / sp st sk /) holds at the beginning of syllables. A greater symmetry, pattern congruity or neatness of structural pattern will be obtained (Hockett 1955: 158-159, 164-165).

Another argument is that the oral stops after tautosyllabic [s] are phonetically more similar to / p t k / in initial position than to the so-called voiced ones / b d g / (e.g. Gleason 1955: 263; Pike 1947: 141; Swadesh 1934: 123; Trager & Smith 1951: 33). What might this phonetic similarity be? Apparently, oral stops after initial [s] are supposed to be more similar to initial / p t k / because the allophones of these phonemes (including those in word-initial position) are voiceless stops while / b d g / are voiced (in the phonological sense and phonetically in their prototypical realisations). As oral stops after initial [s] are voiceless, they appear to be more phonetically similar to / p t k /. More recently, it has been claimed that oral stops after / s / produce pitch perturbations that are similar to those produced by voiceless aspirated stops but distinct from those produced by word-initial [ɸ ɸ ɸ] (Caisse 1981, mentioned by Treiman 1985b; Ohde 1984). Ohde (1984), for example, measured fundamental frequency (*F*₀) in utterances containing voiceless aspirated stops (i.e. [p^h t^h k^h]), voiceless unaspirated (i.e. [sp st sk]), and devoiced stops in word-initial position (i.e. [ɸ ɸ ɸ]). He found that *F*₀ contours were nearly identical for voiceless unaspirated and voiceless aspirated stops, and both types of voiceless stops were associated with significantly higher *F*₀ values than were devoiced stops.

In addition, on the basis of the criterion of reversibility, or capacity of phonemes to appear sequentially in the opposite order, it is argued that oral stops after tautosyllabic [s] are / p t k / because when the order is reversed, / ps ts ks / are obtained but not / bs ds gs / (Fudge 1969; see also Davidsen-Nielsen 1975: 7).⁵⁵ Such an argument seems to be reinforced by Stampe's (1987) claim that when people are asked to say words like “spin” backwards, they say [nɪps], not [nɪbz], but not sometimes one, sometimes the other.

A further argument according to which stops in [s]-clusters should be classified as / p t k / comes from an analysis of such sequences into “simultaneous components”. According to Harris (1944), phonemes can be obtained as a result of the single operation of analysing



utterances into simultaneous components of different lengths which, in many cases, extend over several phonemes. Performing such an analysis, Harris arrived at / sp st sk /, three long components in which the phonemes share the common component of unvoicing and fortisness, which extends across the three bisegmental clusters. Therefore, the long components / sk / and / zg / may be found, but not */ sg / and */ zk /.

Apart from these structural and phonetic criteria (i.e. pattern congruity, phonetic similarity, reversibility, presence of common components along sequences of phonemes), there is also experimental evidence supporting the claim that the oral stops in [s]-clusters should be classified as instances of / p t k /. One source of evidence is the data obtained from spelling tests by adults and children (e.g. Fink 1974; Treiman 1985b). In this respect, Treiman (1985b) examined the way in which comparable subjects spelt syllables like [spa], [sta] and [ska] and Fink (1974) the way in which they spelt two-syllable nonsense words containing voiced or voiceless stops occurring after [s]. As expected, English-speaking adults consistently identified and spelt the oral stops as voiceless (i.e. with the letters “p”, “t”, and “k/c/q”). Consequently, these researchers argued that this spelling behaviour reflected subjects’ knowledge of English orthography. These subjects know that stops in [s]-clusters are almost always spelled with the letters “p”, “t” and “c/k/q”,⁵⁵ the same letters used to spell voiceless aspirated stops in initial position (e.g. “pet”, “tea”, “key”, “queue”, “cat”). Children, in particular those with some reading ability, also showed to behave like the adults.

Apart from the spelling tests, there is further evidence supporting the view that oral stops in [s]-clusters should be classified as / p t k /. Using a 9-point scale with a similarity-rating technique, Nearey (e.g. Derwing & Nearey 1986; Nearey 1981) played subjects pairs of words containing bilabial stops and asked them for similarity ratings. The results showed that greater similarity was rated between [p^h]-[sp] pairs than between [b]-[sp] pairs spelling-supported (e.g. “pill-spill”, “bill-spill”), and even nonsense words (e.g. “spif-pif”, “spif”-“bif”). In other words, subjects rated oral stops after / s / as more similar to initial voiceless stops than to initial voiced stops. The perceived phonetic similarity between oral stops after tautosyllabic [s] and pre-vocalic stressed realisations of / p t k / may also underlie Donegan and Stampe’s (1979: 162) claim that “*pin* alliterates perfectly with *s’pose* but not with *s’bbatical* even if they are pronounced alike with [sp]”.

However, the strongest evidence in support of the classification of oral stops after tautosyllabic [s] is the experimental data obtained from different categorisation tasks. Using

⁵⁵ Fudge claims that / sp st sk / are to be seen as a unitary phonemic element due to their limited reversibility (in particular in the case of / ts / and / ps /). On the contrary, Davidsen-Nielsen (1975) claims that the reversibility argument points towards a non-unitary interpretation of / sk / and perhaps also of / st / (see e.g. “lapse”, “traipse” “blitz”, “Ritz” and “tax”, “fax”).

⁵⁶ This generalisation holds within syllables and across syllable boundaries, except when a morpheme boundary intervenes (as in the word “disband”).



the CF technique, Jaeger (1980a, 1980b, 1986) and Ohala (1983, 1986) provided evidence that subjects consistently categorised velar stops in [s]-clusters as instances of / k /, but not / g /.⁵⁷ Jaeger, like Fink and Treiman, claimed that such findings could be due, at least in part, to orthography for the same reason. Other experiments using as heterogeneous experimental techniques as identification tasks (e.g. Sawusch & Jusczyk 1981) or the classical conditioning paradigm (Jaeger 1980a: experiment 1) provide further support to the claim that the oral stops in [s]-clusters should be classified as / p t k /. In Sawusch and Jusczyk' (1981) study, for example, subjects labelled a syllable with an initial 10-msec VOT as / ba / but, when fricative noise corresponding to / s / was added to the stimulus, subjects identified the stop as a / p /.

The second view on the phonological classification of the oral stops in [s]-clusters is that such oral stops should be treated as / b d g /. There are different sources of evidence to support this view. To start with, different perceptual experiments have shown that oral stops after word-initial [s] are perceptually indistinguishable from the so-called voiced stops in initial position. In these experiments, when the [s] is removed from words like "spy", "store", "scold", for example, and subjects are made to identify the resulting syllables as "pie" or "by", "tore" or "door", and "cold" or "gold" respectively, they identify them as "by", "door", and "gold" (e.g. Davidsen-Nielsen 1969; Lotz *et al.* 1960; Reeds & Wang 1961).

One of the reasons for the perceptual similarity between oral stops in tautosyllabic [s]-clusters and word-initial / b d g / is the lack of aspiration in both in contrast with the aspirated nature of / p t k / in word-initial position. This is actually one of the explanations given in the aforementioned perceptual studies (e.g. Reeds & Wang 1961: 80). Phoneticians and phonologists have regularly emphasised that English stops after tautosyllabic / s / are unaspirated, meaning that they are not pronounced with a following puff of air, even if the syllable carries a strong accent (e.g. Bloomfield 1978: 80; Davidsen-Nielsen 1969, 1974; Gimson 1980: 48, 151-152; Gleason 1955: 22; Jones 1984: 69, 1989: 138; Klatt 1975; Ladefoged 1993: 47-48, 84; O'Connor 1973: 132-133; Ohde 1984; Roach 1983: 30; Swadesh 1934: 23).⁵⁸

The greater physical similarity between stops in [s]-clusters and / b d g / in initial position of words is also reinforced by the fact that the realisations of / b d g / in that position

⁵⁷ Ohala made one group categorise together words containing [k^h] and the words "ghoul", "gate", "gold", and "grape" created by splicing the [s] from the beginning of the words "school", "skate", "scold", and "scrape". He then made another group categorise together words containing [g] (e.g. "game", "glitter"), [g] (e.g. "together"), and the words "ghoul", "gate", "gold", "grape" again formed by splicing the [s] from "school", "skate", etc. These words (i.e. "school", "skate", etc.) appeared intact (i.e. with the [s]) in the test sessions of both groups as test items. The results showed that the first group included them in the category in the category with [k^h] just as decisively as the second group excluded them from the category / g /, in spite of what might be considered conflicting phonetic evidence during the training session. According to Ohala, this showed that the criterion of assigning allophones to phonemes is *distributional*, not purely phonetic.

⁵⁸ However, aspiration is usually frequent in words containing a prefix with -s followed by an intuitively transparent morpheme boundary, cases with "mis-", "ex-", "dis-" (e.g. "miscalculate", "discourteous", "mistake", "disqualify"). Only in these cases can one talk about a syllable (and morpheme boundary) between [s] and the following stop (Davidsen-Nielsen 1974).



are rarely voiced during the closure stage, being instead wholly or partially voiceless (e.g. Davidsen-Nielsen 1969; Gimson 1980: 152; Lisker & Abramson 1964; Nearey 1981; O'Connor 1973: 133; Reeds & Wang 1961).⁵⁹ In short, word-initial / b d g / are devoiced (or voiceless) unaspirated stops (i.e. [b̥ d̥ g̥]), just as those after initial [s] whereas the allophones of / p t k / in initial position are voiceless aspirated stops (i.e. [p^h t^h k^h]).

In addition, voice onset time (VOT), the interval between the release burst of a stop consonant and the beginning of vocal cord voicing, is extremely similar between [b̥ d̥ g̥], [sp st sk], and [b d g], but highly dissimilar between [p^h t^h k^h] and [sp st sk]/[b̥ d̥ g̥] (e.g. Davidsen-Nielsen 1969, 1974; Klatt 1975; Ohde 1984; Treiman 1993: 141-142; see also Davis 1995; Lisker & Abramson 1964). More specifically, while long VOT signals voiceless aspirated stops, shorter VOT signal both voiceless unaspirated stops and voiced stops.⁶⁰

On the basis of these phonetic facts, it is not surprising that linguists like Twaddell (1935: 30-31), Bloch and Trager (1942: 44), or Schane (1968: 711) explicitly claimed that the decision to assign oral stops in [s]-clusters to / p t k / has to be made quite arbitrarily as phonetic similarity might justify the inclusion of the stops with the voiced stops. Not surprisingly, Davidsen-Nielsen (1969) claims that following the criterion of phonetic similarity / sb sd sg / is a legitimate analysis and according to Roach, "there could be a strong argument for transcribing them as / sb sd sg /" because word initial / b d g / are unaspirated, / p t k / are aspirated and that / sp st sk / are unaspirated (Roach 1983: 100).⁶¹ However, the phonetic similarity-based solution is seldom adopted perhaps because of the convenience of continuing the traditional conventional spelling (Bloch & Trager 1942: 44; Hubbell 1950: 21; Pike 1947: 141; Roach 1983: 100; Twaddell 1935: 30-31). Finally, the / b d g / solution has been strongly criticised by Donegan and Stampe (1979: 173) on the basis of its logical argumentation.⁶²

Finally, another reason why / b d g / might be supported as a valid transcription of the oral stops in [s]-clusters is that, at some particular point during development, some children consistently classify these stops as / b d g /, as revealed by spelling tests (e.g. Fink 1974;

⁵⁹ It is usual to label stop consonants as voiced if they are characterised by the presence of glottal buzz during the interval of articulatory closure, while absence of buzz during this interval is a mark of voiceless stops. However, while this difference is adequate basis for the physical separation of stop categories in many languages, it works only in part in English. Although in medial position English / b d g / are usually voiced and / p t k / voiceless, in initial position both sets are commonly produced with silent closure intervals and should therefore be classed as voiceless.

⁶⁰ Studies have found out that VOT lasts between 60 and 100 msec for voiceless aspirated stops, and between 10 and 40 msec for both voiceless unaspirated and voiced stops. Specific values depend on the phonetic character of the following sound (vowel, liquid, approximant), the place of articulation of the stop, and the occurrence of the word in isolation or in context.

⁶¹ It has also been frequently mentioned that / p t k / are fortes, or strongly articulated while / b d g / and oral stops in [s]-clusters are lenes or weakly articulated (e.g. Bloch & Trager 1942: 43-44; Bloomfield 1979: 99; Pike 1947: 140; Swadesh 1934: 119; Twaddell 1935: 30-31). This again would make oral stops in [s]-clusters more similar to / p t k / than to / b d g /. However, the duration of the hold stage of oral stops, one of the supposedly important measures of the force of articulation criterion (long hold and short hold stages being associated with fortes and lenes consonants respectively) has not revealed such differences. Davidsen-Nielsen (1969) did not reach any conclusion regarding force of articulation in English from the duration of the hold stage as the average duration of the hold stage of [p^h t^h k^h] in initial position was 11 cs, 10 cs in the case of [b̥ d̥ g̥] and 7 cs for [sp st sk].

⁶² Donegan and Stampe claim that to argue that the oral stops in [s]-clusters are / b d g / because after electronic removal of [s] the residue is heard as / b d g / rather than / p t k / is a reasoning on a par with claiming that lizards are snakes because if you cut off their legs people will think that they are snakes.



Treiman 1985b) and observation of naturally-occurring spellings (e.g. Read 1971, 1975, 1986) in beginning spellers. The consensus obtained from these studies is that children not yet influenced by standard orthography sometimes spell clusters like [sp st sk] as “sb” “sd” and “sg” (e.g. “sbek” –speak–, “sda” –stay–, “sgie” –sky–). In doing so, their spellings either represent low-level phonetic characteristics not reflected in standard English spelling or a phonological system somehow different from adults’ but towards which they will eventually accommodate; as a result of exposure to the written system and familiarisation with it, children start producing standard spellings and are increasingly more likely to spell stops after [s] in the conventional manner. In addition, the decrease in unconventional spellings is closely tied to a child’s reading ability than to his/her age.

The third alternative to the classification of oral stops in [s]-clusters is that these stops instantiate neither / p t k / nor / b d g /, but a third category.⁶³ For example, Twaddell (1935: 48-49) claims that stops after tautosyllabic [s] are not members of any of the (macro)phonemes / p t k / or / b d g / because such oral stops are articulatory complexes that do not have *all* the characteristics that all the different realisations (or “microphonemes”) of those (macro)phonemes share. Consequently, for Twaddell, oral stops in [s]-clusters are different phonemes, although he does not provide a specific symbol for them.

Also, Hockett (1955: 165), who was not completely satisfied with his earlier “pattern-congruity” solution, proposed another alternative analysis similar to the previous one suggested earlier by Harris (1944). This consists in dividing / sp st sk / horizontally into simultaneously-occurring components. In this way we obtain / SP /, / ST /, or / SK / + / H / (i.e. voicelessness) where / P T K / are neither / p t k / nor / b d g / but voicing-irrelevant stops.⁶⁴

The most famous “third-category” view is perhaps that maintained by Prague School-oriented phonologists. These researchers claim that when a contrast between a pair of phonemes which exists in other places in the language is suspended in a given context, the contrast is neutralised and the phonological unit that occurs in that position is an archiphoneme. Archiphonemes are phonological units that share the features common to the phonemes involved in the neutralisation but they are neither of those units. As there is no contrast or opposition between / p t k / and / b d g / after initial / s /, the contrast is neutralised and the oral stops after that initial / s / in words like “spin”, “stay”, or “sky, are neither / p t k / nor / b d g /, but the archiphonemes / P T K / (e.g. Akamatsu 1988: 285, 299, 302; Trubetzkoy 1969: 210; see

⁶³ According to Bloch and Trager (1947: 49) such a solution is wise when the object of the classification is to exhibit in detail not only the possibilities of contrast, between phonemes, but also the positions where particular contrasts are suspended.

⁶⁴ This solution is different from Harris’s as the latter treated such sequences as / sp st sk / with the second member of the component being identified as / p t k / while Hockett treats / SP /, / ST /, or / SK / + / H / as indivisible components.

also Lass 1984: 52 for a discussion) so that the phonological transcription of those lexical items would be / sPm sTeɪ sKaɪ /.⁶⁵

Apart from theoretical discussions, the archiphonemic (or archisegmental) solution has been supported (and criticised) on the basis of speech error data (Davidsen-Nielsen 1975, 1978; Stemberger 1983), and hinted by spelling data (Treiman 1985b).

As Stemberger (1983: 12) claims, the main prediction of archiphonemes (or archisegments) is “that there will be variation associated with an archisegment”.⁶⁶ However, Fromkin (1973: 23-24) argues against archisegments, stating that their predictions for speech errors are wrong. For example, Fromkin reasons that if the / s / disappears from [s]+oral stop clusters, a / t / should result half the time and a / d / the other half. However, she claims, no slips reported in the literature or in her corpus reveal the voiced obstruants (e.g. long and strong → *trong* and *slong* (not **drong*); steak and potatoes → *spake* and tomatoes (not **domatoes*); etc.). On the contrary, Davidsen-Nielsen (1975) induced subjects to make speech errors while pronouncing invented words like “gaspate”, “maskate”, “kaspate”, etc. As the oral stops, when moved out of the position of by a slip of the tongue, are thereby disambiguated and emerge as either [p^h t^h k^h] or [b̥ d̥ ɡ̊], Davidsen-Nielsen claimed that speakers encode a voicing-irrelevant archisegment / P T K /. The distribution of these two-stop series is regulated in the following way: [p^h t^h k^h] occur if the interfering segment, i.e. the segment which is interchanged with the archisegment or which attracts the [s] is unvoiced, and [b̥ d̥ ɡ̊] occur if the interfering segment is voiced. Davidsen-Nielsen’s findings have been strongly criticised (e.g. Stemberger 1983: 12) because the study relied on an experimental task of unknown constraints, with no controls of any sort, and the number of errors induced was fairly small. For these reasons Stemberger (1983: 12-13) conducted a larger study. In it, out of 31 errors involving deletion of the [s], / p t k / occurred 28 times (e.g. “who *tole* (stole) the spoon?”) and / b d ɡ / only 3 times (e.g. “in your really *gruffy* (scruffy) clothes). This, Stemberger suggested, could be interpreted in two ways. First, we could assume that archisegments predict that voiced and voiceless stops will be equally frequent, and use this data to reject the hypothesis that archisegments exist. The few errors can be accounted for, for instance, as feature errors. However, it may also be the case that the fact the archisegment usually disambiguates as a voiceless stop again simply because voiceless stops are far more frequent than voiced stops in spoken English. This explanation was stronger, according to Stemberger, in the analysis of speech errors in which / s / is added to stops (e.g. “.... is that *sprices* (prices) are still expensive”). Again, Stemberger found more errors involving / p / added to voiceless stops (48

⁶⁵ For the different manners and criteria used to represent archiphonemes throughout the history of phonology see Akamatsu (1988: 314-331).

⁶⁶ This prediction is different from the view that oral stops after tautosyllabic / s / is neither / p t k / nor / b d ɡ / (e.g. Hockett 1958: 109), a view strongly criticised by Akamatsu (1988: 310).



vs. 27) but taking into account the different frequencies of both fortis and lenis stops in spoken English, he claimed that / s / was as likely to be added to / p t k / than to / b d g / and that these data showed that stops after tautosyllabic [s] were neither / p t k / nor / b d g / but the archisegments / P T K /.

Finally, the fourth position on the categorisation of oral stops after tautosyllabic [s] does not require a segmental (bisegmental or unisegmental) interpretation of such clusters. Rather, it makes the phonemic problem disappear by adopting a binary distinctive feature approach allowing for unique phonological representations of contrastive morphemes. The phonological representation of those stops specifies only distinctive features and leaves blank any features which can be predicted and filled in by redundancy rules. In this respect, syllable-initial aspirated stops, unaspirated devoiced stops, and unaspirated stops after tautosyllabic [s] are [+consonantal], [-vocalic], [-continuant], [-nasal], [-compact], and [+grave]. However aspirated stops are [+tense] whereas voiced stops are [-tense] and stops in [s]-clusters are unspecified as far as the feature of tenseness, which includes aspiration, is concerned. Finally, the feature [voice] is left unspecified for the three of them (Jakobson *et al.* 1952: 6-39; Schane 1968).

Having reviewed the extensive body of literature related to the particular phonological problem under investigation, it is clear that this issue has mainly been explored on a theoretical basis. The claim that the voiceless unaspirated oral stops in [s]-clusters are members of a given phoneme category has seldom been experimentally tested. Exceptional “external” evidence is to be found in the CF experiments by Jaeger (1980a, 1980b, 1986) and Ohala (1983, 1986), in the identification tasks by Sawusch and Jusczyk (1981), the spelling tests by Fink (1973) and Treiman (1985b), and in the speech error data reported by Davidsen-Nielsen (1975) and Stemberger (1983). Except for the interpretations by Davidsen-Nielsen and Stemberger of their speech data, the aforementioned experimental studies seem to indicate that at least for adult speakers, the most plausible solution is / p t k /. However, more research is needed to support such a view. In the first place, Jaeger and Ohala’s studies which used real English words, only looked at the classification of oral “velar” stops in [s]-clusters leaving aside the question of whether equivalent bilabial or alveolar oral stops elicit the same pattern of responses. Sawusch and Jusczyk used / p /, but theirs were synthesised stimuli, which are very useful to control for certain acoustic characteristics but which may, on the contrary, lose many of the important characteristics of real speech utterances. Finally, Fink (1974) and Treiman (1985b) provided spelling data supporting the / p t k / solution for adult speakers, but they used non-words.

The purpose of the present investigation was to provide further evidence on two specific issues: speakers’ ability to classify different allophones of the same phoneme as instances of the same category and the classification of oral bilabial stops after tautosyllabic [s]. As Wells claims (1982a: 53), the classification of bilabial stops after tautosyllabic [s] is a case in which



“psycholinguistic experiments could lead to a preferred solution”. In the present experiment, the phoneme / p / was chosen to be the focus of the experiment for several reasons. First, the status of / p / with the concept formation technique has not been previously carried out and it may shed further light on the categorisation of oral stops in [s] clusters. In addition, / p / seems to be one of the most important consonants in the initial linguistic stages in infants and children (Jakobson 1968: 69-70). Finally, / p / has a wide distribution in English as it can appear at the initial, medial, and final positions of words and in many different phonetic contexts (in the onset and coda positions of syllables, in the environment of different vowels and in different consonantal clusters). In other words, / p / has then a great variety of subphonemic variation.

The specific research questions that the present experiment addressed were:

- 1) Can speakers categorise different allophones of / p / as members of the same category?.
- 2) If so, how do they consider the voiceless unaspirated bilabial stop after tautosyllabic / s /?.

The first hypothesis is then that the experimental subjects will be able to classify different allophones of / p / as members of the same category. This hypothesis is based on the fact that a phoneme-sized categorisation has been extensively shown to be easily available to subjects in phoneme monitoring experiments, absolute identification and discrimination tasks as well as, to some extent, in concept-formation tasks. Therefore, this task may be relatively easy for the subjects. The second hypothesis is that they will consider the voiceless unaspirated bilabial stop as a member of / p /. The second hypothesis is based on the fact that most of the experimental psycholinguistic literature supporting any of the four different views favours the claim that subjects regard oral stops in [s]-clusters as members of the phonemes / p t k /.

Underlying the whole study is the operationalisation of the phoneme as a conceptual category instantiated by phonetically different sounds that language users classify as instances of the same sound type or category.

3.4.2. Method.

3.4.2.1. Subjects.

Twenty subjects between the ages of 20 and 34 (mean age 23) took part in this experiment. The subjects had been selected randomly from the 80-subject group described in

3.1.1.2. There were 10 men and 10 women. Subjects were paid for their participation and reported having no hearing problems.

3.4.2.2. Stimuli.

The stimuli of the present investigation consisted, like in the previous two experiments, of 100 monosyllabic English words. These exemplified 12 different canonical syllable structures. These stimuli had a mean duration of 650 msc.

In the learning session, the positive instances (32 items in total) had different syllable structures: VC (2 items), CV, CCV, VCC, CCVCC, and CVCC (4 items each), CVC and CCVC (5 items each). There were 16 examples of pre-nuclear / p / and 16 post-nuclear ones. The phonetic environments in pre-nuclear positions were [p^hV-] (8 items), [p^h_ɪV-] (4 items), and [p^h_ʊV-] (4 items). The phonetic contexts in post-nuclear position were [-Vp] (8 items), [-Vmp] (3 items), [-Vp^ot] (2 items), [-Vsp] (2 items), and [-Vpθ] (1 item). Negative items (28 in total) exemplified different canonical forms of syllable structure: VC (2 items), CCV and CCVC (3 items each), and CVC, CV, VCC, CCVCC, and CVCC syllables (4 items each). There were 22 non-interfering negative words, including no phonetic realisation of / p /. There were 3 orthographically interfering (i.e. “sphere”, “graph”, and “psalm”) and 3 phonetically interfering (i.e. “bay”, “blast”, and “bet”). Orthographic interference derived from the fact that the letter “p”, the paradigmatic representation of / p / is usually pronounced / f / in the digraph “ph”, and it is utterly silent in the digraph < ps > in word-initial position.⁶⁷ Phonetic interference may be caused by the presence of / b / in word-initial position, which is, in principle, the most phonetically similar to most instances of / p / as it is also an oral bilabial stop that is partially or wholly devoiced (i.e. [b̥]).⁶⁸

In the test session, there were 19 positive instances, 12 negative ones, and 9 test tokens. The syllable structures of the positive stimuli were CV, CCVCCC, CVCC (1 instance each), CCV, CVC, CCVC, CCVCC (3 instances each), and CVCC (4 instances). There were 10 pre-nuclear / p /'s and 9 post-nuclear / p /'s. The phonetic environments in pre-nuclear position were [p^hV-] (4 items), [p^h_ɪV-] (3 items), [p^h_ʊV-] (2 items), [p^h_jV-] (1 item); in post-nuclear position the contexts were [-Vp] and [-Vmp] (3 items each), [-Vp^ot], [-Vpst], and [-mps] (1 item each). 17 items had a syllable structures already encountered in the training session.

⁶⁷ All positive stimuli in this experiment were spelled with < p >. < pp > is also a possible spelling of / p /, although it is less frequent (e.g. “apply”, “suppose”, “Lapp”). Although “ph” is a very frequent spelling of / f / (e.g. “phantom”, “physics”, “graphic”, “Ralph”, etc.), in the word “Stephen” the digraph “ph” stands for / v /, and variably with / f / in “nephew”. “Ph” has the phonological value / p / in “shepherd”, and variably with / f / in “diphtheria”, “diphthong”, or “naphtha”. In addition, “p” is a silent letter in the word-initial digraphs “ps” (e.g. “psychology”, “psoriasis”), “pt” (e.g. “Ptolemy”, “pterodactyl”), and “pn” (e.g. “pneumatic”, “pneumonia”). < p > is silent in some words like “cupboard”, “raspberry”, “Thompson” or “receipt”.

⁶⁸ In studies using the phoneme monitoring technique (Dell & Newman 1980; Newman & Dell 1978; Stemberger *et al.* 1987), in which the phoneme to be detected was / p /, words containing / b / created some phonetic interference.



There were 2 positive control tokens whose syllable structures (and phonetic context in which / p / was realised) had not been previously encountered (1 CVCCC -“lapsed”-, and 1 CCVCCC -“glimpse”-) and 1 positive control whose phonetic context (but not its syllable structure) had not been encountered by subjects (i.e. [p^hjV-]). Thus positive items included all the phonetic contexts presented in the learning session as well as a new one, a palatalised [p^j] as in “pew” in pre-nuclear position and two instance followed by the fricative [s] (i.e. [-mps] and [-Vpst]). Negative items exemplified the following syllable structures: CVC, CCV, CCVC, CCVCC, CVCC, and CV (2 items each). There were 8 non-interfering items, 2 orthographically interfering (i.e. “phone”, “nymph”) and 2 phonetically interfering (i.e. “bear”, “slob”). One of the phonetically interfering was a negative control (it included an allophone of / b / not previously encountered, that is, in word-final position).

Test tokens exemplified the following syllable structures: CCVCC, CCV, CCCV, CCCVC (2 instances each), and CCVC (1 instance). The phonetic contexts (always in pre-nuclear position) were [spV-] (5 items), [sp^hjV-] (2 items), and [sp^jV-] (2 item).

The positive stimuli of both the learning and the test sessions exemplified many of the possible phonetic realisations of / p / but not all.⁶⁹ / p / was instantiated by strongly aspirated realisations followed by vowels or by devoiced [ɸ], [ɸ̥], and [j̥], with simultaneous [ɹ], [l] and [j] articulations respectively. These represent the ideal third stage of a plosive (i.e. the “release” stage), characterised by a sudden audible oral central release of air either in the form of aspiration or as an immediately following vowel. Aspiration is found most notably before a stressed vowel in the same word, that is, in initial stressed positions followed by vowel (e.g. “pet”) or by / l r w / (in these latter cases aspiration is manifested in the devoicing of / l r j /). However, other realisations of / p / were weakly aspirated instances in final position preceded by vowels, by [m] (in which case the oral closure slightly precedes the velic closure), and by the fricative [s]. It was also followed by the fricatives [s] and [θ] or by the plosive [t], in which case / p / has an inaudible release (i.e. [-Vp^ot]). Despite their phonetic differences, all

⁶⁹ The fact that stimuli were monosyllabic single words and spoken at a normal rate restricts the range of possible phonetic realisations of / p /. These include different manifestations in non-stressed syllables, like weakly aspirated stops followed by vowels (e.g. “lepper” [ˈlepə], “polite” [pəˈlaɪt]), labio-dental realisations when / p / is immediately followed by / f / or / v / as in “hopeful” or “cup-full”, short bilabial fricative sounds when / p / is not completely formed, as in “super” [ˈsʊfə], etc. / p / can also be realised with a delayed release when followed by homorganic oral stop (/ p / or / b /) as in “top people”, “top boy”, or with nasal plosion (i.e. [p̚]) when it is followed by its homorganic nasal, that is, / m / (e.g. “topmost”, “cheap meat”) or with a slightly different nasal release followed by a non-homorganic nasal (e.g. “cheap nuts”). In word-final position preceding silence or preceded by a vowel or consonant (e.g. “cap”, “help”), / p / is weakly aspirated or unaspirated, produced with inaudible release or with strong aspiration (typical of emphatic speech), with glottal reinforcement (e.g. [ʃɪp̚]), or with a single glottal stop (e.g. [ʃɪʔ]). The glottal variants also occur before consonants, as in “apse” or “hopeful”. For details about all these allophones see Abercrombie (1967: 142-147), Bloch & Trager (1942: 31-34), Gimson (1980: 46, 151, 155-159, 167-169), Jones (1989: 138-139), Kleidler (1989: 139), Ladefoged (1993: 49-52, 85), Lass (1984: 20), O'Connor (1973: 130-136), Roach (1983: 44).



the realisations of / p / in the present experiment can be described as voiceless bilabial plosives.⁷⁰

Table 22 offers a summary of the types of stimuli used and table 23 the actual list of words used and their category status.

Table 22. Types of Stimuli and Number of Items per Subtype of Stimuli Used in Experiment 2.				
Type of Stimulus	LEARNING SESSION		TEST SESSION	
	Subtype of Stimulus	Number of Items	Subtype of Stimulus	Number of Items
POSITIVE ITEMS	<i>Pre-nuclear</i>		<i>Pre-nuclear</i>	
	[p ^h V-]	8	[p ^h V-]	4
	[p ^h _ɪ V-]	4	[p ^h _ɪ V-]	3
	[p ^h _ɪ V-]	4	[p ^h _ɪ V-]	2
			[p ^h _ɪ V-] ⁷¹	1
	<i>Post-nuclear</i>		<i>Post-nuclear</i>	
	[-Vp]	8	[-Vp]	3
	[-Vmp]	3	[-Vmp]	3
	[-Vp ^o t]	2	[-Vp ^o t]	1
	[-Vsp]	2	[-Vp st]	1
	[-Vpθ]	1	[-mps]	1
	TOTAL	(32)	TOTAL	(19)
NEGATIVE ITEMS	<i>Non-interfering:</i>		<i>Non-interfering:</i>	
	Any other type of consonant	22	Any other type of consonant	8
	<i>Interfering:</i>		<i>Interfering:</i>	
	Orthographically < ps- > (= / s /),	1	Orthographically < ph > (= / f /)	2
	< ph > (= / f /)	2		
	Phonetically [b̥]	3	Phonetically [b̥]	2
	TOTAL	(28)	TOTAL	(12)
TEST ITEMS	----	---	[spV-]	5
			[sp _ɪ V-]	2
			[sp _ɪ V-]	2
			TOTAL	(9)

⁷⁰ Though still present as an articulatory movement, the release stage is not accompanied by release of compressed air (Abercrombie 1967: 146-147; Gimson 1980: 157; Ladefoged 1993: 84; O'Connor 1973: 133). The closure stage of the second stop takes place during the hold stage of / p /, so that when the release stage of / p / comes there can be no burst of escaping air or explosion, with its accompanying noise, since the stricture for the second stop prevents this, the air held back by the [t] closure. When it is necessary to indicate whether a stop is unexploded or unreleased the diacritic [°], which is not an official IPA symbol, can be used (Abercrombie 1967: 150). We can also mark an unreleased with the diacritic [˘] (e.g. [p˘]).

⁷¹ Underlined allophones in bold letters represent control stimuli.

Table 23. Stimulus List for Experiment 3: Category / p /					
Order	Stimulus	Positive (+) Negative (-)	Order	Stimulus	Positive (+) Negative (-)
LEARNING SESSION					
1.	pet	+	53.	print	+
2.	sell	-	54.	fee	-
3.	up	+	55.	pond	+
4.	egg	-	56.	end	-
5.	pay	+	57.	grant	-
6.	plea	+	58.	top	+
7.	drip	+	59.	fist	-
8.	die	-	60.	trap	+
9.	apt	+	TEST SESSION		
10.	tray	-	1.	pit	+
11.	priest	+	2.	pear	+
12.	depth	+	3.	prow	+
13.	drill	-	4.	sheet	-
14.	path	+	5.	plane	+
15.	ape	+	6.	spend	test
16.	old	-	7.	near	-
17.	drift	-	8.	slow	-
18.	golf	-	9.	clamp	+
19.	pie	+	10.	pulse	+
20.	fish	-	11.	bear	- int
21.	pray	+	12.	cap	+
22.	ash	-	13.	spa	test
23.	bay	-int	14.	ground	-
24.	place	+	15.	prayer	+
25.	opt	+	16.	false	-
26.	stamp	+	17.	drop	+
27.	sphere	-int	18.	spy	test
28.	post	+	19.	glimpse	+ control
29.	graph	-int	20.	spoon	test
30.	blast	-int	21.	prince	+
31.	shop	+	22.	phone	- int
32.	east	-	23.	paste	+
33.	pea-p	+	24.	ship	+
34.	play	+	25.	sly	-
35.	self	-	26.	lapsed	+ control
36.	psalm	-int	27.	slob	- control (int)
37.	proud	+	28.	sponge	test
38.	sea-see	-	29.	plot	+
39.	asp	+	30.	spray	test
40.	clasp	+	31.	cross	-
41.	dry	-	32.	tramp	+
42.	damp	+	33.	sply	test
43.	clean	-	34.	nymph	- int
44.	keep	+	35.	spring	test
45.	paw	+	36.	pure	+ control
46.	act	-	37.	lamp	+
47.	bet	-int	38.	rapt	+
48.	trust	-	39.	split	test
49.	plough	+	40.	stealth	-
50.	group	+			
51.	fond	-			
52.	imp	+			

3.4.2.3. Procedure.

3.4.2.3.1. Instructions.

The instructions were very similar to the ones employed in experiments 1 and 2. However, there were some variations in the second, third, and fourth paragraphs. The variations consisted in replacing the references to the occurrence of the target sound in word-initial position with a reference to the fact that the target sounds could appear anywhere in the word since the instances exemplifying the to-be-learned category appeared in different phonetic contexts. In addition, it was considered that description of the to-be-learned concept, “a certain type of consonantal sound” was more appropriate because a more general description (e.g. “a certain type of sound”) might have led subjects to look at differences between vowels in the stimuli presented given the relative disadvantage of not having their attention focused on a specific position in the word. These modified paragraphs are reproduced here for convenience:

You are going to be listening to a list of words over the headphones. Some of these words contain a certain type of consonantal sound somewhere in the word (at the beginning, in the middle, at the end) while other words do not have this sound anywhere in their pronunciation. Your task is to pay attention to the words and try to figure out the type of consonantal sound that some words have in some part of their makeup and others lack.

At first, you will not know which type of consonantal sound we are talking about. Here is how you can figure it out. There are two rectangles, a green one and a red one on the screen of the computer. There are also two keys or buttons, a green one and a red one, on the keyboard. Both the rectangles and the keys mean the same. Green means “the word contains the type of consonantal sound somewhere in the word”. Red means “the word does not contain the type of consonantal sound anywhere”.

Whenever you hear a word, one of the two rectangles on the screen will remain on display while the other will disappear after a few seconds. When the word contains the type of consonantal sound, the green rectangle will stay but the red one will disappear. When the word does not contain the sound, the red rectangle will stay but the green one will disappear. Rectangles disappear automatically. (Every rectangle that disappears will eventually reappear just before a short tone which will indicate that a new word is about to come).

The training session began, again, only when the experimenter had guaranteed that subjects had understood the instructions reasonably well. This was inferred from the subjects’ paraphrasing of the contents of the instructions given and a discussion with the experimenter of what their task was would involve.

3.4.2.3.2. Training Session, Test Session, and Post-Experimental Interview.

All experimental events in the training and test sessions proceeded as described in 3.1.1.5. Subjects were run individually in the same sound-treated room as in the previous experiments. After the computer-based task finished, a post-experimental interview was conducted. No incidences were registered.

3.4.3. Results.

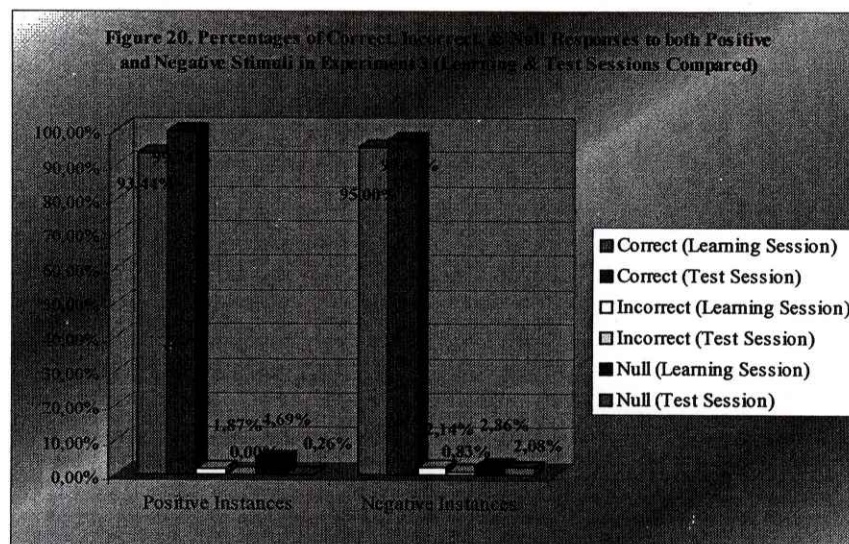
3.4.3.1. Training and Test Sessions.

Of the 20 subjects who participated in experiment 4 all were considered to have formed the category correctly as they reached the 37-correct response criterion pre-established for deciding that subjects' categorising behaviour had not been random in the learning session (P -value $0.03 < 0.05$). The average of correct responses in that session was 56.55 correct responses (range 51-60, s.d.= 2.06).

The number of correct (C), incorrect (I), and null responses (NR) to both positive and negative stimuli and percentages of correct responses to each stimulus type in the learning session are shown in Table 24. The table also shows the number of items per type of stimulus, and the number of responses elicited, which results from multiplying the number of items by the number of criterial subjects. The same information obtained from subjects' performance in the test session is shown in Table 25.

Table 24. Number of Correct/Incorrect/Null Responses and Percent Correct Responses to Positive and Negative Stimuli in Experiment 3 (Learning Session).						
Stimulus Type	Type of Response			% Correct Responses	Items	Number of Responses Elicited
	C	I	NR			
Positive Instances	598	12	30	93.44%	32	640
Negative Instances	532	12	16	95%	28	560
TOTAL	1130	24	46	94.17%	60	1200

Table 25. Number of Correct/Incorrect/Null Responses and Percent Correct Responses to Positive and Negative Stimuli in Experiment 3 (Test Session).						
Stimulus Type	Type of Response			% Correct Responses	Items	Number of Responses Elicited
	C	I	NR			
Positive Instances	379	0	1	99.74%	19	380
Negative Instances	233	2	5	97.08%	12	240
TOTAL	612	2	6	98.71%	31	620



A close comparison of above tables shows that the percentages of correct responses to both positive and negative stimuli substantially increase in the test session as compared with those in the learning session. More specifically, correct responses were significantly more frequent in the test session than in the learning session (94.17% versus 98.71%; $P\text{-value} = 0.000 < 0.05$ by a contrast of proportions). Subjects' overall better performance in the test session can also be observed in figure 20, in which an increase of correct responses to positive stimuli and a decrease of incorrect and null responses to positive and negative stimuli from the learning session to the test session are observed. All of these data indicate that subjects were performing better in the test session than in the learning session, where they were already doing very well.

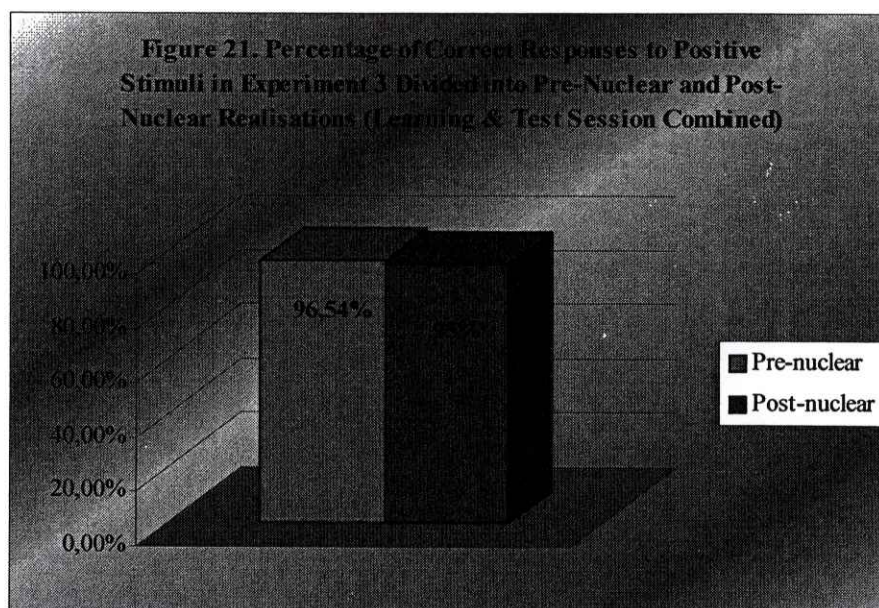
Looking at possible relationships between the variable "type of stimulus" and "type of response", it was found out that, in the learning session, there was not a significant relationship between type of stimulus (i.e. negative, positive) and type of response (correct, incorrect, null) ($P\text{-value} = 0.25 > 0.05$ by a test of independence or chi-square test). In the test session, the expected minimum frequency was below 5 and so the test of independence was not reliable.

Collapsing the values of the variable "type of response" (i.e. correct, incorrect, and null response) into two categories (i.e. correct and incorrect/null), it was shown again that there was not a significant relationship between type of stimulus and type of response in the learning session ($P\text{-value}: 0.30 > 0.05$ by a test of independence using the statistic Yate's corrected chi-square test). However, in the test session a significant relationship was found between type of stimulus and type of response ($P\text{-value}: 0.01 < 0.05$ by a test of independence using again the statistic Yate's corrected chi-square test). More specifically, focusing on the corrected residuals of the test, it became clear that positive stimuli were in inverse proportion to the number of

incorrect/null responses and in direct proportion to the number of correct responses while negative stimuli were in inverse proportion to the number of correct responses and in direct proportion to the number of incorrect/null responses. This indicates that it was harder to be right with negative stimuli than with positive stimuli, which, again, may be explained in terms of the poorer performance in a concept learning task with negative stimuli as compared with positive stimuli (see sections 3.2.3.1., & 3.3.3.1).

Table 26 shows the number of correct, incorrect, and null responses to the overall amount of positive stimuli of both the learning and test sessions combined. The stimuli are grouped by type of allophonic variant and/or phonetic context as well as by their position in the syllable (i.e. pre-nuclear vs. post-nuclear). Table 26 also shows the percentage of correct responses to each type of allophonic variant, the number of items per type of variant, and the number of responses elicited in each case.

Table 26. Number of Correct/Incorrect/Null Responses & Percent Correct Responses to Positive Stimuli in Exp. 3 Grouped by Position in the Syllable & Type of Allophonic Variant/Phonetic Context (Learning & Test Sessions Combined).							
Position in the Syllable	Allophonic Variant/Phonetic Context	Type of Response			% Correct Responses	Items	Number of Responses Elicited
		C	I	NR			
Pre-nuclear	[p ^h V-]	225	0	15	93.75%	12	240
	[p ^h _ɪ V-]	138	1	1	98.57%	7	140
	[p ^h _ɪ V-]	119	0	1	99.17%	6	120
	[p ^h jV-]	20	0	0	100%	1	20
SUBTOTAL		(502)	(1)	(17)	(96.54%)	(26)	(520)
Post-nuclear	[-Vp]	199	9	12	90.45%	11	220
	[-Vmp]	120	0	0	100%	6	120
	[-Vp ^o t]	58	2	0	96.67%	3	60
	[-Vsp]	40	0	0	100%	2	40
	[-Vpθ]	19	0	1	95%	1	20
	[-Vpst]	19	0	1	95%	1	20
	[-mps]	20	0	0	100%	1	20
SUBTOTAL		(475)	(11)	(14)	(95%)	(25)	(500)
TOTAL		977	12	31	95.78%	51	1020



The results of a *t* test reveal that the differences between the percentages of correct responses to the two subsets of negative stimuli according to the position that their instances occupy in the syllable (i.e. pre-nuclear and post-nuclear), which can also be seen in figure 21, were not statistically significant ($t(9) = 0.252$, P-value: $0.907 > 0.05$). This means that the mean percentages of correct responses to positive stimuli appearing before the nucleus of the syllable and after that nucleus were homogeneous. In other words, the position of the realisation of / p / in the syllable did not have a significant impact on subjects' accuracy in the task. However, the results of an ANOVA carried out on the percentages of correct responses to each type of allophonic variant showed that the differences were significant ($F^*(10, 209) = 3.59$, P-value: $0.0002 < 0.05$). The results of two post-hoc analyses using the Duncan and SNK Multiple range tests showed that all the allophonic variants except for [p^hV-] and [-Vp] represented an homogeneous subset on the one hand and [p^hV-] and [-Vp] another homogeneous subset on the other (by the Duncan and SNK tests). The reason for this difference can be due, in the absence of any other obvious explanation, to the fact that many of the examples of [p^hV-] and [-Vp] appear at the beginning of the task, the part of the test where subjects make more mistakes.

Table 26 also shows that subjects' accuracy in their responses to the control positive stimuli, in bold letters, was very high. A comparison of the percentages of correct responses to each type of control positive allophonic variant (see table 27) shows that there were no

significant differences between the three types (100% versus 95%; $P\text{-value} = 0.000 < 0.05$ by a contrast of proportions).

Table 27. Number of Correct/Incorrect/Null Responses to Positive Control Stimuli in the Test Session in Experiment 3.						
Subtype of Positive Control	Type of Response			% Correct Responses	Items	Number of Responses Elicited
	C	I	NR			
[p ^h jV-]	20	0	0	100%	1	20
[-Vpst]	19	0	1	95%	1	20
[-mps]	20	0	0	100%	1	20
TOTAL	59	0	1	98.33%	3	60

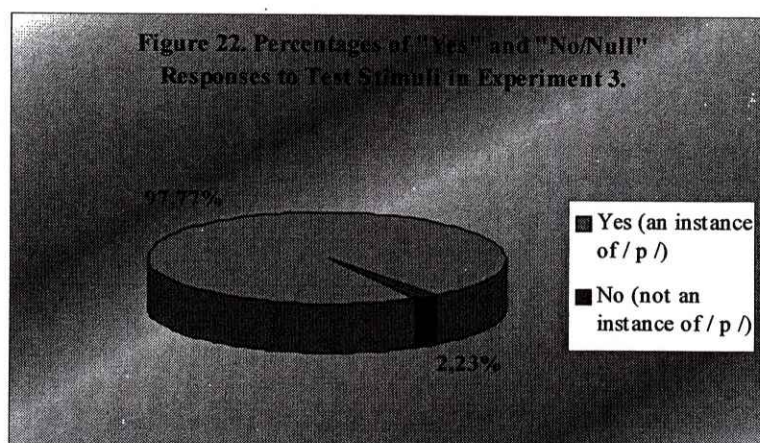
Regarding the negative stimuli containing possible phonetic or orthographic interference, table 28 shows the number of correct, incorrect, and null responses to the negative stimuli to the two types of possibly interfering stimuli, namely, stimuli with phonetic interference (the devoiced realisation of / b /, namely [b̥]), and stimuli with potential orthographic interference (the fricatives / f / and / s / spelled with “ph” and “ps” respectively). The percentages of correct responses to each type of stimuli are also indicated. A comparison of the percentages of correct responses to the potentially phonetically interfering stimuli and those containing potential orthographic interference (“ph”= / f / and “ps”= / s / combined) shows that there was not a statistically significant difference between both groups (100% versus 86%, $P\text{-value} = 0.000 < 0.05$ by a contrast of proportions). In other words, the two groups behaved similarly. In view of the fact that the average percentage of correct responses to such stimuli was 93%, very near the 95% of correct responses to negative stimuli in the learning session and the 97.08% in the test session we can safely argue that the predicted phonetic and orthographic interference did not occur.

Table 28. Number of Correct/Incorrect/Null Responses & Percent Correct Responses to Negative Interfering Items in Experiment 3 (Learning & Test Sessions Combined).							
Type of Potential Interference	Type of Stimuli	Type of Response			% Correct Responses	Items	Number of Responses Elicited
		C	I	NR			
Phonetic	[b̥]	100	0	0	100%	5	100
Orthographic	< ph > (= / f /)	66	9	5	82.5%	4	80
	< ps- > (= / s /)	20	0	0	100%	1	20
	SUBTOTAL	(86)	(9)	(5)	(86%)	(5)	(100)
TOTAL		186	9	5	93%	10	200

Finally, as regards the test stimuli, the number of “yes”, “no”, and null responses to the test words grouped by allophonic variant/phonetic context (i.e. [spV-], [sp_ɪV-], and [sp_ɪV-]) are shown in table 29.

Table 29. Number of Yes/No/No Response and Percent Correct Responses to Different Test Items Grouped by Type of Phonetic Context in Experiment 3.						
Phonetic Context	Type of Response			% Positive Responses	Items	Number of Elicited Responses
	Y	N	NR			
[spV-]	99	1	0	99%	5	100
[sp _ɪ V-]	40	0	0	100%	2	40
[sp _ɪ V-]	37	1	2	92.5%	2	40
TOTAL	176	2	2	97.77%	4	180

Subjects responded affirmatively (i.e. saying “yes” to the question of whether [sp] is an instance of / p /) 99%, 100%, and 92.5% to each of the three subtypes of test stimuli. Moreover, a close comparison of the percentages of “yes” (or positive) responses to the test shows that the percentage differences between the three subtypes of the target sound were not statistically significant ([spV-] 99% versus [sp_ɪV-] 100%, $P\text{-value} = 0.315 > 0.05$; [spV-] 99% versus [sp_ɪV-] 92.5%, $P\text{-value} = 1.29 > 0.05$; [sp_ɪV-] 92.5% versus [sp_ɪV-] 100%, $P\text{-value} = 0.072 > 0.05$, by three different contrasts of proportions). These results clearly indicate that subjects were overwhelmingly including the bilabial stops after tautosyllabic [s] as members of the category that we have called “phoneme / p /” (97.77% of all the responses elicited). This results can also be graphically seen in figure 22. In short, subjects treated oral bilabial stops after tautosyllabic [s] as instances of the category “phoneme / p /” equally often irrespective of the following speech segment.



3.4.3.2. Post-Experimental Interview.

The completion of the test session of the CF task was followed, as was the case in the previous three studies, by a series of individual post-experimental interviews conducted in order to gather useful information about subjects' formation of the category. Table 30 shows a summary of the verbal feedback obtained from the subjects in such post-test interviews. The table reveals that subjects provided very similar descriptions of the target category, calling it either "a *p* sound" (6 subjects), "the sound *p*" (9 subjects), or simply "*p*" (3 subjects). Only two subjects provided different descriptions. Subject 2 (KS) called the category "the strong sound *pe*". This subject seemed to be willing to specify, according to his description, the perceived "strong" nature of the sound and to try to represent the [i:] of the pronunciation of the sound in spelling. The most different description was provided by subject 7 (MD), who called the category "labial, 'lipped' sounds", although he also explicitly referred to the specific sound "*p*'s". Subject 7 (DM) was the only one that used the name of the sound as employed by the experimental subjects (i.e. *p*) in the plural form.⁷² No reference was made to the contrasting category (i.e. words not containing / *p* /), probably because, unlike in experiment one, this category was not a psychologically pre-established category but simply an ad-hoc grouping created for the specific purposes of experiment 3.

Table 30 also shows that subjects provided a considerable amount of information about the articulatory gestures needed to produce the members of the category under investigation. As in the case of the previous category (i.e. oral plosives), subjects also referred to the 4 stages in which the production of the prototypical / *p* / can be divided, that is, closing, hold, release, and post-release. As far as the *closing* stage is concerned, subjects said that the lips, the articulatory organs involved in the production of the realisations of / *p* / in the experimental task, "squeeze a bit" (12, CG), "get together" (12, CG; 16, BW; 18, TO). The subject also said that the speaker has to "purse" the (14, FW) or "close" them (15, RS). Regarding the *hold* stage subjects said that "lips are closed" (2, KS; 4 AN), that "lips are together" (20, LT), or that *p* "stops the air, the sound is blocked by the lips" (8, AM). In relation to the *release* stage, it was claimed that the production of / *p* / required the speaker's lips to "part" (12, CG; 16, BW) as he/she had to "breath out, open lips quickly" (2, KS), to "push lips forward and open them very slightly" (11, DJ). Finally, subjects mentioned the post-release stage in the following ways: "(you) burst it (the sound) out" (1, LS; 6 MS), "blow a little" (14, FW), or "project air forward" (20, LT). Consequently, the sound was characterised by "air puff" (5, LG).

⁷² The numbers given to the subjects in Table 30 do not necessarily indicate the order in which subjects were run in the experiment.



Subjects provided abundant references to the auditory impression caused to them by the instances of the category. The rich descriptions of the target sounds contained many different adjectives referring to the perceived rapid duration of the realisations of / p /. These include the adjectives “short” (4),⁷³ “abrupt” (6), “instant”, “sudden”, “clipped”, and “quick” (1 instance each). Other adjectives pointed to the forceful or explosive character of the release as well as to other perceptual nuances of the post-release stage: “hard” (6), “forceful” (4), “harsh” (3), “popping” (2), “sharp” (2), “bursting”, “exhaling”, “crashing”, “strong”, and “dry” (1 instance each). It is interesting, however, to mention that many subjects (9) spontaneously mentioned differences between various members of the category. More specifically, subject 9 (MD) felt that some instances of the category / p / (e.g. *prayer*, *pour*) were stronger than others (e.g. *depth*); subject 11 (DJ) said that *p* “suffers” in *spray*, unlike in *pray*, and that it is almost silent in *tramp*, *drip*, or *stop*; subject 12 (CG) mentioned that *p* was “kind of hidden” in *clasp* and that in words like *clamp* or *tramp* the [m] somehow “swallows” the end of the word, which “fades away”. On the contrary, subject 12 claimed, / p / was “stronger” and “more obvious” in [p+V] or [p+l] sequences. In addition, subject 14 (FW) said that, depending on the position of the *p* in the word or the following consonant, *p* was “more difficult to hear”, as in *apt*. Subject 15 (RS) claimed that the *p* in *tramp* was “less dominant” than in *pet*, which was “stronger”. Curiously, subject 16 (BW) mentioned almost the same example, saying that *p* in *pet* was stronger than in *stamp*. Also, subject 18 (TO) said that *p* had “more emphasis” in initial position than in final position (e.g. *ship*) and subject 19 (CK) claimed that [p+V] sequences were “harder” than [p+C] sequences, which were “softer” because the consonants “come together” although these were “harder” than the *p* in *tramp*. In general, subject 19 said, some *p*’s were “softer” than others. Finally, subject 20 (LT), said that *p* in *pet* had “more definition” than the *p* in *damp* or *clamp*, in which *p* was “more silent”. In short, all these comments indicate that subjects perceived within-category differences despite the fact that the instructions had not encouraged such awareness of the differences but instead had emphasised cross-category differences between the positive and the negative stimuli.

The examples of the positive category that subjects gave included most of the realisations presented in the learning session, the test session, or both. These were [p^hV-] (9 instances), [p^hɹV-] (5 instances), [p^hɹV-] (4 instances), [-Vp] (6 instances), [-Vmp] (15 items), [-Vp^ot] (2 instances), [-Vsp], [-Vpθ], and [-Vmps] (1 instance each). [sp-] was mentioned as a positive instance twice (“spy”, “spray”). It is also interesting to note that subjects did not mention any example of the negative stimuli except a few that were perceived as being somehow related to the positive category in the sense of causing potential orthographic interference. More specifically, subject 15 (RS) pointed out that “sphere” and “nymph” were not

⁷³ The numbers indicated between parentheses mean the number of times that each adjective was used.

examples of the sound but that they were misleading. Also, subjects 4 (AN) and 5 (LG) mentioned two of the negative stimuli with potential orthographic interference (i.e. “psalm”, “phone”) once each.

Table 30. Name Given to the Category, Articulatory and Auditory Characteristics of the Category Instances, and Examples of the Category Provided in the Post-Experimental Interview.					
	Subj.	Name Given to the Category	Actions to Produce the Sound	Auditory Impression	Examples
1	LS	The sound <i>p</i>	You burst it out	Abrupt	
2	KS	The strong sound “pe”	Lips are closed, breathe out, open lips quickly	Hard	+ < stop >, + < praise >, + < play >
3	KA	The sound <i>p</i>			
4	AN	The sound <i>p</i>	Lips are closed	Bursting	+ < pray >, + < tramp >, - < psalm >, - < phone >
5	LG	A <i>p</i> sound	Air puff	Short	+ < imp >, - < phone >
6	MS	<i>p</i>	Burst it out	Sharp	- < psalm >
7	DM	The sound <i>p</i>		Crashing, hard, forceful	+ < damp >
8	AM	The sound <i>p</i>	Stops the air, the sound is blocked by the lips	Abrupt, hard, forceful, strong	+ < paste >
9	MD	Labial, “lipped” sounds (<i>p</i> ’s)	More delineate than, say, a “b” (finer)	Some <i>p</i> ’s were stronger (e.g. <i>prayer</i> , <i>pour</i>) than others (e.g. <i>depth</i>)	+ < prayer >, + < pour >, + < depth >
10	JS	The sound <i>p</i>			+ < prayer >, + < stamp >
11	DJ	A <i>p</i> sound	Push lips forward and open them very slightly	Clipped, dry, hard sharp, short; <i>p</i> suffers in <i>spray</i> vs. <i>pray</i> ; almost silent in <i>tramp</i> , <i>drip</i> , <i>stop</i>	+ < tramp >, + < drip >, + < stop >, + < pray >, + < spray >
12	CG	A <i>p</i> sound	Lips squeeze a bit, get together then part	Harsh, hard, abrupt. Kind of hidden in <i>clasp</i> , <i>P+V</i> stronger, more obvious. <i>P+V p+l</i> , the easiest, <i>clasp</i> , <i>tramp</i> , <i>m</i> swallows the end of word, which “fades away”.	+ < pet >, + < play >, + < clasp > + < tramp >, + < clasp >
13	CJ	A <i>p</i> sound			+ < lamp >, + < tramp >
14	FW	<i>p</i>	Purse your lips and blow a little	Depending on the position of the <i>p</i> in the word or the following consonant (e.g. <i>apt</i>) is more difficult to hear	+ < apt >, - < nymph > (misleading)
15	RS	A <i>p</i> sound	Close lips	Forceful, popping, quick, short <i>p</i> in <i>tramp</i> less dominant than in <i>pet</i> , which is stronger	+ < drop >, + < camp >, + < pea >, + < tramp >, + < pet >, - < sphere >
16	BW	The sound <i>p</i>	Lips get together, then part	Harsh, abrupt <i>p</i> in <i>pet</i> is stronger than in <i>stamp</i>	+ < glimpse >, + < lip >, + < part >, + < pet >
17	CM	The sound <i>p</i>			
18	TO	The sound <i>p</i>	Lips get together	Harsh, hard, abrupt, short More emphasis in initial position than in <i>ship</i>	+ < plane >, + < plant >, + < ship >
19	CK	<i>p</i>		More sudden, popping, abrupt, instant, <i>p+V</i> harder than <i>p+C</i> (softer because they come together) but <i>p</i> harder than in <i>tramp</i> . Some <i>p</i> ’s softer than others.	+ < apt >, + < pet >, + < tramp >
20	LT	A <i>p</i> sound	Lips are together, project air forward	Forceful, exhaling. <i>p</i> in “pet” has more definition than in <i>damp</i> , <i>clasp</i> , which is more “silent”	+ < stamp >, + < pet >, + < spy >, + < damp >, + < clasp >

3.4.4. Discussion.

The above investigation was conducted in order to obtain experimental evidence on the formation of the category “phoneme / p /” as it is understood by phonetically naive native speakers of English. The study also examined subjects’ classification of oral bilabial stops after tautosyllabic [s].

The results of both the learning and test sessions indicate that all the subjects who took part in the experiment formed the category correctly, reaching the pre-established 37-correct-response criterion in the learning session, with a mean of 56.55 correct responses (range 51-60, s.d.= 2.06) in that session. Subjects’ successful formation of the category is also evident in their significant increase in the mean percentage of correct responses in the test session (i.e. 98.71%) in comparison with the percentage of correct responses in the learning session (i.e. 94.17%). In other words, these percentages show that subjects made few mistakes in the learning session (i.e. 5.83%) but even fewer in the test session (1.69%). In addition, subjects responded correctly to positive control stimuli (i.e. [p^hjV-], [-mps], and [-Vpst]) 98.33% of the time (100%, 100%, and 95% respectively), and to negative controls (i.e. [-Vb]) 100% of the time. These percentages represent substantially high levels of accuracy in their responses to new subtypes of stimuli in the absence of feedback. Finally, subjects’ reports in the post-experimental interview revealed a relatively homogeneous labelling of the category that is also in accordance with the concept intended by the experimenter. Also, the articulatory, acoustic, and auditory characteristics implicit in subjects’ reports correspond to many of the features described in the specialised literature. Therefore, the answer to the first research questions (i.e. whether subjects can categorise different allophones of / p / as members of the same category) is a clear “yes” and the hypothesis entertained (i.e. that subjects would be able to do so) is then confirmed. In other words, the results of the experimental task clearly show that the phoneme / p /, operationalised as a conceptual category, concept, or scheme instantiated by phonetically different sounds that language users classify as instances of the same “type of sound” or category, is a robustly pre-established category in memory. The claim that the fact that “allophones are perceived as instances of phonemes is ... simply an instance of categorisation” (1999: 319) seems then to gain support from the results of experiment 3, which strengthen the view that the phoneme is a psychologically real unit and add new evidence to the data already obtained supporting this view (e.g. Jaeger 1980a, 1980b, 1986; Ohala 1983 1986).

The results of both the learning and test sessions also show that that subjects responded correctly to [b̥], the devoiced realisation of / b / in either word-initial or in word-final position on every occasion. Consequently, no predicted phonetic interference took place in subjects’ formation of the category. In addition, the results show that the predicted orthographic interference had almost no effect at all (91.25% of correct responses to all stimuli with potential

orthographic interference). More specifically, words spelled with word-initial < ps- > caused no interference at all (0% errors) and subjects failed to provide a correct response to words including the digraph “ph” (standing for / f /) only 17.5% of the time. A fact supporting the claim that orthography was little interfering is that some unpredicted phonetic interference may have occurred in subjects’ responses to one of the four words containing “ph”, namely “nymph”. This word, which was responded to correctly on 13 occasions, negatively in 2, and obtained no response on 5 occasions, represented 50% of all incorrect/null responses to the group of negative stimuli with “ph”=/ f /, a fact that is curious in view of the fact that “phone”, the previous word containing “ph”=/ f / had a 100% of correct responses. Regarding “nymphç”, it could be argued that although its nasal consonant is not bilabial but a labio-dental consonant that anticipates the voiceless labio-dental fricative (i.e. [nɪmf]), the labial part of the nasal may have biased subjects towards believing to have heard another labial sound, that is, [p]. In fact, if that were the case, the orthographic interference of “nymph” would be more questionable and would not appear to be sufficient, by itself, to explain why “nymph” had the highest percentage of incorrect responses of all “< ph >=/ f /” stimuli.

Finally, the results of the test session also show that subjects overwhelmingly considered oral bilabial stops in [s]-clusters as members of the category under investigation (97.78%) with no significant differences between the different types of clusters (i.e. [spV-], [spɹV-], and [spɿV-]). Therefore the answer to the second research question (i.e. whether subjects would consider oral stops after tautosyllabic [s] as instances of the category / p /) is again a clear “yes”. Consequently, the hypothesis entertained (i.e. that subjects would consider the target oral stops as instances of / p /) is again confirmed.

Why do subjects categorise oral bilabial stops after tautosyllabic [s] as instances of / p /? As mentioned in 3.4.1., researchers like Fink (1974) and Treiman (1985b) claim that adults spell non-words with the letters “p”, “t”, and “k/c/qu” because they know that stops in [s]-clusters are almost always spelled with those letters, also used to spell voiceless aspirated stops in initial position. A similar spelling-based criterion might explain why subjects assign the oral bilabial stops after tautosyllabic [s] to / p /: people know that those stops are spelled in the same manner as stops in word-initial position that are classified as / p / (e.g. “pit”-“spit”). Related to this spelling-based explanation of the classification of oral stops after tautosyllabic / s / is the fact that literacy is a possible source of psychologically real linguistic knowledge and that orthography may play a very important part in the formation of speakers’ unconscious speech sound categories. In other words, spelling is, with all likelihood, instrumental in originally forming people’s unconscious speech sound categories (Jaeger 1980a: 156-158, 350-352, sections 3.4.4., and 5.2.2. respectively). According to Jaeger:

Certain entities can be psychologically real either because they have been brought to speakers' conscious attention as part of their education or because they have been intuited from the orthographic system of their language.

Jeri J. Jaeger (1984: 34).

To this we should add that such knowledge has also been shown to be closely related to the classification of allophones as members of different phoneme categories (e.g. Jaeger 1980a, 1980b; Mompeán-González, in press; Skousen 1982). It might be argued, then, that the use of real words with spellings familiar to the subjects biased them towards treating the oral bilabial stops in word-initial [s]+oral bilabial stop clusters as instances of / p /. However, as mentioned earlier, even when non-sense syllables are used (e.g. Fink 1974; Treiman 1985b) the same results are obtained. It might then be useful to obtain further evidence on the precise role of orthography from completely illiterate subjects, as some researchers recommend (e.g. Jaeger 1980a; Nearey 1981). Given that beginning spellers occasionally spell oral stops after tautosyllabic [s] with “b”, “d”, “g” as well (Fink 1974; Read 1971, 1975, 1986; Treiman 1985b) and that even literate adults perceive the phonetic similarity between those stops and the word-initial realisations of the series / b d g /, ⁷⁴ it would not be surprising to find that illiterate subjects could focus on such similarities between word-initial / b d g / and oral stops after tautosyllabic [s] and classify the latter as instances of / b d g /.

However, there is evidence that orthography may not be telling the whole story about adult literate subjects' classification of oral stops after tautosyllabic [s]. In fact, it is one thing to claim that subjects' classification reflects their knowledge of spelling conventions; but it is a much stronger claim to argue that spelling rules are the original and exclusive source of their classifying behaviour.

In this respect, it might be interesting to recall the results obtained in Ohala's (1983) CF experiments. Ohala taught one experimental group the category “words containing [k^h]”. In the learning session, this category was exemplified by words like “cat”, “key”, etc. The negative instances for this group were, amongst different items including word-initial / g / (e.g. “get”, “game”, etc.), the words “ghoul”, “gate”, “gold”, and “grape”. These words had been created by splicing the [s] from the beginning of the words “school”, “skate”, “scold”, and “scrape”. These four words appeared intact (i.e. with the [s]) in the test session. The interesting finding was that subjects assigned these words to the target category even though those words -or the crucial part of them- had been presented as non-category items in the training session. Ohala also taught a second group the category “words containing [ǵ] or [g]”, exemplified in the learning session by word-initial / g / (e.g. “glitter”) and intervocalic instances (e.g. “digger”),

together with the words “ghoul”, “gate”, etc., also formed by splicing the [s] from the beginning of “school”, “skate”, etc. The interesting finding in this second CF task was that when the words appeared intact in the test session subjects rejected them from the category even though the [s]-less fragment of them had been given in the training session as positive category items. Ohala claimed that the apparently controversial results obtained could not be explained on purely phonetic groups, as the fact that both stops after initial [s] and word-initial / b /’s created by removing the [s] from the beginning of the word were phonetically identical. Instead, subjects’ criterion of assigning allophones to phonemes seemed to be *distributional*, not purely phonetic or orthographic. The lesson to be learned from Ohala’s (1983) CF studies is that subjects assign velar stops after tautosyllabic [s] to / k / no matter how different those velar stops are from word-initial instances of / k / or how similar they are to word-initial instances of / g /. This hypothesis is further encouraged by Sawusch and Jusczyk’ (1981) study, in which subjects labelled a syllable with an initial 10-msec VOT as / ba / but, when fricative noise corresponding to / s / was added to the stimulus, subjects identified the stop as a / p /. It seems then that any oral stop will be taken as a / p /, / t /, or / k / (depending on its point of articulation) whenever it appears after tautosyllabic / s /. The subjects in experiment 3 may have well followed the above distributional strategy.

There is a further piece of evidence from experimenter 3 that might be taken as supporting the fact that subjects’ criterion for assigning oral bilabial stops after tautosyllabic / s / is not exclusively spelling-based. In experiment 3, there were four words in which the digraph “ph” had the phonological value / f /, namely “sphere”, “graph”, “phone”, and “nymph”. The first two appeared in the learning session and the last in the test session. The positions of these word in the 100-stimulus list were 27, 29, 82, and 94 respectively. The word “sphere” is particularly interesting because the spelling of its initial consonantal cluster (i.e. < sph >) is, except for the final “h”, the same as those of tautosyllabic [s]+bilabial stop clusters. Although it is true that the digraph “ph” does not usually have a bilabial phonological value, except for a few exceptions like “shepherd”, it could be argued that the mere presence of the letter “p” could have made subjects classify “sphere” as an instance of / p /. However, subjects erred on “sphere” 25% of the time, making 15 correct responses and 5 incorrect ones. In other words, 15 out of the 20 criterial subjects considered as early as item 27 in the 100-stimulus list, that “sphere” was not an example of the category under investigation (i.e. / p /). By word 29 (i.e. “graph”), two stimuli later, 18 out of the 20 subjects considered that “graph” did not contain any instantiation of / p / either, and by word 22 in the test session (i.e. “phone”) every subjects

⁷⁴ In Treiman’s (1985b) study, subjects consistently spelled oral stops in [s]-clusters as voiceless but, when asked to give a phonetically plausible alternative spelling, two thirds were able to spontaneously notice the phonetic similarity between stops after / s / and / b d g / in word-initial position or be induced to show such an awareness.



answered correctly. This shows that subjects did not classify stimuli as containing an instance of / p / just because they were spelled with the letter “p”. This means that subjects were increasingly basing their responses, as their answers to the words “graph” or “phone” indicated, on phonetic grounds. If oral stops after / s / are included as instances of / p /, those stops have, *at least*, to be bilabial, an important phonetic criterion. .

The orthographic, distributional, and phonetic criteria discussed are not mutually exclusive and it is likely that all of them play some role in subjects’ classifying behaviour regarding the test stimuli. Be that as it may, the results of the present investigation seem to indicate that the answer to the phonological question of the classification of oral stops after tautosyllabic / p t k / indubitably points to / p t k /. We might then tend to think that this issue is already settled, given the experimental results obtained from different studies on the classification of oral stops after tautosyllabic [s] (e.g. Jaeger 1980a, 1980b; Ohala 1983, 1986; Sawusch & Jusczyk 1981). However, there remain three alternative possibilities that we should discuss. First, subjects could consider oral bilabial stops after tautosyllabic [s] as / b / and it might be argued that the excessive presence of instances of / p / in the experimental task may have somehow biased subjects towards treating [sp] sequences as including an instance of / p /. However, from what we know about Ohala’s (1983, 1986) studies, it is likely that if subjects had to form the category / b / and were made to classify test items of the type used in experiment 3 (i.e. “spy”, “spoon” etc.) they would exclude them from the category / b /. However, although actual experimental evidence is needed to confirm such a hypothesis, it is likely that subjects would reject [s]+oral bilabial stop clusters as members of the category. If that were the case, as it seems most likely (although), the archiphonemic solution would remain as the sole (and second) competing *segmental* alternative.

The present experiment does not provide definite evidence against archiphonemes but it does show that oral stops in [s]-clusters are *classified* as instances of / p / and does not lend support the probable prediction of the archiphonemic solution (in the Praguean sense). If, for instance, oral bilabial stops in [s]-clusters were Praguean archiphonemes, this means that they would represent a third category different from either / p / or / b /. This further implies that subjects should *not* have included the test items as members of the category they had formed in the learning session, that is, / p /.⁷⁵

It might still be argued that oral stops after tautosyllabic / s / are not actually a third category but *both* / p / and / b /. In this respect, it should be mentioned that this is an inaccurate view of the original notion of the archiphoneme that has been strongly criticised precisely for

⁷⁵ Further experiments are needed to clarify whether archiphonemes are psycholinguistically real categories for speakers. A kind of absolute discrimination experiment of the type used in speech perception experiments might, for example, be devised in



falsifying the genuine idea embodied in the theory of the archiphoneme (Akamatsu 1988).⁷⁶ However, even if we neglect such a common misunderstanding and we examine the possibility that oral stops after tautosyllabic / s / may instantiate both / p t k / and / b d g /, we come to the conclusion that, as Donegan and Stampe suggest (1979: 162), “uncertainty” or “variation” seems to be the most powerful argument supporting such a view. However, no variation was obtained in experiment 3: subjects did not classify oral bilabial stops after tautosyllabic [s] as instances of / p / 50% of the time and as non-instances the other 50% of the time. If oral stops were, for language users, both / p / and / b /, more inconsistency in classification would be expected but this is not what our results showed. Instead, subjects overwhelmingly took such oral stops as examples of / p /.

At this point it could be argued, as it has been in the past (e.g. Davidsen-Nielsen 1975; Stemberger 1983), that oral stops after tautosyllabic [s] sometimes emerge as / b /, as when “spell mother” turns into [smel 'bʌðr] and that this is enough to believe that there is something forcing the interpretation that oral bilabial stops after tautosyllabic [s] should be classified as / b /. To explain such behaviour of oral stops in [sp] clusters, we can resort to the interpretation proposed by Davidsen-Nielsen (1975). According to the Danish phonetician, the reason that when the oral plosives after tautosyllabic [s] are moved out of their position emerge as [p^h t^h k^h] when the interfering segment (i.e. the segment which is interchanged with the archisegment or which attracts the [s]) is unvoiced, and they emerge as [b̥ d̥ g̥] when the interfering segment is voiced. However, reasonable as the explanation seems to be, it brings to mind the question of whether speakers’ productions need correspond with the entities that he/she classifies as instances of a given category or with his/her intentions. To give an example from another speech error, if an English speaker tries to say “bit and fat” but ends up saying “pig and vat” (Fromkin, 1973: 17), this does not mean that the subject’s classification of the labio-dental fricative [f] when he/she pronounces “fat” appropriately has to be an instance of / v /, although it may be taken as a / v / in the speech error even by himself. Regarding examples like [smel 'bʌðr] we could also regard them as a case of voicing reversal but with no implications as to the classification of the oral stops after tautosyllabic [s]. The speaker may end up saying [smel 'bʌðr] but he intended to say [spel mʌð] and when asked to classify the stop after [s] in “spell mother”, he/she will identify it as a / p /. If the speech error data (e.g. Davidsen-Nielsen 1975; Stemberger 1983) and spelling data (Treiman 1985b) are interpreted in this way, there is, as Stampe claims (1987: 290, 297) no other empirical evidence in support of archiphonemes.

which phonetically naive or even illiterate speakers would have to respond by pressing either of three buttons labelled “/ p /”, “/ b /” or “/ any other /”.

Finally, we should like to consider the possibility that speakers could classify the oral stop after tautosyllabic [s] as the featural approach of Jakobson (e.g. Jakobson *et al.* 1952: 6-39) or Schane (1968) would predict. According to such an approach, syllable-initial aspirated stops, unaspirated devoiced stops, and unaspirated stops after tautosyllabic [s] are [+consonantal], [-vocalic], [-continuant], [-nasal], [-compact], and [+grave]. However aspirated stops are [+tense] (the feature of tenseness includes aspiration) whereas voiced stops are [-tense] and stops in [s]-clusters are unspecified as far as tenseness is concerned. Finally, the feature [voice] is left unspecified for the three of them.

If it were true that the categories to which stops are assigned were exclusively specified by those distinctive features, and that subjects used those features to classify a given speech segment as a member of a given category, the subjects in experiment 3 should have not classified oral bilabial stops after tautosyllabic [s] together with [p^h] on any occasion because word-initial / p / (as well as many of the other realisations of / p /) are specified as [+tense]. It may still be the case that syllable-initial aspirated stops, unaspirated devoiced stops, and unaspirated stops after tautosyllabic [s] retain different conceptual representations but when it comes to classifying different types of stops they are grouped on the basis of their phonetic similarity or any other reason. However, if the binary feature approach is interpreted in classificatory terms, it is not at all clear why unaspirated stops after tautosyllabic [s] are classified together with syllable-initial aspirated stops unless a more abstract binary distinctive featural representations applies to the two types of stops and such a featural representation is more at the level of the phoneme than the allophone.

To conclude this discussion, we should like to recall Derwing and Wang's claim that "in science, no doors are ever closed forever, as new observation may come to light at any time or new theoretical developments may serve to put old problems in an entirely new light" (1986: 113). However, for the time being, we can safely argue that subjects classify the oral stops after tautosyllabic [s] as instances of the category that they call "the sound *p*". In doing so, we do not deny the fact that subjects may perceive the phonetic similarity between word-initial / b / and oral stops after tautosyllabic [s] and we acknowledge that spelling, distributional, and phonetic criteria may be equally important in the categorising behaviour of adult literate subjects. We do not deny either that the oral stops after tautosyllabic [s] occasionally disambiguate as stops that, for the phonologists, should be classified as / b d g /. However, we can be more or less certain that subjects' intentional classifying behaviour treats the stops in [sp] clusters as instances of / p /.

⁷⁶ This seems to be, by the way, the actual interpretation of the archiphoneme in Davidsen-Nielsen's (1975) and Stemberger's (1983) speech error studies and in Treiman's (1985b) spelling study.



3.5. Experiment 4. The Category “Allophone [p^h]” in English.

3.5.1. Design: Rationale, Purpose of the Experiment, Research Questions, and Hypotheses.

The different realisations of a phoneme are often referred to as *allophones* (e.g. Abercrombie 1967: 86; Bloch & Trager 1942: 40; Gimson 1980: 43, 47, 146-147; Gleason 1955: 263; Hyman 1975: 61-64; Jones 1984: 172; Kreidler 1989: 9, 98; Lass 1984: 18; Roach 1983: 35; Trager 1942: 221; Wells 1982a: 45-46; etc.). The phoneme / p /, for example, is realised in different ways depending on the context: as an aspirated voiceless bilabial plosive (i.e. [p^h]) in stressed pre-nuclear, pre-vocalic positions, as an unaspirated voiceless bilabial plosive [p] after tautosyllabic pre-nuclear [s], as an unreleased voiceless bilabial plosive (i.e. [p^o]) before another plosive, etc. The various allophones of a phoneme are said to be in *complementary distribution*. This means that allophones occur in (and are predictable from) specific and mutually exclusive phonetic contexts, that is, where one allophone occurs, the others cannot, and the reverse. Occasionally, two or more different allophones occur in the same position variably depending on a host of factors like speech rate, articulatory emphasis, etc. These allophones are then said to be in *free variation*. For instance, although / p / is usually realised as a voiceless weakly aspirated plosive in pre-vocalic non-stressed positions (e.g. “copper” ['k^hɒpə], it can also be pronounced with a bilabial fricative sound (i.e. ['k^hɒɸə]) when pronounced in a careless manner. Another case of free variation is provided by post-vocalic, word-final / p /, which is usually weakly aspirated or unaspirated plosive in final position (e.g. [læp]), can also be pronounced with an unreleased variety [p^o], typical of colloquial speech, with a strongly aspirated variety [læp^h], typical of emphatic speech, or with either a glottalised stop [læʔp], or a glottal stop [l^hæʔ], typical of some varieties of English like popular London speech.

In the phonetic and/or phonological literatures, it is usually claimed that, despite articulatory and/or acoustic differences, the allophones of any given phoneme are accepted, considered, felt, heard, etc., as the “same” (e.g. Bloch & Trager 1942: 40; Brown 1977: 13; Gimson 1980: 48; Hockett 1955: 144; Jones 1984: 171; Nathan 1986: 216; O'Connor 1973: 121; Pike 1947: 65; Stampe 1987: 295). These seem to be statements about categorisation in that they imply that somehow different speech sounds are treated or responded to in the same manner. However, although every author claims that so-called contrastive phonological units (e.g. / p /-/ b /, / m /-/ n /, etc.) are *perceived* as different, there appear to be contradictory statements as to the extent to which speakers are aware (or unaware) of allophonic variation. Some researchers claim that awareness of allophonic variation is severely limited and that it can only be reached with a special (and often long) phonetic training; speakers have, according to

this view, serious difficulties in distinguishing between allophonic variants (e.g. Baudouin de Courtenay 1972: 174ff; Donegan & Stampe 1979: 162-164; Nathan 1996: 112, 1999: 312-313; Pike 1943: 115, 1947: 65; Stampe 1987: 293; Swadesh 1935: 118; etc.). On the contrary, others (fewer in number) mention that speakers are able to distinguish such variants even without special training. These researchers further claim that language users are not generally aware of allophonic variation simply because they need not be so in everyday life, not because they cannot perceive or hear such variations. According to this view, speakers simply feel that the differences between allophonic variants do not matter but they can hear them under careful listening circumstances (when people may be induced to focus their attention on such variations) or may even notice them spontaneously (e.g. Abercrombie 1967: 85, 87; O'Connor 1973: 121). Still, other researchers are more cautious and simply maintain a more neutral position: they claim that ordinary native speakers are usually unaware of the allophonic variations (e.g. Gimson 1980: 48; Kreidler 1989: 98; Wells 1982a: 41).

The fact that subjects may be perceptually limited in distinguishing allophonic variants may, in principle, seem surprising given the extensive literature now available on the issue of categorical perception vs. continuous perception. **Categorical perception** refers to a mode of perception in which changes along a stimulus continuum are not perceived continuously, but in a discrete manner. Categorical perception is in direct opposition to **continuous perception**, which refers to a relatively continuous relationship between changes in a stimulus and changes in the perceptual experience of that stimulus. Categorical perception studies (e.g. Liberman *et al.* 1957; Studdert-Kennedy *et al.* 1970; see also Repp 1984 for a review) claim that listeners can discriminate stimuli only to the extent that they can recognise them as members of different categories. Advocates of categorical perception claim that people are better able to distinguish between speech sounds that belong to different categories than between different speech sounds that belong to the same category.

However, several studies have shown that the discrimination of stimuli from a given phonetic category with a relatively high degree of accuracy is not at all limited; under certain specific experimental conditions, listeners can discriminate stimuli within a category remarkably well (e.g. Barclay 1972; Carney & Widin 1976; Carney *et al.* 1977; Massaro & Cohen 1983; Pisoni & Lazarus 1974; Pisoni & Tash 1974; Samuel 1977). It is now generally agreed that categorical perception results do not imply that stimuli are perceived categorically. Instead, it is argued that task specifics make subjects' responses look as though they perceived stimuli categorically when, under more appropriate experimental conditions, they can be shown to have actually perceived them continuously (e.g. Barclay 1972; Healy & Massaro 1982; Massaro & Cohen 1983).

Furthermore, growing evidence suggests that within-category stimuli are not only discriminable from one another but are perceived as varying in typicality, with some members



of a phonetic category perceived as better examples than others. Studies including tasks that assess overt judgements of category goodness have shown that subjects can provide typicality ratings for different within-category speech sounds with statistical reliability (e.g. Davis & Kuhl 1992, 1994; Grieser & Kuhl 1989; Hodgson & Miller 1991, 1992, 1996; Kuhl 1991; Miller & Volaitis 1989; Miller *et al.* 1997; Samuel 1982; Volaitis & Miller 1991, 1992; Wayland *et al.* 1994). In addition, several typicality “effects” have been revealed by tasks that assess the functional or differential effectiveness of members of a category in such phenomena as dichotic competition, selective adaptation, generalisation, and category verification. For example, some stimuli are more effective adaptors than others in selective adaptation experiments (Miller 1977; Miller *et al.* 1983; Samuel 1982) or are more effective competitors in dichotic listening competition experiments (Miller 1977; Repp 1977). Moreover, some stimuli elicit greater generalisation than others in transfer sessions when the former have served as the referent stimuli in training sessions (Grieser & Kuhl 1983, 1989; Kuhl 1991). Finally, it has been shown that typical stimuli take less time to be verified as category members in category verification tasks than less typical members of the category (e.g. Massaro 1987; Miller 1994; Pisoni & Tash 1974).

In short, what these studies show is that if people treat phonetic segments in a categorical way (e.g. a particular sound may be either / p / or / b /), it is because the status that speech sounds have as phonetic segments in natural language may force listeners to respond to them in a *categorical* or *phonetic* rather than a *continuous* or *auditory* mode of perception. However, the fact that phonetic segments are categorical in their linguistic function does not imply that they are also categorical in the way they are perceived. A clear distinction should then be made between perception and the utilisation of perceivable differences (Carney *et al.* 1977; Massaro 1987; Massaro & Cohen 1987).⁷⁷ If this distinction is not made, one essentially asserts that all stimuli that are treated similarly, e.g., stimuli that are identified as the same in some sense, are in fact perceived as being identical stimuli.

Despite the enormous body of literature in the field of speech perception on the issue of categorical vs. continuous perception, we know of no experimental studies in phonology or any related discipline that have explicitly addressed the issue of whether the different allophones of a given phoneme can be actually perceived as different sounds despite “belonging” in the same category. In fact, in the speech perception literature experiments are typically conducted on a fixed phonological context (e.g. a pre-vocalic, word-initial [p^ha-ba] continuum), but contextual allophonic differences are not usually dealt with. There is some evidence that speakers group different allophones into phoneme categories (Jaeger 1980a, 1980b; Ohala 1983,

⁷⁷ Massaro (1987; Massaro & Cohen 1983: 32) distinguishes between two types of processes in phonetic categorisation: sensory and decisional. Massaro claims that all sensory processes are continuous, and categorical-perception boundary effects arise only because of discrete “decision” processes that cause stimuli to be “partitioned” categorically into either member or not member

1986; experiment 3 in this dissertation) but not yet whether speakers can classify instances of a given allophone to the exclusion of other allophonic variants of the same phoneme.

The purpose of the present experiment was then to provide empirical evidence on the issue of whether allophonic variation can be detected and whether there is any psychological validity for the notion behind the concept of allophone. More specifically, the present study tried to find out whether speakers can discriminate realisations of the allophone “aspirated / p /” (i.e. [p^h]) from other realisations of the same phoneme.⁷⁸ Much has been written on aspiration in phonetics and/or phonology (e.g. Abercrombie 1967: 148-149; Bloch & Trager 1942: 32-34; Chomsky & Halle 1968: 326-328; Hoard 1971; Hurch 1988; Kim 1970; Klatt 1975; Ladefoged 1993: 258-259, 268; Lass 1984: 91, 178; Nathan 1996: 114-115; etc.). Aspiration, which should be understood in this study as the voiceless interval consisting of strongly expelled breath between the release of the plosive and the onset of a following vowel, has been claimed to be the distinguishing factor between the pairs / p /- / b /, / t /- / d / and / k /- / g / in word-initial position (e.g. O’Connor 1973: 132; Reeds & Wang 1961; Winitz *et al.* 1975; etc.) or it has been claimed to be a fortition, one of the types of processes explored in natural phonology, which make segments more perceivable by emphasising specific phonetic features, etc. The issue concerning aspiration and aspirated plosives in the present study was whether this specific realisation of a plosive is perceptually salient to speakers. The specific research questions that this study addressed were:

- 1) Can native speakers of English classify instances of a particular allophone of a phoneme category like [p^h] as a member of a category to the exclusion of instances of other allophones of the same phoneme category?.
- 2) Are phonetically naive speakers of English able to perceive such allophonic variation?

The hypotheses entertained in this study were that subjects would be able to classify aspirated realisations of / p / to the exclusion of other realisations of the phoneme (and generalise successfully in a post-training session). This hypothesis was based on the fact that the state of affairs in the categorical vs. continuous debate in speech perception studies seems to support the idea that subjects may actually have no problems to perceive allophonic variation under appropriate experimental conditions (as the ones of the present investigation) although subjects’ overall performance may be less successful than in a “phoneme-based” task like the one in experiment 3 due to their habituation to treat segments phonemically instead of allophonically.

of a phonetic category but they do not imply that these stimulus were perceived categorically. Therefore, Massaro draws the distinction between “categorical partitioning” (or categorical responding) and “categorical perception”.

3.5.2. Method.

3.5.2.1. Subjects.

Twenty subjects between the ages of 18 and 45 (mean age 24) and selected randomly from a group of 80 recruited for the CF experiments reported in this work took part in this experiment. There were 9 men and 11 women. Subjects were paid for their participation and reported no hearing problems. The subjects had the characteristics mentioned in 3.1.1.2.

3.5.2.2. Stimuli.

The stimuli for the present experiment consisted, like in the previous experiments, of 100 monosyllabic English words. These exemplified 12 different canonical syllable structures including open syllables (e.g. CV, CCV, and CCCV), checked syllables with no consonant before the nucleus (VC and VCC), and syllable structures whose nuclei were flanked at both sides by different single consonantal phonemes or permitted consonantal clusters (e.g. CVC, CCVC, CCVCC, etc.). The stimuli had a mean duration of 632 msc.

In the learning session, positive instances (32 in total) included different syllable structures, namely CVCC (9 items), CVC (9 items), CV (5 items), CCV (4 items), CCVCC (3 items), and CCVC (2 items). All examples of [p^h] appeared in the post-nuclear position of words. The phonetic environments in pre-nuclear positions were [p^hV-] (23 items), [p^hɹV-] (5 items) and [p^hɻV-] (4 items). Thus [p^h] was exemplified by strongly aspirated realisations followed by vowels or by either devoiced / r / or devoiced / l / (i.e. [ɹ̥] and [l̥]) with simultaneous [ɹ] and [l] articulations.⁷⁹ Negative items (28 items in total) included different syllable structures, namely CVCC (9 items), CCVC (5 items), CVC, CCV, and CCVCC (3 items each), CCCV (2 items), and VCC, VC, and CCVC (1 item each). All sounds were orthographically non-interfering. The phonetic environments in pre-nuclear positions were [spV-] (6 items), [spɹV-] (2 items), and [spɻV-] (2 items). In post-nuclear positions the contexts were [-Vp] (6 items), [-Vp^ot] and [-Vmp] (3 items each), [-Vps] and [-Vɪp] (2 items each), and [-Vpθ] and [-Vsp] (1 item each).

⁷⁸ As it closely resembles a / h / sound, an aspirated stop is marked with a superscript, usually a small “h” (e.g. [p^h]), although it can also be shown with the sign ['] (e.g. [p']).

⁷⁹ As mentioned in section 3.4.2.2., aspiration (in relation to / p /) is found most notably before a stressed vowel in the same word, that is, in initial stressed positions followed by vowel (e.g. “pet”) or by / l r j / (in these latter cases aspiration is manifested in the devoicing of / l r j /). Medially also, the plosive is aspirated if the following vowel has the louder stress, as in “appal” (e.g. Bloch & Trager 1942: 42-43; Gimson 1980: 151; Jones 1984: 68-69, 1989: 138-139; Kleidler 1989: 116; O'Connor 1973: 132; Roach 1983: 31, 34-35), but this realisation was not included as only monosyllabic words were used. In unaccented syllable (e.g. / p / in

In the test session, there were 22 positive instances and 18 negative ones. There were neither test nor control items in this experiment. Positive ones included the following syllable structures CV (5 items), CVC (9 items), and CVCCC (8 items). In pre-nuclear position there were 21 instances of [p^hV-] and 1 instance of [p^hjV-] –a positive control- (no instances of [p^hɹV-] or [p^hɻV-] were included this time). Negative instances included different syllable structures, namely CCVC and CCVCC (4 items each), CVCC (3 items), CVC (2 items), CCV, CVCC, VCC, CCVCCC, and VC (1 item each). There were no interfering items like in the learning session. In pre-nuclear contexts there were 4 examples of [spV-]. The phonetic contexts in post-nuclear positions were [-Vp] (5 items), [-Vmp] (3 items), [-Vp^ot] and [-Vsp] (2 items each), [-Vpst] and [-Vm^ops] (1 item each). The last item (i.e. [Vm^ops]) represented a negative control.

Table 31 offers a summary of the types of stimuli used and Table 32 presents the actual list of words used and their category status.

Table 31. Types of Stimuli and Number of Items per Subtype of Stimuli Used in Experiment 4.				
Type of Stimulus	LEARNING SESSION		TEST SESSION	
	Subtype of Stimulus	Number of Items	Subtype of Stimulus	Number of Items
POSITIVE ITEMS	<i>Pre-nuclear</i>		<i>Pre-nuclear</i>	
	[p ^h V-]	23	[p ^h V-]	21
	[p ^h ɹV-]	5	[<u>p^hjV-</u>] ⁸⁰	1
	[p ^h ɻV-]	4		
	TOTAL	(32)	TOTAL	(22)
NEGATIVE ITEMS	<i>Non-interfering:</i>		<i>Non-interfering:</i>	
	<i>Pre-nuclear</i>		<i>Pre-nuclear</i>	
	[spV-]	6	[spV-]	4
	[spɹV-]	2		
	[spɻV-]	2		
	<i>Post-nuclear</i>		<i>Post-nuclear</i>	
	[-Vp]	6	[-Vp]	5
	[-Vmp]	3	[-Vmp]	3
	[-Vp ^o t]	3	[-Vp ^o t]	2
	[-Vps]	2	[-Vps]	1
	[-Vɹp]	2	[-Vsp]	2
	[-Vpθ]	1	[-Vm ^o ps]	1
	[-Vsp]	1		
	TOTAL	(28)	TOTAL	(18)

“polite”, “upper”) and finally (e.g. / p / in “lip”) such aspiration as may occur is relatively weak, with some speakers almost imperceptible. In general, we can say that the plosives are not aspirated in word-final position.

⁸⁰ Underlined phonetic sequences in bold letters represent control stimuli.

Table 32. Stimulus List for Experiments 4:
Category [p^b].

Order	Stimulus	Positive (+) Negative (-)	Order	Stimulus	Positive (+) Negative (-)
LEARNING SESSION					
1.	paw	+	53.	poise	+
2.	spend	-	54.	drop	-
3.	push	+	55.	pan	+
4.	rapt	-	56.	spy	-
5.	post	+	57.	clasp	-
6.	pray	+	58.	pots	+
7.	power	+	59.	ship	-
8.	stamp	-	60.	paled	+
9.	pass	+	TEST SESSION		
10.	up	-	1.	payer	+
11.	pulse	+	2.	pass	+
12.	plot	+	3.	pegs	+
13.	spray	-	4.	span	-
14.	purr	+	5.	pen	+
15.	pin	+	6.	clamp	-
16.	spoon	-	7.	wept	-
17.	spear	-	8.	top	-
18.	damp	-	9.	pin	+
19.	pond	+	10.	pea	+
20.	depth	-	11.	spa	-
21.	proud	+	12.	pool	+
22.	help	-	13.	palms	+
23.	shop	-	14.	camp	-
24.	pay	+	15.	pie	+
25.	pill	+	16.	glimpse	- control
26.	pence	+	17.	peace	+
27.	cap	-	18.	keep	-
28.	play	+	19.	poles	+
29.	split	-	20.	spill	-
30.	apt	-	21.	pear	+
31.	par	+	22.	tramp	-
32.	spring	-	23.	pun	+
33.	pain	+	24.	pubs	+
34.	piles	+	25.	lapsed	-
35.	spice	-	26.	pure	+ control
36.	caps	-	27.	group	-
37.	prince	+	28.	pause	+
38.	gulp	-	29.	paved	+
39.	pace	+	30.	sponge	-
40.	pelt	+	31.	gasp	-
41.	trap	-	32.	peel	+
42.	plea	+	33.	opt	-
43.	Splay	-	34.	ape	-
44.	paste	+	35.	punch	+
45.	print	+	36.	pull	+
46.	spare	-	37.	pave	+
47.	kept	-	38.	pant	+
48.	lamp	-	39.	drip	-
49.	plough	+	40.	grasp	-
50.	priest	+			
51.	ropes	-			
52.	puff	+			

3.5.2.3. Procedure.

3.5.2.3.1. Instructions.

The instructions given to the experimental subjects in experiment 4 were more similar to the ones used in experiments 1 and 2 than to those in experiment 3 but not the same as any of them. The actual paragraphs (or the relevant parts of them) where the differences were included are reproduced below for convenience.

.... You are going to be listening to a list of words over the headphones. All of these words contain a certain type of consonantal sound either at the beginning, somewhere in the middle of the word, or at the end. However, although all words contain basically the same consonantal sound, there are some differences between different examples of the consonantal sound just as there are differences between different types of chairs (although all items classified as chairs are chairs). To put it differently, some of the examples of the consonantal sound in the words you'll hear have a certain characteristic (or characteristics) while other examples of the consonantal sound don't. Your task is to pay attention to the words and try to figure out, once you know what type of consonantal sound all words contain, what characteristic(s) some of the examples of the consonantal sound share while other examples of the consonantal sound lack.

At first, you will not know which examples of the consonantal sound contained in every word are characterised by a certain feature (just as at very beginning of the task you will not know what consonant all words include, although after a few words you'll find out).....

Whenever you hear a word, one of the two rectangles on the screen will remain on display while the other will disappear after a few seconds. When the word contains the consonantal sound with a specific characteristic, the green rectangle will stay but the red one will disappear (go off). When the word contains the consonantal sound without that characteristic, the red rectangle will stay but the green one will disappear...

However, we are not only interested in showing you which words contain the consonantal sound with or without a certain characteristic by the disappearance of one of the two rectangles. We are also interested in your guesses and how you show it. So, when you have some idea of what the characteristic certain examples of the consonantal sound have but others lack might be, you should start pressing the keys on the keyboard according to what you think. You'll find out whether you were right by looking at the behaviour of the rectangles.

All subjects reported having understood the instructions fairly well. As in the previous experiments, subjects usually paraphrased the contents of the instructions spontaneously when the experimenter asked them whether they had understood them or not. An accurate explanation of the experimental events indicated that they were ready to begin the task.

3.5.2.3.2. Training Session, Test Session, and Post-Experimental Interview.

The training and test sessions proceeded as described in 3.1.1.5. All subjects were also interviewed in a post-test session. It should be emphasised that in experiment 4 there was a test session but its purpose was not exactly the same as it had been in experiments 1, 2, and 3. The test session in experiment 4 pursued to find out whether subjects generalised their responses without the provision of feedback. However, its goal was not to obtain information about how speakers classified certain instances deemed ambiguous as to their category status. Therefore the test session, which is actually an optional part of a CF investigation (see 2.3.3.), was not absolutely necessary in this investigation but considered, at the same time, as providing

substantive evidence in order to answer the research questions posed and confirm or not the hypotheses entertained.

3.5.3. Results.

3.5.3.1. Training and Test Sessions.

The 20 subjects who participated in this experiment were considered to have formed the category correctly because they reached the pre-established 37-correct response criterion that guaranteed that they had not responded randomly in the learning session. The criterial subjects had a mean of 51.40 correct responses (range 37-59, s.d. 5.15) in the learning session.

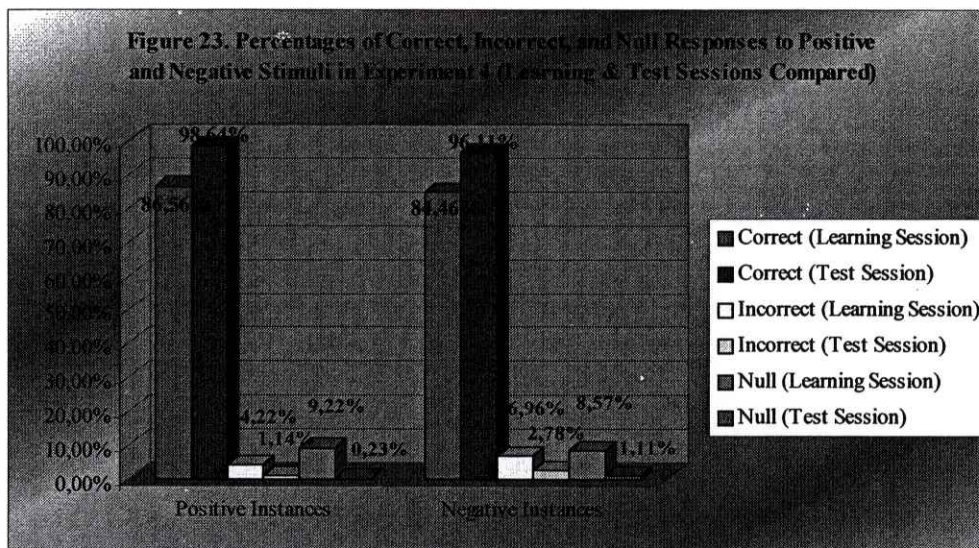
The numbers of correct (C), incorrect (I), and null responses (NR) to both the positive and negative stimuli and the percentages of correct responses to each stimulus type in the learning session are shown Table 33. The table also shows the number of items per type of stimulus and the number of responses elicited in each case. The same type of information obtained from subjects' performance in the test session is shown in Table 34.

Table 33. Number of Correct/Incorrect/Null Response and Percent Correct Responses to Positive and Negative Stimuli in Experiment 4 (Learning Session).						
Stimulus Type	Type of Response			% Correct Responses	Items	Number of Responses Elicited
	C	I	NR			
Positive Instances	554	27	59	86.56%	32	640
Negative Instances	473	39	48	84.46%	28	560
TOTAL	1027	66	107	85.58%	60	1200

Table 34. Number of Correct/Incorrect/Null Responses and Percent Correct Responses to Positive and Negative Stimuli in Experiment 4 (Test Session).						
Stimulus Type	Type of Response			% Correct Responses	Items	Number of Responses Elicited
	C	I	NR			
Positive Instances	434	5	1	98.64%	22	440
Negative Instances	346	10	4	96.11%	18	360
TOTAL	780	15	5	97.5%	40	800

A close comparison of the above tables shows that the percentages of correct responses to both positive and negative stimuli increased from the learning session to the test session. In fact, correct responses were significantly more frequent in the test session than in the learning session

(85.58% versus 97.5% respectively; $P\text{-value} = 0.000 < 0.05$ by a contrast of proportions). The significant increase in subjects' correct responses reveals that they performed much better in the test session than in the learning session. Subject's better overall performance is also visually represented in figure 23, which shows a comparison of the percentages of correct, incorrect, and null responses to positive and negative instances between the learning and the test session. Figure 23 shows an increase of correct responses as well as a decrease of incorrect and null responses to both positive and negative stimuli from the learning session to the test session.

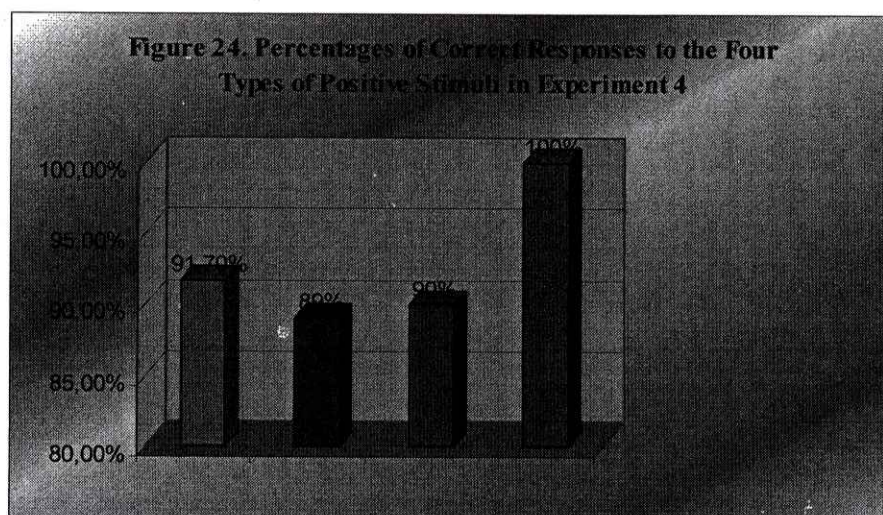


An analysis of the data included in tables 33 and 36 further revealed that there was not a not a significant relationship between type of stimulus (i.e. negative, positive) and type of response (correct, incorrect, null) in the learning session ($P\text{-value} = 0.11 > 0.05$ by a test of independence or chi-square test). Collapsing the values of the variable "type of response" (i.e. correct, incorrect, and null) into two categories (i.e. correct and incorrect/null), no significant relationship between type of stimulus and type of response was found either ($P\text{-value}: 0.34 > 0.05$ by a test of independence using the statistic Yate's corrected chi-square test). In the test session, the expected minimum frequency was below 5 and so the test of independence was not reliable when the values of the variable "type of response" (i.e. correct, incorrect, and null) were not collapsed. However, when such values were collapsed, a significant relationship was observed ($P\text{-value}: 0.04 < 0.05$ by a test of independence using the statistic Yate's corrected chi-square test). More specifically, focusing on the corrected residuals of the test, it was found that the number of positive stimuli were in inverse proportion to the number of incorrect/null responses and in direct proportion to the number of correct responses while negative stimuli were in inverse proportion to the number of correct responses and in direct proportion to the number of incorrect/null responses. The explanation suggested for similar patterns observed in

experiments 1, 2, & 3 (see sections 3.2.3.1., 3.3.3.1, & 3.4.3.1. respectively) may also be applied in this case: a poorer performance in concept formation for negative stimuli than for positive ones as it seems easier, for people, to respond to positive stimuli than to negative stimuli.

Focusing next on subjects' answers to the overall number of positive stimuli in the experimental task, it seems interesting to remark that the positive stimuli were divided into four different groups according to the phonetic context in which [p^hV-] occurs, namely [p^hV-], [p^h_ɹV-], [p^h_{ɹ̥}V-], and [p^hjV-]. The numbers of correct, incorrect, and null responses to each type of positive stimulus as well as the percentages of correct responses, number of items, and number of responses elicited in each case are shown in table 35. Figure 24 also shows the percentages of correct responses to each of the four types of positive stimuli.

Table 35. Number of Correct/Incorrect/Null Responses & Percentage of Correct Responses to Positive Stimuli in Experiment 3 (Learning & Test Sessions Combined).						
Stimulus Type	Type of Response			% Correct Responses	Items	Number of Responses Elicited
	C	I	NR			
[p ^h V-]	807	25	48	91.70%	44	880
[p ^h _ɹ V-]	89	4	7	89%	5	100
[p ^h _{ɹ̥} V-]	72	3	5	90%	4	80
[p ^h jV-]	20	0	0	100%	1	20
TOTAL	988	32	60	91.48%	54	1080



[p ^h V-]	■
[p ^h _ɹ V-]	■
[p ^h _{ɹ̥} V-]	■
[p ^h jV-]	■

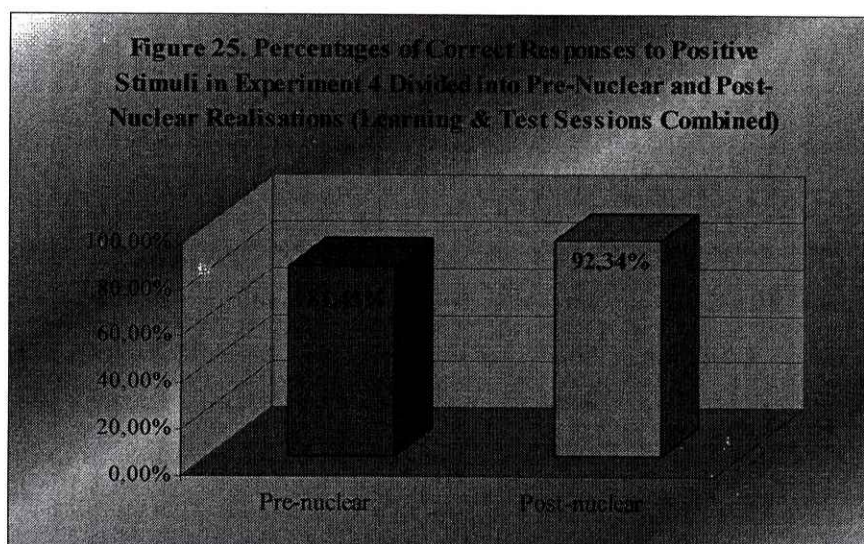
The results of different contrasts of proportions reveal that the differences between the percentages of correct responses to each pair of subtypes of positive stimuli except for [p^hjV-] were not statistically significant ([p^hV-] 91.70% versus [p^hɹV-] 89%, P-value: 0.407 > 0.05; [p^hV-] 91.70% versus [p^hɹV-] 90%, P-value: 0.624 > 0.05; [p^hɹV-] 89% versus [p^hɹV-] 90%, P-value: 0.827 > 0.05). However, there were statistically significant differences between the percentages of correct responses to [p^hjV-], a new uncontroversial instantiation of the category under investigation that served as a positive control in the test session, and other two subtypes of positive stimuli, namely [p^hV-] and [p^hɹV-] ([p^hjV-] 100% versus [p^hV-] 91.70%, P-value: 0.000 < 0.05; [p^hjV-] 100% versus [p^hɹV-] 89%, P-value: 0.000 < 0.05). The only exception was the comparison between [p^hjV-] (100% of correct responses) and [p^hɹV-] (90%), in which the difference turned out to be significant (P-value: 0.003 < 0.05). Carrying out next an ANOVA on the percentages of correct responses to the four types of positive stimuli it was also found out that the differences were significant ($F^*(3, 39) = 8.84$, P-value: 0.0001 < 0.05). Two post-hoc analyses carried out with the Duncan and SNK multiple range tests showed that [p^hV-], [p^hɹV-], and [p^hɹV-] represented an homogeneous subset but [p^hjV-] did not belong to that group. The findings from the contrasts of proportions and the analysis of variance seem to indicate subjects' increasingly better performance in the task (and actual formation of the category) as the higher percentage of correct responses to [p^hjV-] may be due to its positive control status in the test session, where subjects made fewer incorrect responses and where their classifying criteria were already well-established.

Finally, an analysis of the negative stimuli (i.e. the words that did not contain an aspirated realisation of / p /) was also made. Subjects' responses to each type of negative stimulus grouped by allophonic variant and by the position of the variants in the syllable (i.e. pre-nuclear or post-nuclear) are shown in table 36.



Table 36. Number of Correct/Incorrect/Null Responses and Percent Correct Responses to Negative Stimuli Grouped by Position in the Syllable and Allophonic Variant in Experiment 4 (Learning and Test Sessions Combined).							
Position in the Syllable	Allophonic Variant	Type of Stimulus			% Correct Responses	Items	Number of Responses Elicited
		C	I	NR			
Pre-nuclear	[spV-]	158	19	23	79%	10	200
	[sp _l V-]	36	0	4	90%	2	40
	[sp _l V-]	34	3	3	85%	2	40
SUBTOTAL		(228)	(22)	(30)	(81.43%)	(14)	(280)
Post-nuclear	[-Vp]	206	9	5	93.64%	11	220
	[-Vmp]	115	3	2	95.83%	6	120
	[-Vp ^o t]	79	7	14	79%	5	100
	[-Vps]	57	3	0	95%	3	60
	[-Vsp]	57	3	0	95%	3	60
	[-Vtp]	39	0	1	97.5%	2	40
	[-Vpθ]	18	2	0	90%	1	20
	[-mps]	20	0	0	100%	1	20
SUBTOTAL		(591)	(27)	(22)	(92.34%)	(32)	(640)
TOTAL							

The results of a *t* test reveal that the differences between the percentages of correct responses to the negative stimuli divided into a pre-nuclear and a post-nuclear group, which can also be seen in figure 25, were not statistically significant ($t(9) = -1.954$, P-value: $0.082 > 0.05$). This finding indicates that the mean percentages of responses to positive stimuli appearing before the nucleus of the syllable and after that nucleus were homogeneous. In other words, the position of the realisation of / p / in the syllable did not have a significant impact on subjects' accuracy in the task.



However, when an ANOVA was performed on the percentages of correct, incorrect, and null response to each type of positive stimuli it was found out that the differences between the allophonic variants were significant ($F^*(10, 111) = 5.01$, P-value: $0.0000 < 0.05$). The Duncan post-hoc analysis showed that there were three homogeneous subsets. 1) [$sp_{\text{ɪ}}V-$], [$sp_{\text{ɪ}}V-$], [$-Vp$], [$-V_{\text{ɪ}}p$], [$-Vps$], [$-Vmp$], [$-Vsp$], and [$-Vm_{\text{ɪ}}ps$]; 2) [$-Vp$], [$-V_{\text{ɪ}}p$], [$-Vps$], [$-Vmp$], [$-Vsp$], [$-Vm_{\text{ɪ}}ps$], and [$-Vp^{\text{ɔ}}t$]; and 3) [$spV-$] and [$-Vp^{\text{ɔ}}t$]. The SNK test also showed three homogeneous subgroups: 1) [$sp_{\text{ɪ}}V-$], [$sp_{\text{ɪ}}V-$], [$-Vp$], [$-V_{\text{ɪ}}p$], [$-Vps$], [$-Vmp$], [$-Vsp$], [$-Vm_{\text{ɪ}}ps$], and [$-Vp^{\text{ɔ}}t$]; 2) [$sp_{\text{ɪ}}V-$], [$sp_{\text{ɪ}}V-$], [$-Vp$], and [$-V_{\text{ɪ}}p$]; 3) [$spV-$] and [$-Vp^{\text{ɔ}}t$]. These findings suggest that [$spV-$] and [$-Vp^{\text{ɔ}}t$] were the most difficult allophones and, most importantly, that [$-Vm_{\text{ɪ}}ps$] was as easy to respond correctly as any other although it had the highest percentage of correct responses of all types of allophones of / p /.

3.5.3.2. Post-Experimental Interview.

A considerable amount of information was obtained from subjects' reports in the individual post-experimental interviews conducted after the completion of the test session. In order to better discuss this information, table X contains, like similar ones in the previous experiments, the information provided by the subjects divided into four different sections: name or description of the category (and occasionally the contrasting category), articulatory information, auditory impressions, and examples of both negative and positive stimuli. The table shows that, although subjects described the target category in many different ways, the expressions used were more or less synonymous. The aspirated realisations of / p / were referred to as "strong p " (8, HS; 19, DJ), "stronger p " (1, LA; 7, AG), "stronger p 's" (18, RO), "explosive p " (13, JC) "more explosive p " (2, AR), " p ... more of an explosive sound" (4, DW), " p requiring more explosion of air" (5, BM), "more pronounced p " (1, 20), "hard p " (3, HT; 6, OK; 17, KD), "harder p " (12, AS), "stressed p " (9, NW), and "active p " (6, OK).⁸¹ On the contrary, the negative stimuli were referred to as "soft p " (6, OK), "softer p " (1, LA; 3, HT; 7, AG; 8, HS; 12, AS; 13, JC; 17, KD; 19, DJ), "softened p " (9, NW), "quieter p " (1, LA), "calmer p " (2, AR), or "passive p " (6, OK). Subjects also referred to the negative stimuli qualifying the descriptions already used for the positive stimuli. According to this strategy, the negative stimuli were claimed to be "less explosive p " (2, AR), "not as powerful p " (4, DW), " p ... not as strong" (4, DW), " p not so strong" (18, RO), "not as emphatic p " (4, DW), "... requiring not an active expression of the air" (5, BM), "less strong p " (7), "less pronounced p " (20), or "less stressed p " (9).

⁸¹ The numbers given to the subjects in Table 37 do not necessarily indicate the order in which subjects were run in the experiment.

It is also important to emphasise that subjects did not exclusively refer to the auditory phonetic characteristics of the stimuli in the post-experimental interview, but also to the distributional characteristics of negative and positive items. In this respect, 14 out of the 20 subjects mentioned that the positive items were at the beginning of the stimulus word in contrast with the negative ones, which appeared, according to the subjects, elsewhere, “coming later” (in the middle of the word or at the end). Out of these 14 subjects, only 5 (10, RV; 11, TS; 14, WK; 15, PH; & 16, DB) exclusively referred to the positional features of the / p / sounds because they did not provide any other phonetic characteristic of the sound. The rest of those 14 subjects (i.e. 2 to 5, 6 to 9, 13, & 17) described both the positive and negative items using some of the adjectives mentioned in the preceding paragraph as well as the information referring to the position of the realisations of / p / in words.

It could be argued that the five subjects who described the positive stimuli without mentioning any specific phonetic aspect simply formed a distributional characteristic in the sense of associating the positive stimuli with word-initial position and negative stimuli with any other position in the word. It is, however, important to emphasise that those who only referred to the position of the sounds in the stimuli mentioned later some of the characteristics of those sounds similar to the descriptions provided by all other subjects. The references to the articulatory characteristics of the sounds and, more importantly, the auditory impressions that subjects reported having perceived (see the paragraphs below) clearly indicate that no subject simply associated the positive stimuli with the initial position of words nor did he/she fail to notice auditory differences between the positive and negative stimuli. Consequently, the hypothesis that a merely clear-cut category distinction based on the position of the realisations of / p / in the positive stimuli in contrast with the positions in the negative stimuli receives little support.

Looking at the information regarding the articulation of [p^h] in the interviews, it is clear that, although some subjects provided information about the articulation of the positive stimuli that also applied to the negative as well as to other phonemes like / b / or / m / but obviously important for the positive stimuli like the fact that it has a bilabial point of articulation (“close lips” -3, HT-; “close mouth”, -13, JC-), three main aspects or features perceived as characteristic of or peculiar to the realisations of [p^h] were observed. In the first place, the stronger puff of air involved in the production of aspirated plosives seems particularly salient for subjects. When asked to describe what they have to do to produce the sounds belonging to the category under investigation, subjects said that these sounds “come up with more breath” (1, LA), that in these sounds there is “more air than in any other p” (2, AR), that “more air passes through the lips” (6, OK; 12, AS), that “the amount of air or puff of air is bigger” (7, AL), that “more air is expelled” (16, DB), that the speaker should “force a bit of air out” more in the positive stimuli than in the

negative (13, JC), or that the way the lips are positioned provides an opportunity to “aspire the sound” (4, DW). In short, subjects seemed to rely heavily on the fact that when an aspirated / p / is uttered, a fairly strong puff of air is released, but a rather weaker puff of air when unaspirated realisations of / p / occur. In the second place, subjects seemed to perceive the need for a considerable amount of energy to produce [p^h]. Subject 11 (TS), for instance, said that in order to produce the sound one acted as if “spitting out”, subject 8 (HS) claimed that “more effort is made to expel the air”, subject 20 (MM) said that the speaker makes “more emphasis” to produce [p^h], and subject 6 (OK) that he/she has to “work it out”. In the third place, the lip posture perceived and emphasised as an important characteristic of aspirated / p / was one of relatively strong lip rounding. However, they also perceived a special kind of lip posture for the positive items, mentioning with synonymous expressions, that the lips should be more strongly pursed (4, DW; 15, PH; 19, DJ), curled (5, BM), or pushed forward (6, OK). Still, others mentioned that the action of the lips was important but they were not so specific and said that aspirated / p /'s were due to “the way you press the lips” (9, NW), or “the way the mouth is shaped” (10, RV).

Subjects’ abundant references to the auditory impression caused by the positive stimuli in the experiment referred, like in experiment 1, to the perceived rapid duration of the aspirated realisations of / p /. In this respect, subjects said that the positive items were “short” (9, NW; 15, PH; 17, KD) or “very short” (18, RO), “quick” (18, RO), “abrupt” (7, AG; 9, NW), “instant” (10, RV), “sudden” (7, AG), “precipitated” (6, OK), “staccato” (4, DW; 9, NW). In addition, the comments also referred to the perceived forceful or explosive character of the aspirated instances: “explosive” (4, DW), “popping” (2, AR), “hard” (3, HT; 15, PH), “not flowing” (7, AG; 10, RV), “aggressive” (9, NW; 13, JC), “strong” (10, RJ), “sharp” (3, HT; 9, NW), “emphatic” (13, JC), “blown” (15, PH), “punchy” (17, KD), “forthright” (13, JC), “dominant” (6, OK) “crisp” (3, HT), “muscular” (6, OK). Using comparisons, the explosive nature of [p^h] was described as sounding “like a punch” (17, KD), “like the popping of a bubble” (5, BM), “as if you were pushing something” (8, HS), “sort of spitting” (3, HT).

Occasionally, subjects provided a very rich variety of terms to describe the auditory impression caused by the positive stimuli in contrast with the negative ones. Using comparatives, subjects perceived the positive stimuli as “louder” (3, HT), “sharper” (3, HT), “more firm” (14, WK), “stronger” (16, DB; 19, DJ; 20, MM), “more stressed” (16, DB), “harder” (20, MM), “more punching” (16, DB), “more forceful” (16, DB), or “more in four face” (14, WK) than the negative stimuli. These were considered or perceived to be weaker or less forceful: “unreleased as in *apt*” (5, BM), “not emphasised in *spa* or *glimpse*” (15, PH), “(less stressed and) weaker” in other positions (16, DB; 19, DJ). At times, subjects explained why they perceived this to be so. According to subject 9 (NW), “in *sp*- the *s* takes over and is

more dominant". The "s" was also perceived as more "dominant" in word-final position, as in *grasp* (8, HS), and in words like *lapse* or *caps* "the *s* shadows the *p*" (20, MM). Negative stimuli were also perceived as less hearable: "almost unheard in some cases like *grasp* or *depth*" (6, OK), "unheard but natural in *camp* and *laps*" (18, DB), and "barely heard in *camp*" (19, DJ), "blunter, duller, and more mute after word-initial *s*" (5, BM), or "longer, more subtle, and duller" (14, WK), "more hidden *p*'s" (11, TS).

It is interesting to mention that 3 subjects (4, DW; 5, BM; 8, HS) explicitly reported having perceived a scale of strength between different realisation of the phoneme / p /. DW said that there was a continuum of explosiveness from [p^hV-] through [sp] and [Vp] to [mp]. Also, subject 5 (BM) said that the continuum ranged from "stronger" [p^hV-] through [sp] to [Vp], and subject 8 (HS) said that the continuum went from stronger [p^hV-] through [sp] and [-sp] to [-Vtp]. In addition, there were frequent comparisons of two given allophones. Subject 2 (AR) said that the positive instances were explosive sounds but that the sound was "less strong in *pray*" than when followed by a vowel. In a similar vein, subject 11 (TS) said that [p^hV] sequences were stronger than [p^hC] ones"; 15 (PH) claimed that "[p^hV] sequences were harder than [p^hV-]" as "*p* slides into *l*" in the latter; finally 17 (KD) said "[p^hV] is stronger than [p^hC] sequences". In addition, two subjects compared the negative items with the sound *b* saying that they were "more like a *b*" (3, HT; 6, OK). That [sp] was considered as nearer in the "strength" scale to [p^h] was further supported by subject's 1 (LA) comment that at the beginning of the task he considered that [sp] sequences "looked like positive" but then he ruled them out when informed that they were actually negative instances of the target category.

Finally, it should be mentioned that out of the 27 positive instances mentioned, 23 were aspirated realisations followed by a vowel, 2 by / l / and 2 by / r /; Examples were also provided of all the allophonic variants of / p / that were negative stimuli in both pre-nuclear position ([spV-], 9 instances; [sp_lV-], 2 instances; [sp_rV-], 2 instances) and post-nuclear position ([-Vp], 4 instances; [-Vmp], 5 instances; [-Vp^ot], 3 instances; [-Vps], 7 instances; [-Vtp], 2 instances; [-Vsp], 4 instances; [-Vpθ], 5 instances; and [-Vmps], 1 instance).

Table 37. Name Given to the Category, Articulatory and Auditory Characteristics of the Category Instances, and Examples of the Category Provided in the Post-Experimental Interview.

S.	Name Given to the Category	Artic. Actions	Auditory Impressions	Examples	S.	Name Given to the Category	Artic. Actions	Auditory Impressions	Examples
1	LA Stronger <i>p</i> , more pronounced vs. softer, quieter <i>p</i>	These sounds come up with more breath	At the beginning < sp > looked like positive but them I ruled them out	< pin > < steep > < camp >	11	TS	Spitting out vs. more "hidden" <i>p</i> 's	<i>p</i> + <i>V</i> sequences stronger than <i>p</i> + <i>C</i> ones	+ < paved > + < pull > - < speed >
2	AR more explosive (at the beginning) vs. calmer, less explosive <i>p</i> .	More air than in any other <i>p</i>	Popping sound	- < spray >	12	AS	More air passes through the lips		+ < piles > - < depth >
3	HT Hard <i>p</i> (at the beginning) vs. softer <i>p</i> , more like a <i>b</i> .	Close lips	Crisp, sharp sort of spitting. Louder, hard, sharper than the negatives	+ < pin > - < camp > - < depth >	13	JC	Close mouth, force a bit of air out but less in the middle/at the end and softer	Aggressive, emphatic, and forthright word-initially	+ < push > - < up > - < steep > - < clasp >
4	DW <i>P</i> at the beginning, more of an explosive sound vs. <i>p</i> elsewhere & not as powerful, strong, or emphatic	Purse lips gives opportunity to aspirate the sound	Explosive, staccato. It interrupts the rhythm if too much emphasis is made. CONT	+ < pull > - < camp > - < wept >	14	WK	<i>P</i> at the beginning vs. <i>p</i> in the middle or at the end	More firm, more "in your face" vs. longer, more subtle and duller	+ < pull > - < opt > - < grasp > - < spear >
5	BM <i>P</i> requiring more explosion of air followed by a vocal sound vs <i>p</i> not requiring an active expression of the air	More curling of the lips, lip movement	Like the popping of a bubble. The release of air produced the <i>p</i> . unreleased in <i>apt</i> After <i>s</i> it's blunter, duller, more mute. CONT	+ < pull > + < push > - < apt > - < spa > - < depth > - < help >	15	PH	<i>P</i> coming at the beginning vs. <i>p</i> coming later on, in the middle or at the end.	Short, blown, hard. <i>p</i> + <i>V</i> harder than <i>p</i> + <i>L</i> (<i>p</i> slides into <i>l</i>). In <i>spa</i> or <i>glimpse</i> it's not emphasised.	+ < pull > + < push > - < glimpse > - < spa > - < clasp >
6	OK Hard, active <i>p</i> , usually at the beginning, vs. soft, passive <i>p</i> , more like a <i>b</i> .	More air passes through the lips. Push them forward. work it out	Dominant, muscular., precipitated. In <i>grasp</i> or <i>depth</i> <i>p</i> is almost unheard. Like a <i>b</i> after <i>s</i>	+ < plead > - < grasp > - < depth > - < lapse >	16	DB	More air is expelled	Stronger, more stressed, punching, forceful initially, less stressed and weaker in other positions	+ < push > + < priest > - < spy > - < camp >
7	AG Stronger <i>p</i> not between consonants vs. <i>p</i> word-medially & less strong, softer	The amount of air or puff of air is bigger	It is a sudden sound, not flowing, but abrupt	+ < pull > + < piles > - < lapse >	17	KD	A hard <i>p</i> , the first sound, vs. the 2 nd or the 3 rd sound and softer <i>p</i>	Like a punch, short, punchy. <i>p</i> + <i>V</i> is stronger than <i>p</i> + <i>C</i>	+ < power > + < pence > + < price >
8	HS Strong <i>p</i> , appearing in initial position vs. softer <i>p</i> , normally at the end of the word	More effort is made to expel the air	As if you were pushing something. In <i>grasp</i> the <i>s</i> is more dominant. CONT	+ < pull > + < grasp > - < spend >	18	RO	Stronger <i>p</i> 's vs. not so strong <i>p</i> 's	Very short and quick. In "camp" or "laps" it is unheard, but natural.	+ < post > - < lapse > - < depth >
9	NW Stressed <i>p</i> at the beginning vs. less stressed <i>p</i> at the end and softened.	The way you press the lips	Aggressive, sharp, short, abrupt, staccato. In <i>sp</i> , <i>s</i> takes over, is more dominant	+ < pull > + < please > - < spy > - < help >	19	DJ	Strong <i>p</i> vs. softer <i>p</i>	Stronger in initial position. Weaker elsewhere. Barely heard in <i>camp</i>	+ < paled > - < ship > - < spy > - < lapse >
10	RV <i>P</i> at the beginning of the word vs. <i>p</i> somewhere else	The way the mouth is shaped	Kind of strong, instant sound, not flowing (as "s" would be)	+ < push > + < pear > - < spy >	20	MM	<i>P</i> more pronounced vs. less pronounced <i>p</i>	Word-initial <i>p</i> is stronger and harder. In <i>lapse</i> or <i>caps</i> , <i>s</i> shadows the <i>p</i> .	- < spill > - < caps > - < lapse >

3.5.4. Discussion.

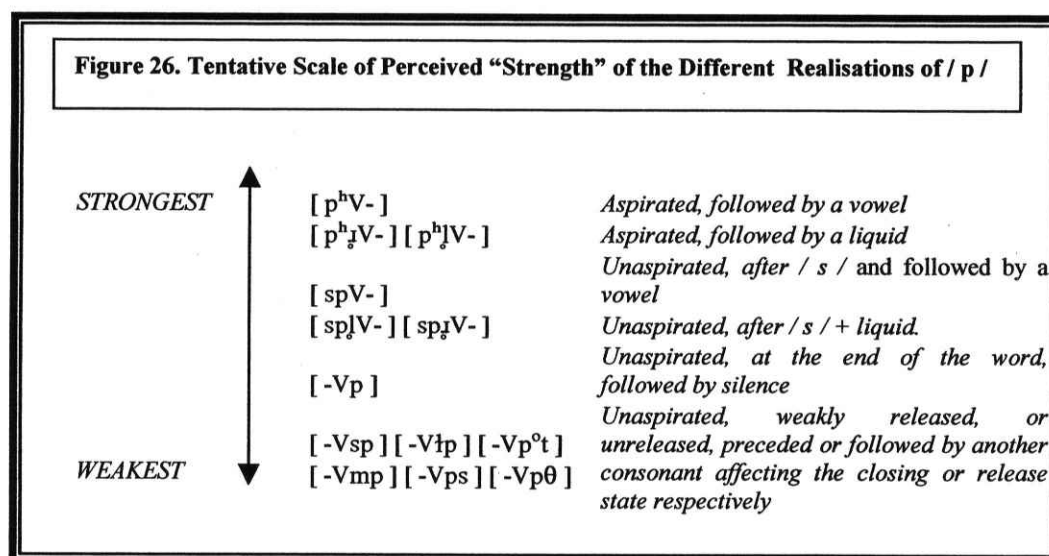
The experiment conducted above tried to find out whether the allophonic variant [p^h] of the phoneme / p / is perceptually salient for speakers and whether the latter are able to classify different instances of that realisation of the phoneme / p / to the exclusion of other phonetically different allophones. The investigation was motivated by the long-standing controversy in the phonological and/or phonetic literature over subjects' presumably limited ability to "hear" or discriminate the allophonic variants of phonemes. As far as voiceless stops are concerned, for example, Donegan and Stampe suggested (1979: 162), not long ago, that "speakers are... totally unconscious of the difference between e.g. initial [k^h] and non-initial [k], even in alternations like *crunch* : *scrunch*". Earlier, Pike had suggested (1947: 65), in reference to / p /, the voiceless plosive under investigation in the study above, that "speakers of English have difficulty in learning to distinguish between the two [p] sounds in *paper*".

However, the results of both the learning and test sessions of the experiment 4 clearly indicate that not only did all the experimental subjects realise which type of consonantal sounds all words contained (i.e. / p /), but also that they could discriminate between the positive instances and the negative ones remarkably well.

The results of both the learning and test sessions indicate that all the subjects who took part in the experiment formed the category correctly, reached the pre-established 37-correct-response criterion in the learning session with a mean of 51.40 correct responses (range 37-59, s.d.= 5.15). Subjects' successful formation of the category was further supported by the fact that they made significantly more correct responses in the test session (i.e. 97.5%) than in the learning session (i.e. 85.58%) and, consequently, fewer errors in the test session (i.e. 2.5%) than in the learning session (14.42%). In addition, subjects responded correctly to the positive and negative control stimuli (i.e. [p^hjV-] and [-Vmps] respectively) with extreme accuracy (100% correct responses in each case. Thus, the answer to the first research question, namely, whether native speakers of English can classify instances of a particular allophone of a phoneme category like [p^h] together to the exclusion of instances of other allophones of the same phoneme category is "yes" and the hypothesis initially entertained (i.e. that subjects would be able to behave in such a way) is confirmed.

However, the results of the present investigation show more than subjects' ability to set aspirated instances aside from those containing weak or absence of aspiration. The verbal reports produced by the subjects in the post-experimental interview show clearly reveal that

subjects were able to perceive the allophonic variation between the positive stimuli and different types of allophones representing negative stimuli. Subjects provided, during the post-experimental interview, similar descriptions of the type of sound that positive instances had in opposition to the negative items, emphasising the stronger, harder, or powerful nature of [p^h] as a distinguishing feature from other realisations, perceived to be softer, calmer, and weaker. Consequently, the answer to the second research question, namely whether subjects could *perceive* allophonic variation, is “yes” and the hypothesis entertained (i.e. that they would be able to show such an awareness) is again confirmed. It is also interesting to remark, in this respect, that the reports of a few subjects’ in the post-experimental interview indicate that, at least 3 subjects (4, DW; 5 BM; 8 HS) felt that the different realisations of / p / lay across a continuum of “strength” ranging from the strongest types of / p / included in the positive stimuli in the experiment to the weakest realisations (negative stimuli preceded or followed by another consonantal sound). Taking also into account different comments made by other experimental subjects (e.g. 5, BM; 6, OK; 11, TS; 13, JC) who did not made an explicit reference to the existence of a continuum but who compared specific pairs of allophones or referred to a specific characteristic of a given allophone (see section 3.5.2.2.), a scale of perceived strength in either the production or perception of the different realisations of the phoneme / p / could be tentatively proposed. Figure 26 shows this scale.



In retrospect, it might not be surprising that subjects found an allophonic variant like the oral aspirated bilabial plosive particularly salient. Given the extensive body of evidence on subjects’ ability to discriminate within-category differences in the “categorical perception” literature, it would have been puzzling to find that subjects are utterly insensitive to all of the allophonic variation inherent to any phoneme category.

The results of the present investigation reveal subjects' awareness of allophonic variation under experimental conditions. However, these are not the sole laboratory investigation in which such awareness has been obtained. In Treiman's (1985b) spelling study, subjects consistently spelled oral stops in [s]-clusters as voiceless but, when asked to give a phonetically plausible alternative spelling in a post-test interview, two thirds were able to spontaneously notice the phonetic similarity between stops after / s / and / b d g / in word-initial position or be induced to show such an awareness. In addition, looking back at the reports in the post-experimental interview of experiment 3 (see section 3.4.3.2.), many references were also made to the differences between different realisations of the phoneme / p /, although subjects were not required to focus on allophonic variation at any moment on that occasion.

However, an important issue should be addressed in relation to the results obtained in experiment 4. As mentioned in 2.2.2.1., a frequent criticism about experimental results is the fact that the unusual circumstances of the experimental situation (i.e. being in a laboratory, using a given apparatus, the feeling of being examined, etc.) may cause subjects to give unusual responses that they would not under normal circumstances. This would indubitably cast doubt on the validity of the results outside of the experimental setting and the external validity of the results obtained (i.e. the extent to which they are applicable to entities or people other than those tested in the experiment) would be questionable. Applied to the results of the present investigation, it might be the case that, as the experiment has shown, allophonic variation may be relatively perceptually salient under careful listening circumstances but that such awareness might be severely limited (or completely lacking) under normal listening circumstances in a real-world communicative situation. However, it should be remembered that this (i.e. showing that subjects perceive allophonic variation and are constantly aware of it) *was not* the aim of experiment 4. The aim of the experiment was simply to provide evidence on subjects' ability to perceive allophonic variation. However, a comparison of the experimental results with external evidence obtained in more natural or ecological circumstances might support the view that perception of allophonic variation is not a strange phenomenon occurring only in very special experimental situations or an artefact of the experimental situation. If experimental results are in accordance with conclusions drawn from observations of naturally-occurring phenomena believed to show the psychological and perceptual pertinence of a given phonological construct, then the validity of the experimental results are more strongly supported. Such type of evidence could be obtained, for example, from observations of spontaneous spellings by beginning spellers. For instance, although the English writing system does not represent allophonic variation, beginning spellers usually represent low-level phonetic variation in their spellings. Kindergarteners and first graders usually symbolise / t / and / d / before / r / with letters that are more appropriate for / tʃ / and / dʒ / and so, they spell "trap" as "chrap" and "drowned" as

“jrad” (Treiman 1985c, 1993: 142-145). These facts seem to reflect that beginning spellers try to represent allophonic variation (i.e. the affricated nature of / t / and / d / before / r /) in their spelling form, a behaviour that decreases as exposure to standard writing system increases and reading ability improves. In short, children are often sensitive to phonetic differences *within* phonemic categories (as well as to the differences *among* phoneme categories). As Treiman puts it:

Just as children hear the differences between the first sounds of *tap* and *cap*, so they hear the difference between the first sounds of *trap* and *tap*. Children sometimes misspell *trap* as *CHRAP* because they try to symbolize phonetic variation in their spelling.

Rebecca Treiman (1993: 145-146)

When could awareness of allophonic variation occur outside the laboratory occur?. Leaving aside phoneticians’ ability to perceive such differences, which presumably derives from an acquired perceptual sensitisation in the process of learning new concepts,⁸² O’Connor (1973: 121-123) discusses four non-laboratory circumstances under which allophonic variations may be spontaneously noticed by the average phonetically naive language user. These are (1) when an allophone occurs outside the sequences where it is customarily found,⁸³ (2) when allophones are produced in isolation, although this is presumably a rather marginal phenomenon in everyday speech, (3) when gross accentual differences occur, and (4) on some exceptional cases, like when allophonic variants of the same phoneme are distinctly different in their phonetic features. In addition, Nathan (1999: 313) mentions an additional circumstance: when the presence of an allophonic variant happens to activate a different lexical item, as when the voiced alveolar flap (i.e. [r]) in words like “latter” and “ladder” makes the hearer notice that the realisations of both / t / and / d / are very similar, if not identical.

To conclude this discussion, we would like to draw attention to the fact that the results of the present experiment merely suggest that speakers *can* perceive allophonic variants. This does not imply that such variants are perceived more or less frequently or that the claim that speakers usually discard allophonic variation is not correct. It is probably true that subjects do not focus on allophonic variation in most cases because the goals, needs, and characteristics of everyday speech do not require this. However, this does not mean that allophonic awareness or the ability to distinguish between allophonic variants is limited as many phonologists and phoneticians

⁸² The process of perceptual sensitisation was discussed in section 1.1.1., devoted to the relationship between perception and categorisation (see also Goldstone 1994a; MacKintosh 1995: 196-1998). Applied to the perception of allophonic variation, we could summarise the point by saying that experts like n other words, experts like phoneticians or phonologists may perceive allophonic variants because they have learned many different phonetic concepts while developing their expertise.

⁸³ As Lass says (1984: 17), substituting the [k^h] of *kill* for the [k^h] of *cool* may give us a rather odd version of *cool*, but one that’s still *cool* and no other word. But if we substitute [g] for [k^h] we get *ghoul*. So the difference between front and back velars in English is not lexically information-bearing but it may provide indexical information (i.e. about speaker-characteristics, like sex, effeminacy, foreignness, etc.). and may thus be noticed by the speaker.



seem to imply.⁸⁴ An example drawn from a different issue may clarify the point. Discussing the perception and/or processing of words, Cutler and his co-workers claim:

Words appear to be apprehended as wholes, and we certainly do not have the impression of processing them phoneme by phoneme. Certainly, listeners *can* pay attention to phonemes; for instance, we can easily notice a speech error involving an exchange of phoneme ("worken spurds"), or realize that a speaker has consistent difficulty with a particular sound. But, in general, our attention is not devoted to the level of individual sounds.

Anne Cutler, van Ooijen, Brit., Norris, Dennis, & Rosa Sánchez-Casas (1996: 807).

By analogy we may say that speakers can detect allophonic variants, that some of them may be particularly salient on certain occasions for various reasons but, in general, our attention is not devoted to the allophonic level. However, we might be justified in claiming some psychological and/or perceptual reality to this level and, consequently we can argue that allophones are psychologically and/or perceptual units that a psychologically-oriented phonology should strive to represent.

⁸⁴ It is also important to emphasize that the results of the present experiment do not indicate whether words are stored allophonically or not, neither was this one of the aims of the present investigation. In this respect, differences also exist with two opposite views. One claims that allophonic variants are not stored in long-term memory (e.g. Donegan & Stampe 1979: 162-165; Nathan 1996, 1999: 312) while others claim that words are stored with allophonic variants (e.g. Bybee 1994, 1999).

CHAPTER IV.

CONCLUSIONS.



4.1. Introduction

The purpose of the present and final chapter is to briefly summarise the connections between the previous chapters. The chapter also aims at comparing the results of the four psycholinguistic phonological experimental studies. It also suggests directions for the future application of the CF experimental technique in phonology for addressing psycholinguistic issues as well as other in other disciplines concerned with the study of sounds or the sound patterns of languages. The chapter ends with a final reflection on the role of experiments in phonology practised with an actual psychological commitment.

4.2. Categorisation, The Categorisation of Sounds, Phonology, and Experimentation.

Chapter I of the present dissertation started with an emphasis on the importance of the cognitive ability of categorisation for human perception, memory and cognition, behaviour, and language. The chapter moved on to a review of the substantial experimental research conducted with human adults on their categorising abilities when confronted with (mainly) visual stimuli. In this respect, the chapter laid out the history of categorisation research for the past 80 years and the categorisation models proposed to account for categorisation and the nature of mental categories (i.e. the classical view, the probabilistic view, the exemplar view, the mixed view, and the theory-based view). Chapter I also reviewed the extensive body of evidence gathered about the categorising abilities of other subject populations like children, infants, and non-human species, showing that categorisation is basic for all these populations as well as pointing out similarities and differences between human adults and the other subject populations mentioned.

Chapter II began with a review of different studies on the categorisation of sounds that different subject populations (human adults, children, infants, and non-human species) make.

The enormous amount of studies conducted throughout the years on subjects' ability to categorise sounds revealed that the ability to categorise has been as closely examined in the auditory domain as in the visual one, although usually with different purposes. Special emphasis was made on the experimental techniques used to obtain such evidence. Next, the bearing of the categorisation of sounds for the field of phonology was explained. It was argued that the study of the categorisation of the sounds of a given language by its users might help the phonologist interested in providing phonological descriptions with psychological claims shed light on controversial phonological problems and thus, help provide phonological descriptions that reflect the knowledge (in a broad sense) that speakers have about those controversial phonological issues. At this point a distinction between the goals of descriptive phonology and psychological phonology was explicitly made. As categorisation is such an important cognitive function, it was reasoned that categorisation behaviour might shed some light on the knowledge speakers have about the sound pattern of their language. It was argued that experimentation is an appropriate *modus operandi* if a true psychological phonology is to be pursued and that experimental evidence is the most useful type of evidence to validate psychological claims in phonology. This was done after the most frequent criticisms about the use of experiments in phonology were reviewed (and answers to those criticisms were provided) and the advantages of experimentation in phonology were discussed. The criticisms and advantages of experimentation in phonology were dealt with in order to better understand the point of experiments in the practising of a phonology of the type advocated in this dissertation.

4.3. CF as an Experimental Paradigm

Among the most promising experimental techniques used to study the auditory categorising abilities of adults and children is the well-known experimental technique known as "concept formation". The specifics of this technique were extensively discussed at the end of Chapter II. In addition, an specific implementation of the CF paradigm used to address the four different phonological issues addressed in this dissertation was described at the beginning of Chapter III. The motor-response computer-based version of the CF used in this work was used to obtain empirical psycholinguistic evidence on four given phonological questions over which considerable controversy has arisen throughout the years in the phonological and/or phonetic literature. The characteristics of the experimental subjects, the stimuli used, the general arrangement of the tasks, etc., were also discussed at the beginning of chapter III.

4.3.1. Evidence for Specific Categories: Within-Category Facts.

Having then discussed the characteristics of the CF experimental technique and the specific implementation used in this dissertation, the four phonological questions under investigation, Chapter III provided an account of four different experiments conducted with the CF experimental technique and laid out the results of such investigations. The data obtained provided answers to the different research questions posed at the beginning of each experiment. The results suggested that, at least for educated and phonetically native English-speaking adults, the four categories investigated were somehow psychologically relevant in that the subjects can form them, satisfactorily generalise their responses to new instances without the provision of feedback (i.e. making significantly more correct responses in the test session than in the learning session), and later verbalise their knowledge in a post-experimental interview.

As far as the category “consonantal sounds” is concerned, 12 out of the 20 subjects who took part in the experiment 1 formed the category. The results of experiment 1 also showed that subjects tended to categorise word-initial pre-vocalic [h], [w], and [j] as members of the category “consonantal sounds”. The results also revealed that the conventional spelling of the target sounds seemed to be a factor conditioning subjects’ responses to some extent. More specifically, when the target sounds were spelled with letters of the alphabet typically referred to as “vowels” (e.g. “o” and “u”), as in “once” or “use”, subjects tended *not* to classify the target sounds as “consonantal sounds” significantly more often than when the target sounds were spelled with “consonant letters” (i.e. “y” and “w”). Orthography appeared then to be leading subjects to include or exclude some of the target sounds as instances of the category “consonantal sounds”, even if subjects are aware that they only have to pay attention to sounds, which, by the way, they seemed to do throughout the experimental task.

Regarding the category “oral plosives”, the results of experiment 2 showed that 14 out of 20 experimental subjects formed the category. The results of the experiment also show that there may have been some sort of phonetic interference in subjects’ formation of the category (namely, the presence of nasal consonants) although its role may not have been very strong. More interestingly, the results also showed that / tʃ / and / dʒ / were considered more frequently as non-members of the category “plosives” (/ tʃ / = 68.57%, / dʒ / = 60.71%) than as members (/ tʃ / = 31.43% and / dʒ / = 39.28%). This fact seemed to show that, if subjects’ criterion is mainly phonetic, as so it seemed, they may have considered the fricative-like release of the affricates more important than the occlusive and/or hold stage. In other words, although subjects may be sensitive to the two-part nature of affricates, the fricative part seems to divert subjects from considering the affricates as instances of the category “plosives” more often than not.



The results of the third experiment showed that all the subjects who took part in the experiment (i.e. 20) formed the category correctly, with neither phonetic nor (a substantial) orthographic interference in the formation of the category. The results also show that subjects overwhelmingly considered bilabial oral stops in [s]-clusters as members of the category / p / based, presumably, on phonetic (i.e. the fact that the plosives *have to be* bilabial), distributional (the fact that they appear *after* [s]), and orthographic criteria (the fact that they are spelled with “p”).

Finally, experiment four showed that all the experimental subjects who part in that experiment (i.e. 20) could discriminate the allophonic variant [p^h] of the phoneme / p / from other different phonetic realisations of the same phoneme. The results clearly demonstrate that subjects not only learned which type of consonantal sounds all words contained but also that they could discriminate between the differences between the positive instances in opposition to the negative instances, namely the presence of strong aspiration vs. weak or lack of aspiration, and that they associated these perceptual differences with the positions in the syllable where such differences occurred.

4.3.2. Evidence for Between-Category Similarities and Differences.

After having conducted the four different experimental studies reported above and looking at the results of them all it is interesting to note that some coincidences as well as divergences can be detected. Amongst the coincidences we will highlight the type of information used to refer to the stimuli in the post-experimental interviews and the influence of spelling. The differences between the four different studies will be discussed in relation to the possibility that the categories studied are taxonomically structured in the conceptual system.

Subjects’ formation and understanding of the categories under investigation was further supported by the rich and relatively homogeneous reports provided by the experimental subjects in the post-experimental interviews with synonymous “folk” expressions of many of the technical terms used in the specialised literature. The post-experimental interviews showed that subjects could verbalise their knowledge about the sound categories that they had formed.

What aspects of sounds did subjects pay attention to? Bloch and Trager (1942: 11-12) claim that a speech sound is a physical event with three main aspects: 1) *physiological* (or articulatory); 2) *acoustic*, (related to the vibrations of air molecules propagated as “sound waves” through the air); and 3) *auditory* (related to the perception of the sounds). They further claim that, in popular speech, terms descriptive of speech sounds are almost exclusively auditory and they refer, in a vague way, to the hearer’s impression. According to them, for instance, speakers regard “the vowel of *cash* is ‘flat’, the vowel of *calm* ‘broad’; the *g* of *get* is

'hard', the *g* of *gem* 'soft'... Other sounds are said to be 'bright' or 'dull' or 'thin', or 'harsh', even the term 'guttural' is used merely as a vague synonym of 'strange' or 'unpleasant' ”.

The reports of the post-experimental interviews of the four different experiments clearly indicate that the auditory impressions produced by the sounds heard were important for the criterial speakers. However, our results also show that subjects also described the physiological part of the speech sounds they heard, referring to the articulation of sounds in some cases in a very precise way for which speakers seem to possess a large and rich folk phonetic terminology of a rather vague or inaccurate value for the specialised phonetician but quite revealing for the experimentalist interested in the way speakers conceive of speech production and perception.¹

Another point of coincidence between the different experiments is the fact that orthography revealed itself as having a powerful effect on subjects' formation and/or conception of phonological categories. In fact, orthographic interference is an important potential source of artefacts in all phonological experiments involving literate subjects (Nearey 1981),² even those dealing with speech segments also investigated in this dissertation like the so-called semivowel (e.g. Cutler *et al.* 1990; van Ooyen *et al.* 1992) or the plosives after tautosyllabic [s] (e.g. Fink 1974; Jaeger 1980a, 1980b; Wolf 1973)

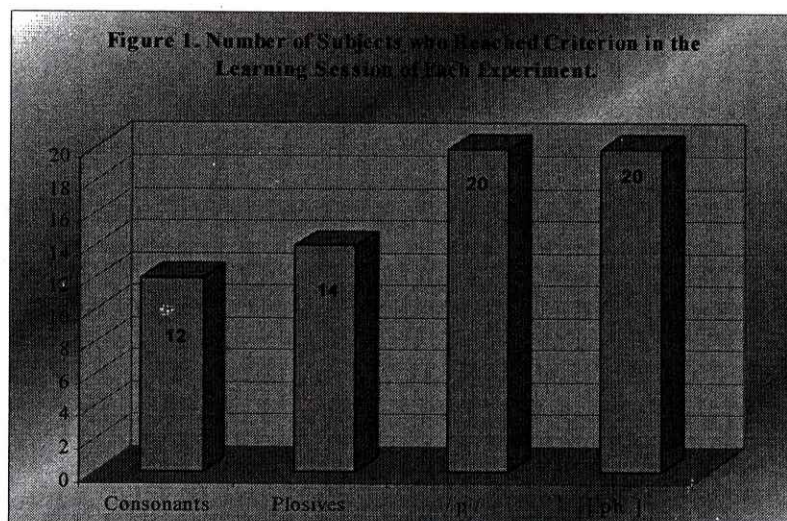
Finally, it should be mentioned that although the four categories investigated in chapter III were psychologically real for the criterial subjects, it is obvious that some categories seemed to be more prominent than others or at least easier to form in an experimental session. This can be inferred from different sources of evidence that are comparable across the four experimental studies, namely 1) the number of subjects who reached criterion in the learning session, 2) the average number of correct responses in the learning session of each experiment, 3) the standard deviation of the individual scores of correct responses in the learning session, and 4) the percentages of correct responses to positive and negative stimuli in both the learning and test sessions in each experiment. The first three pieces of evidence are shown in Table 1 and the percentages of correct responses in Table 2. Figure 1 also shows the number of criterial subjects per experiment.

¹ The acoustic phase is probably conceived of metaphorically in terms of the so-called *conduit* metaphor (Reddy 1979), which is essentially the preferred framework for conceptualising communication. According to that metaphor, language functions like a conduit, transferring thoughts from one person to another. In the first place, people insert their thoughts, ideas, or feelings in the words. In the second place, the words accomplish the transfer by containing the thoughts or feelings and conveying them to others. In the third place, in listening, people extract the thoughts and feelings once again from the words. Words, are, then, containers with "insides" (wherein meanings, thoughts, ideas are inserted, reside, or are extracted) and "outsides" that are ejected by speaking into an "external space". Speakers probably resort to that metaphor in order to talk about the acoustic propagation of sound waves, conceived of as containers or things that travel from the speaker's mouth to the listener's ear.

² There is now an increasing body of evidence that orthography plays an important role in phonological awareness, reshaping of phonological systems in the child as he is exposed to the written language, and in adults' organisation and classification of the sounds of his/her language. For example, different phonological studies on the psychological status of the Vowel-Shift rule (Jaeger 1980a, 1984; Moskowitz 1973; Treiman 1987; Wang & Derwing 1986) coincide to point out that the psychological basis for these vowel alternations stems from English spelling rules. That is to say, what these five alternations have in common is that each one represents what are often taught in grammar school as the "long" and "short" versions of the five written vowels in English.

Table 1. Number of Criterial Subjects per Experiment, Subject's Mean Correct Responses, Range and Standard Deviation in the Learning Session.				
Exp.	Category	Criterial Subjects	Average No of Correct Responses in LS (max: 60)	Range & Standard Deviation
1	Consonantal Sounds	12	48.67	s.d. = 6.50
2	Oral Plosive Sounds	14	47	s.d. = 7.06
3	/p/	20	56.55	s.d. = 2.06
4	[p ^h]	20	51.4	s.d. = 5.15

Table 2. Percentages of Correct Responses (PCRs) to Positive, Negative, and Positive + Negative Stimuli in both the Learning and Test Sessions.							
Exp.	Category	PCRs to Positive Stimuli		PCRs to Negative Stimuli		PCRs to Positive + Negative Stimuli	
		LEARNING SESSION	TEST SESSION	LEARNING SESSION	TEST SESSION	LEARNING SESSION	TEST SESSION
1	Cons. Sound	83.85%	95.61%	85.12%	88.89%	84.44%	93.01%
2	Oral Plosive	81.25%	93.23%	73.98%	83.33%	77.86%	89.40%
3	/p/	93.44%	99.74%	95%	97.08%	94.17%	98.71%
4	[p ^h]	86.56%	98.64%	84.46%	96.11%	85.58%	97.5%



The comparisons are possible because, as was mentioned in section 3.1.1.5., the number of positive and negative stimuli were the same in the learning sessions of each experiment and

the stimuli appeared in the same order.³ In addition, subjects had received very similar experimental treatments.

The results clearly show that the category “phoneme / p /” was the easiest to form of the four. / p / was, together with [p^h], the category that most subjects formed. However, although all subjects also learned the more specific category “allophone [p^h]”, subjects in experiment 3 made more correct responses in the learning session than the subjects in experiment 4 and the standard deviation of the range of correct response scores was higher for experiment 4 than for experiment 3. In other words, although the 20 experimental subjects of experiments 3 and 4 formed the categories / p / and / p^h /, the subjects forming / p / made more correct responses in the learning session as an average and the scores of the different subjects differed less than those of the experimental subjects who formed [p^h]. In addition, the subjects in experiment 3 made more correct responses to both positive and negative items in both the learning and the test sessions than in experiment 4. These two categories are, however, far more salient than the categories “consonantal sounds” and “oral plosives” (experiments 1 and 2 respectively), which were more or less as difficult (or easy) to form (the category “consonantal sounds” had fewer criterial subjects but these performed a little better than the subjects in experiment 2).

In retrospect, it is not at all surprising that the categories “oral plosive” and “consonantal sounds” were more difficult to form than the categories [p^h] and, in particular, / p /. Previous studies have found that, for both adults subjects (e.g. Ohala 1986) and infants (e.g. Fodor *et al.* 1975), categories defined by a given phonetic feature and instantiated by different speech segments that do not belong to the same phoneme category are more difficult to form than categories instantiated by speech segments that are classified as members of the same phoneme-sized category. For example, Ohala (1986: groups 3 & 4) taught two groups of English-speaking subjects two opposing categories: one group was taught to group the [k] in a word like *school* with [k^h] as in *cool*, and the other group with [g], as in *ago* and [ɡ], as in *good*. The first group formed the category easily but many subjects in the second group could not form it at all, and those who did described the category in a disjunctive way: “[gə] sounds or the [k^hə] sound after s”. Ohala concluded that the findings revealed that [k^h, k] was more likely to be a pre-established grouping for subjects than [g, ɡ, k] was, which seems to be a grouping based on the features “velar”, “stop”, and “oral”, but whose instances do not belong to a single phoneme. In addition, Jaeger (1980a: 367), who studied the category / k / as well as the category “two-consonant clusters” found out that the percentage of subjects who formed the monosegmental (i.e. / k /) or bisegmental (i.e. “two-consonant clusters”) categories was higher

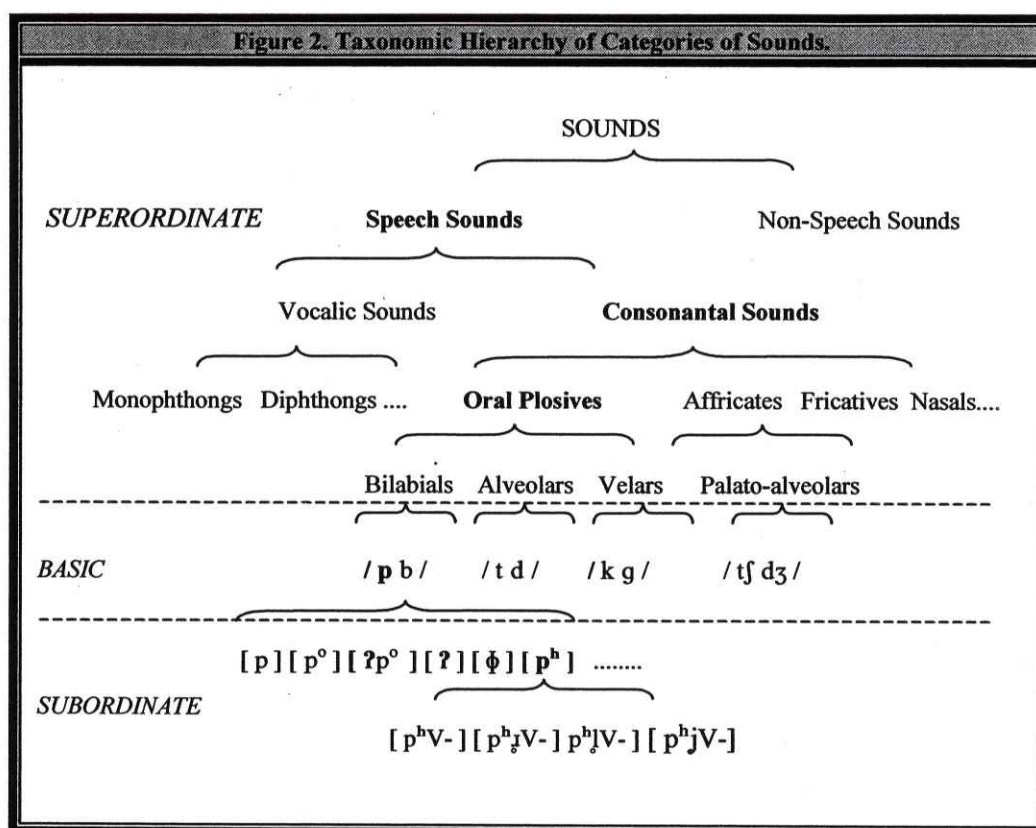
³ There were 32 positive stimuli and 28 negative in the learning sessions of each experiment and 19 positive and 12 negative in the test sessions of all experiments except for experiment 4 where no test items were included. In this experiment, there were 22 positive and 18 negative but the extra positive stimuli and negative appeared in the places where the test items had appeared in the previous experiments so the order was respected.



(i.e. 100% with either a motor-response or a verbal-response CF technique) than the percentages of criterial subjects in feature categories like “[-anterior]” (79%), “[+anterior]” (50%), “[+sonorant]” (50%), and “[+voice]” (50%, in word pairs; 38%, in single words). In addition, the mean number of trials to criterion was fewer in the / k /-experiments and the “two-consonant clusters” task (19 and 23 respectively) than in the feature categories ([-anterior]: 33; [+anterior]: 45; [+sonorant]: 31; [+voice] in word pairs: 54; [+voice] in single words: 43). Finally, Fodor and his co-workers (1975) found that infants grouped syllables beginning with / p / (e.g. / pi /, / pu /) more readily than syllables sharing phonetic features (/ pi /, / ka /, the feature being voiceless plosive) but not a segmental grouping.

The relative greater salience of / p / over the rest of the categories seems to suggest that the “phoneme” level may have some of the characteristics of the basic level of taxonomies (see section 1.2.1.3.5.). This is so because the categories or concepts under investigation in each of the experimental studies can be conceived of as occupying different levels of abstraction in a taxonomy of sounds. Thus, as figure 2 shows, from the top to the bottom of a hierarchy of speech sounds, “sounds” might be conceived of as including the categories “consonants” (or consonantal sounds) and “vowels”. Consonants might, in turn, include different categories like “plosives”, “nasals”, “fricatives”, etc., while the contrasting category (i.e. “vowels”) would include different categories like “monophthongs”, “diphthongs”, etc. The category “oral plosives” might include further subcategories of point of production (e.g. bilabial, alveolar, velar) or fortisness (this is not represented in the figure)⁴ and these could be further divided into the different “phoneme” categories like / p /, / t /, / k /, / b /, / d / or / g / and each phoneme category would, in addition, include allophonic categories like [p^h], [p^o], [ϕ], etc. Finally, some types of allophones like [p^h] might be further divided into contextual realisations of the variant (depending on the following sound). In other words, as far as the categories under investigation are concerned, [p^h] is a type of / p /, which is in turn a type of oral plosive which is a type of consonant and this, a type of speech sound. Sounds may then be classified and subclassified (Swadesh 1934: 121).

⁴ Sounds can also be cross-classified (Swadesh 1934: 121). For instance, / p / can be an oral plosive, but also a bilabial plosive and a fortis plosive.



Why would the “phoneme” level be basic? Apparently, it seems that it is easier to learn in a concept formation task, as the present results suggest. In addition, it seems that the writing system of alphabetic languages like English represent that particular level but do not reflect allophonic variation or higher-order phonetic and/or phonological relations. Be that as it may, and despite of occasional suggestions about the basic-level status of phonemes (Jaeger 1980a; Mompeán-González 1999), it seems that future research should look at this issue in more detail.

4.3.3. Experimental Evidence and the Solution to Phonological Problems.

The findings of the present investigation, although highly suggestive, should be taken with caution. In fact, the results of psycholinguistic experiments should always be evaluated carefully. Definitive conclusions about a given issue should never depend on the outcome of one experiment or small group of experiments. The results of any one study are always tentative and inconclusive in certain respects, and may always be subject to criticism of some sort and possible rejection (Ohala & Jaeger 1986b: 4). In this respect, the present results could be criticised for many different reasons. In the first place, an admitted deficiency in experiment 1 is also that although the positive stimuli were semi-randomly ordered, the first six positive stimuli were plosives and, as a great rate of incorrect or null responses to the positive stimuli occurred precisely in these stimuli (40.28%) –stimuli 1 to 9-, this gives the presumably inaccurate view

that the plosives were more difficult or less clear examples of the category “consonantal sounds” and that the affricates, which occurred much later in the task, were inherently easier examples of the category. Positional effects, which should have been more carefully controlled for, could also have occurred in experiment 2 and, most importantly, in experiment 3, providing the impression that some allophones of / p / are more difficult than others. In the second place, it could be argued that more interfering items should have been used in some of the experiments. In experiment 2, for instance, the digraph “ph” was not included as a possibly interfering spelling (it was included, however, in experiment 3). Earlier, in experiment 1, only one stimulus with potential orthographic interference was used (i.e. “hour”) and, consequently, the possible role of orthographic interference in the formation of the category in the learning session was not sufficiently explored. The reason for the sole inclusion of “hour” was that no other monosyllabic word started with silent *h*, which leads to a third possible criticism of the stimulus lists employed in the experiments, namely their limitation to monosyllabic words. If the stimuli had not been limited to monosyllabic words, more items causing potential orthographic interference like “honour”, “honest”, etc. could have been included in experiment 1, more allophonic variants of the phoneme / p / in experiment 3, or more realisations of [p^h] not restricted to word-initial position (e.g. “appall”) in experiment 4. The criticism is, in short, that the use of monosyllabic words was very restrictive and that polysyllabic words would have been more useful in extending the range of manifestations of category instances and/or negative items containing potentially interfering features. A related criticism is that the use of single, unrelated words does not allow to examine the classification of allophonic variants that occur in longer utterances in which the speaker is presumably more likely to produce realisations of a particular phoneme that will not occur in the careful pronunciation of single words.

Finally, it could also be argued that insufficient test items were included in the test sessions of some of the experiments. For example, more test items including [j], and [w] could have been used in experiment 1 so that the role of the conventional orthographic spelling in the categorisation of those types of speech segments could have been more closely assessed. In addition, the presence of / dʒ / spelled with “g” in experiment 2 might have shed more light on the (apparently minimal) role of spelling in the classification of the affricates.

Given all these possible criticisms and acknowledging the fact that the experimental studies conducted in chapter III pose as many questions as they try to answer, it seems reasonable to claim that final judgements about many of the phenomena observed in those experiments should be suspended until a larger body of evidence from varying sources has been gathered and findings have been replicated several times. However, if contradictory results were obtained in the course of so doing, it should be made clear whether the previous results or the ones obtained later are due to wrong interpretations, poor methodological practice, or the

presence of unanticipated confounding variables. Comparing different experimental studies may help the researcher move onto a new investigation with refined experimental methods, better-devised methods, and better-conducted experiments that will examine the phenomena under investigation in more detail. However, it is precisely one of the virtues of experimentation that, despite somehow inconclusive and sometimes contradictory results, higher and higher quality evidence can be increasingly gathered so that a point may be reached when a specific problem or phenomenon will be more or less solved, a rather unusual situation in phonology at present but by no means impossible.⁵ Once a larger body of experimental evidence is obtained via the spiral process of hypothesis formulating, hypothesis testing, hypothesis revision and revised hypothesis testing, a convergence of results supporting more confidently held beliefs may be reached (Ohala 1995: 721). This should be attained by making recourse to different experimental techniques and trying to improve, if possible, the implementations of the experimental techniques used.

The use of other experimental methods to confirm or not the results of a given investigation will, in turn, help to validate the external validity of the experimental results obtained, that is, the extent to which they are applicable to entities or people other than those tested in the experiment (see section 2.2.2.1.). It seems that in phonology, as in any other field of psychological science, there is an imperative need to obtain different converging lines of evidence for any theory (Hampton & Dubois 1993: 24) as “theories which unify results from many methods are more robust and more predictive, on the average, than those based on fewer methods” (Pierrehumbert *et al.* 2000: 280). This is a claim also made explicit in Goldsmith’s assertion that “it is truer today than ever before that the studies bridging components and methodologies are critical for the testing and refinement of phonological theory” (1995: 20). This will prevent phonology from one of its most common pitfalls: the fact that approaching a problem in an isolated way tends to produce a kind of myopia which prevents one from seeing and taking advantage of hints derived from other domains (Ohala 1990). Myopia starts not so much from an inability to make connections but from neglecting data from multiple sources like different experimental techniques.

For these reasons, in order to provide a more conclusive solution to the phonological problems under investigation new better-designed and more carefully-conducted CF experiments should be devised and performed in an attempt to examine the relevant phenomena in more detail. Also the phenomena under investigation should be tested with other experimental techniques.

⁵ An example can be seen in relation to the psychological validity of the vowel-shift rules of the original Chomsky & Halle (1968) analysis. On the basis of an extensive body of experimental results, Wang and Derwing (1986: 113) claim that the controversy surrounding the nature of this particular alternation set has been empirically resolved so that researchers could then turn their energies to other phonological problems with a sense of satisfaction and relief somewhat unusual in phonology.



4.3.4. The Utility of the Concept Formation Paradigm.

Despite the tentative results obtained in this dissertation and the need to replicate them through cross-methodological research, the present investigation hopes to have shown by itself the utility of the experimental paradigm for testing phonological issues, a point also made earlier (previous statements about the usefulness of the concept formation technique to study psycholinguistic issues in phonology (e.g. Jaeger 1980a: 381, 1986: 216; Ohala 1986: 18)). The dissertation also hopes to have subscribed the statement that the more convincing way to show that experiments can help to answer questions in phonology than by showing how answers to phonological questions can be provided through actual experiments (Ohala & Jaeger 1986a: 6).

In this respect, the technique may be considered a very valuable for different reasons.

First, the technique is very flexible, a relatively simple one and one of the easiest to implement (Jaeger 1986: 216; Ohala 1983: 18). In addition, the technique has advantages over other experimental methods like traditional identification and/or discrimination techniques used in speech perception research.⁶ This makes the technique a potentially attractive one.

Second, the CF technique permits to study speakers' perception, processing and categorisation of either naturally-occurring auditory stimuli or speech-synthesised ones. Although originally and extensively used with artificial visual stimuli (Rosch 1977: 180) and lately with natural or artefactual categories (e.g. Rosch 1973b), the technique has been used in the fields of speech perception (e.g. Weitzman 1990a, 1990b, 1992, 1993), and experimental phonology (e.g. Bertinetto 1992; Jaeger 1980 a, 1980b; Jaeger & Ohala 1984; Ohala 1983, 1986) to study the categorisation of sounds. Future research may shed more light on the categorisation of sounds through the use of this technique.

Third, the technique provides very useful behavioural evidence on the different questions under investigation.

It is therefore surprising that, given the relative advantages (and successes) of this experimental paradigm, it has almost gone unnoticed amongst researchers in phonology or linguistic in general. One cannot but wonder, however, whether such indifference to the technique is the result of a wider indifference to experimentation in linguistics and/or phonology as it has traditionally been practised. In this respect, the present dissertation hopes to have contributed to the consolidation of this technique for psycholinguistic phonological experimentation and to have shown its actual validity and potential.

In short, the CF technique is an important technique with which phonology can make much progress. Utilising it with some frequency may constitute a methodological advance and

⁶ The problem with these paradigms is, Weitzman says (1993: 141), that they allow experimenters no control over the past experience of subjects; they merely reflect their past experience. As a consequence, they are not useful to determine such things as the range of learnable phonological categorisations and what constraints there are on the usability of these categories in natural language.

methodological advances, as Pierrehumbert and her co-workers claim (2000: 280), can be just as important as theoretical ones in the progress of science. In addition, it constitutes an important step in phonologists' striving to "to throw open the gates of our science to the data and methods from any source that proves useful" (Ohala 1983: 232) to provide solutions to problems in phonology. We hope to have carefully followed Ohala's recommendations once again:

Our discipline should continue to augment the arsenal of experimental methods available to grapple with the issues that confront it. But it is also true that currently available methods can already provide solutions to many current issues if phonologists would just use them.

John Ohala (1986: 22).

4.4. Directions for Future Research

The aim of this final section is to suggest directions for the future application of the CF technique to other issues in phonology and related disciplines, for which this experimental paradigm might also be a methodological improvement in the imperative and unavoidable task of obtaining evidence to validate their claims. In fact, the potential usefulness of the CF paradigm has already been envisaged outside the field of experimental phonology. For instance, not long ago, Weitzman claimed that the use of the CF paradigm would help speech perception researchers increase their knowledge of that field of research:

In the future the application of this method can be expected to advance our knowledge and understanding of this still largely underexplored territory.

Raymond Weitzman (1993: 149).

The range of possible phonological topics in phonology that can be investigated experimentally is enormous. In principle, "no claims in phonology are above doubt: the existence of the phoneme, syllable, or the feature.... all of these are potential subjects for experimental study" (Ohala 1983: 714). A few topics that can be examined with the CF technique will be commented on below.

In the first place, the CF technique can be used to continue to provide further and more insightful evidence on the problems with which we have been concerned in the present dissertation and/or closely related ones. The results of the experiments reported above have provided some understanding of the phonological questions studied but the designs and results also suggest a number of issues that need to be investigated through further experimental research. In the first place, the voiced post-alveolar lateral / l / and, in particular, the voiced retroflex continuant / r / are usually realised as frictionless vowel-like sounds that have also been subject to some controversy regarding their status as members of the categories consonants

or vowels.⁷ In addition, as mentioned in section 2.3.4., the role of orthography was barely ascertained and more evidence is needed on its possible influence on subjects' categorisation of [h], [w], and [j] as members of the categories "consonants" or "vowels" or even on the distribution and/or function of those segments in the phonetic and/or phonological syllable.

In the second place, the category status of affricates could (and should) be investigated in more detail. As mentioned in section 3.3.4., the results of experiment 2 suggested that the palato-alveolar affricates were far from clear, 100%-of-the-time members of the category "oral plosives". Consequently, further research should clarify whether affricates are considered as fricatives or whether they deserve, according to language users' judgement, to be grouped into a different category with possibly, other affricated clusters like [tʃ] or [dʒ]. For instance, a CF experiment could teach subjects to form the category "fricatives" in a learning session and then look at subjects' classification of / tʃ / and / dʒ / in the test session as members of that category or not. In the third place, CF studies could be designed to investigate how speakers classify oral alveolar stops after initial [s]. Given that both the oral velar stops after tautosyllabic [s] (e.g. Jaeger 1980a, 1980b; Ohala 1983, 1986) and bilabial stops (Sawusch & Jusczyk 1981; experiment 3 this dissertation) have already been investigated with the CF paradigm, the categorisation of oral alveolar stops after word-initial / s /, which remains untested with the CF experimental, appears to be the immediate step to be followed in the solution to the classification problem of those stops. Finally, the CF paradigm could be used to continue to provide evidence on speakers' perception (and conception) of allophonic variation. As mentioned in 4.3.3., a CF experiment could be conducted in which aspirated realisations of / p / occur as positive members of the category "allophone [p^h]" in word-final position in an attempt to determine whether the concept is more or less independent of distributional factors. Although subjects' reports in the post-experimental interview reveal that even those who simply described the positive stimuli and the negative stimuli referring to their positions in the word perceived auditory differences between the two types of stimuli, no aspirated / p /'s were included in final position because articulatory emphasis on the stimulus words was avoided; consequently, it should be more conclusively established whether subjects associated the positive stimuli with the initial position of words but did not hear any differences between the positive and the negative stimuli. Some other experiments could be conducted to examine whether subjects can discriminate other allophonic variants and where the limits of allophonic perception lie.

Also, the CF experimental methodology could be used to further investigate any number of psycholinguistic issues within the domain of a specific language. First, following the traces

⁷ For all the details concerning / l / and / r / see Abercrombie (1967: 50), Brown (1977: 23-27), Christophersen (1956: 123), Gimson (1978: 151, 200-201, 205-207), Jones (1989: 173-176, 205), Kreidler (1989: 26-28), Ladefoged (1982: 61-62), O'Connor (1973: 148-150), and Roach (1983: 43-44, 48-49).

of experiments 1 and 2 and the studies of Jaeger and Ohala (Jaeger 1980a: experiment 6; Jaeger & Ohala 1984), different “feature” categories could be investigated. Amongst these, it would be interesting to further look at categories of manner of articulation, like “fricatives”, “affricates”, “nasals”, etc. If a feature is, according to Ladefoged (1982: 38) “a phonetic property that can be used to classify sounds”, it may well be that subjects’ may form many different “feature” categories to classify the sounds of their language. Regarding the category “fricatives”, a CF experiment conducted to teach subjects the category “fricatives”, for example, could be conducted to shed light on the controversial issue of the membership of / h / in the category “fricatives”. The results of experiment 1 showed that subjects considered / h / as a consonant. However, as the word-initial realisations therein (e.g. “hot”, “hard”) have no local friction, but cavity friction, which voiced vowels also have, it is dubious whether / h / would be classified as a fricative or not. If word-initial [h] were not classified as a fricative but it is still considered as a consonantal sound, further CF experiments could be conducted to find out whether its position in the syllable or its orthographic representation are responsible for its deserving to be considered as a consonantal sound. Studies on the category “fricatives” could also be conducted, as mentioned earlier, to look at the categorisation of [tʃ] and [dʒ] and compare the results with the ones presented in experiment 3.

Second, in the spirit of experiment 3, CF studies could be designed to study questions about the grouping of allophones into phoneme categories and shed light on the categorisation of controversial allophones as members of certain phoneme categories. One example might be the controversial classification of voiced alveolar flaps (Mompeán-González 2002). CF experiments could be conducted to find out whether flaps in words such as “ready” or “city” are classified as / d / or / t / or any other category, and what criteria (e.g. perceived phonetic similarity, orthographic conventions, lexical influences, etc.) subjects use to classify flaps as members of phoneme categories. A CF could also be devised to find out how speakers classify the short close front vowel in words such as “happy” or “easy”, the actual phonological classification of which is a somehow controversial issue (Wells 1982; Windsor-Lewis 1990). Another problem of classification concerns / ə /, which, according to some phonologists, is an allophone of several other vowels, according to other phonologists, it is an allophone of a phoneme that also includes the realisations of the traditional / ʌ / phoneme, etc. (see Clark & Yallop 1995: 143-144; Roach 1983: 100-101). CF experiments could shed light on the psychological reality of / ə / as an independent conceptual “phoneme” category or to validate or invalidate any of the claims about the category membership of schwas as psychological phonological claims.

Third, in view of the results obtained in experiment 4, the psychological reality of certain allophonic variants or of subphonemic phenomena could also be investigated. For example,



further allophonic variants could be investigated to see whether speakers are able to distinguish these varieties. For example, a CF could be conducted to find out whether speakers can learn to discriminate two of the most characteristic allophones of / l /, the so-called “clear” l and “dark” l. The production of / l / involves a complete closure in the middle of the mouth made by the tip of the tongue touching the centre of the teeth-ridge and with a lowering of one or both sides of the tongue to allow a lateral escape of the air-stream. However, when / l / appears in syllable initial coda pre-vocalic positions (e.g. [let]) in syllable-initial position, the front of the tongue is raised simultaneously while when / l / appears in word-final positions after vowels (e.g. [fi:t]), as a syllabic sound (e.g. ['bɒtɫ]), or before a consonant (e.g. [pɹɪz]), the back of the tongue is considerably raised toward the soft palate, being thus strongly velarised. The rising of the back of the tongue gives “dark” l the resonance of a back vowel. It seems that most allophones will not be pre-established categories for most speakers unless they have a specialised phonetic training background. However, the CF technique may prove useful to investigate to what extent allophonic variations might be perceptually discriminable for speakers and possible degrees of awareness and psychological significance of such variations.

Finally, other CF experiments could be conducted to investigate many interesting and traditionally controversial issues like the interpretation of the sequences [tɹ] (e.g. “tray”) and [dʒ] (e.g. “dread”) as unisegmental (one-phoneme) or bisegmental (two-phoneme) way.⁸ Such studies could be conducted in the spirit of Jaeger’s (1980a) experiment 4, which taught subjects the category “two-consonant clusters” and which found out that phonetically naive native speakers of English considered that / tʃ / and / dʒ / in words like “chip” or “gyp” were each “one sound” and not clusters of two consonants.

The CF technique could also be used to study the psychological reality of traditional generative underlying forms, the most ubiquitous premise of (classical) generative phonological argumentation, by examining speakers’ capacity to categorise words on the basis of generative phonological underlying forms like the common / rɪxt / posited for pairs like “right-righteous”. If, as McCawley suggests (1986: 29), the underlying form / rɪxt / has any psychological relevance at all or “it is a proposition about what (at least some) speakers of English have in their linguistic competence, it ought to have some bearing on those speakers’ categorisation of English words, e.g. the / i / of the underlying form ought to imply the availability to those speakers of classification dimensions on which “right” is classified with other words to which the given scheme of phonological analysis assigns underlying forms /CoɪCo/...”. McCawley continues to claim that “to the extent that other pairs resist neat analyses in terms of generative phonological underlying forms, the neatness of the analysis of right becomes suspect”.

⁸ At the theoretical level, most phonologists treat such sequences as clusters of two phonemes. Gimson is the most remarkable exception. He treats / tr / and / dr / as single units, but not on phonological grounds but simply “on purely phonetic and pedagogical terms” (1978: 173).

The technique could also be applied to different questions regarding suprasegmental phonological questions about lexical stress, rhythm, or intonation.

CF experiments could also be conducted to compare different issues across languages. For example, it is well known that despite the similarities of the vocal apparatus across members of the species -and the ability of people of any genetic background to acquire any language -phoneme inventories are different in different languages. Experimental studies also show that there are no two languages in which the implementation of analogous phonemes is exactly the same (see Pierrehumbert *et al.* 2000 for examples). Results such as these make it impossible to equate phonological inventories across languages. CF experiments could be designed to investigate these issues. For example, Jaeger (1980a: experiment 4) studied the category “words which begin with two consonants”, but did not include [ts] and [dz], which are phonological sequential constraints. Jaeger (1980a: experiment 4). An interesting example of such a path of investigation was provided by Jaeger (1980a: experiments 3), who studied Japanese speakers categorisation of voiceless dental obstruents in Japanese through a CF experiment. Jaeger found out that Japanese subjects tended to include as members of the Japanese phoneme category that can be symbolised, as Jaeger did, as / t /, not only [t] but also [ts] and [tʃ]. The inclusion of the voiceless palato-alveolar affricate as member of the / t / category is, however, unlikely for native English speakers of English.

This CF methodology might also be used to study a variety of second-language phonological learning issues. For example, the CF technique might be used to make subjects learn to discriminate between different phoneme categories whose phonetic realisations may either be non-existent in the learner’s native language or they may be grouped within a different category. For example, in Spanish there is no such binary phonological contrast between two high front vowel phonemes as there is in English with the vowels / i: / and / ɪ /. This contrast is a problem for many learners, notoriously for speakers of Spanish, who have a vowel of approximately the same quality as that of / i: / -although not the with the length peculiarities of the English vowel- but lack a vowel that is similar to / ɪ /. A CF could be designed to teach a number of Spanish learners of English that phonological contrast. This was precisely done recently (Mompeán-González 2002 in press). Other segmental differences (e.g. / ɪ / vs. / i: /, / u / vs. / u: /, / ɒ / vs. / ɔ: /, etc.) could be studied and/or taught in the same way. In addition, the CF technique could be used to teach suprasegmental differences. In the same study by Mompeán-González, a different group learned the suprasegmental contrast between a rising and a falling intonation as it relates to questions tags.⁹ The potential pedagogical value of such the

⁹ Questions tags have specific construction rules and two contrastive intonation patterns that can be considered as having both an attitudinal and a grammatical function (Roach 1983: 147). When question tags have a falling intonation, the speaker expects the other person to agree with what he/she has said. The speaker is comparatively certain that the information is correct, and simply expects the listener to provide confirmation (i.e. expects the answer to be *yes* or *no*, depending on the question). On the contrary, if



CF technique in second language teaching/learning, which has remained unnoticed so far, should be further explored in the future. Its use in the language laboratory may help teachers make their students attain categorical differences at a perceptual level.

In addition, CF experiments might be run with second-language learners of English who have previously learned a given phonological category in English to find out whether they categorise all allophones as members of the already-learned category or whether there are some allophonic variants that are not included in the category and should, therefore, be rehearsed in an instructional setting.

Finally, in the field of speech perception, Weitzman (1993) discussed different possible uses of the CF experimental paradigm, such as exploring which phonological categorisations are more difficult to learn than others (i.e. the degree of learnability of phonological categories based on certain acoustic differences, uncovering thus the contribution of various acoustic differences in the perception of certain kinds of phonological categories. In other words, the CF technique can help us determine the relative perceptual saliency of different acoustic properties (e.g. Weitzman 1992). Other possible uses of the CF experimental paradigm include: 1) examining how differences in learning experience can affect new learning, such as in second language learning, 2) studying how repertoires of categorical distinctions are built up, 3) determining whether physiologically conditioned acoustic effects are perceptually salient, and 4) exploring the perceptual mechanisms involved in historical sound changes, such as the development of tone distinctions in many Asian languages arising from the loss of consonantal distinctions.

Also, the CF paradigm could be used to teach dialectal differences. Just as dialectal differences are discriminated by subjects in experimental settings (e.g. Willis 1971).

In sum, CF experiments may be run to investigate a very wide range of topics and to answer some of the long-standing questions persisting in different fields of research. This technique seems a very promising one in order to obtain valuable experimental evidence on the questions under investigation, researchers interested in field of research related to phonology might benefit from its use.

4.5. Conclusion.

The present dissertation hopes to have emphasised the importance (as well as complexity) of categorisation as a human cognitive ability as revealed by the enormous amount of experimental research conducted with different subject populations and the extensive amount of

intonation rises throughout the question tag, it is because it is a real question and the speaker does not know what the listener's answer will be. The question tag is an actual request for information.

topics involved in such studies. This work also hopes to have shown that the categorisation of sounds is an integral part of categorisation research and necessary for a full understanding of that cognitive ability. However, categorisation behaviour is also a very useful source of evidence for the phonologist interested in the psychological aspect of phonological patterns and entities and the pertinence of reflecting their psychological reality is phonological analysis. The thesis has also tried to show how use of experiments and experimental methods may be used to pursue phonological investigations with psychological claims and, more specifically, it hopes to have shown how the “concept formation” experimental technique can be used to obtain experimental evidence on, in this case, four different phonological issues.

REFERENCES

The following abbreviations for journals, publishers, etc. will be used throughout the reference section.¹

AE	American Ethnologist
ALB	Animal Learning and Behavior
ARP	Annual Review of Psychology
BBS	Behavioral and Brain Sciences
BJDP	British Journal of Developmental Psychology
BJP	British Journal of Psychology
BLS	Berkeley Linguistic Society
BPS	Bulletin of the Psychonomic Society
BRMI	Behavioral Research Methods and Instrumentation
CC	Communication and Cognition
CD	Child Development
CG	Cognition
CGD	Cognitive Development
CL	Cognitive Linguistics
CLS	Chicago Linguistic Society
CP	Cognitive Psychology
CPC	Cahiers de Psychologie Cognitive
CS	Cognitive Science
DP	Developmental Psychology
EJSP	European Journal of Social Psychology
IBD	Infant Behavior and Development.
JASA	Journal of the Acoustical Society of America
JCL	Journal of Child Language
JCP	Journal of Comparative Physiology
JCPP	Journal of Comparative and Physiological Psychology
JEAB	Journal of the Experimental Analysis of Behavior
JECP	Journal of Experimental Child Psychology
JEP:ABP	Journal of Experimental Psychology: Animal Behavior Processes
JEP:HLM	Journal of Experimental Psychology: Human Learning and Memory
JEP:HPP	Journal of Experimental Psychology: Human Perception and Performance
JEP:LMC	Journal of Experimental Psychology: Learning, Memory, and Cognition
JESP	Journal of Experimental Social Psychology
JL	Journal of Linguistics
JMB	Journal of Mind and Behavior
JML	Journal of Memory and Language
JP	Journal of Phonetics

¹ The names of journals for which fewer than three different references have been cited throughout the dissertation are not abbreviated in this section.

JPR	Journal of Psycholinguistic Research
JPSP	Journal of Personality and Social Psychology
JVLVB	Journal of Verbal Learning and Verbal Behavior
LEA	Lawrence Erlbaum Associates
Lg	Language
LM	Learning and Motivation
LS	Language and Speech
MC	Memory and Cognition
MPQBD	Merrill-Palmer Quarterly of Behavior and Development
PB	Psychological Bulletin
PACSS	Proceedings of the Annual Conference of the Cognitive Science Society.
PP	Perception and Psychophysics
PR	Psychological Review
PS	Psychonomic Science
PSS	Psychological Science
PHN	Phonetica
PHY	Phonology Yearbook
QJEP	Quarterly Journal of Experimental Psychology
QJEP:HEP	Quarterly Journal of Experimental Psychology. Section A: Human Experimental Psychology
QJEP:CPP	Quarterly Journal of Experimental Psychology. Section B: Comparative and Physiological Psychology
SC	Social Cognition

Abelson, R. P. 1981. "Psychological Status of the Script Concept". *American Psychologist*, 36:715-729.

Abercrombie, D. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Adelson, B. 1985. "Comparing Natural and Abstract Categories: A Case Study from Computer Science". *CS*, 9:417-430.

Ades, A. E. 1977. "Vowels, Consonants, Speech and Nonspeech". *PR*, 84:524-530.

Ahn, W. K., & Medin, D. L. 1992. "A Two-Stage Model of Category Construction". *CS*, 16:81-121.

Aiken, L. S., & Williams, T. M. 1973. "A Developmental Study of Schematic Concept Formation". *DP*, 8:162-167.

Aitken, M. R. F., Bennett, C. H., McLaren, I. P. L., & Mackintosh, N. N. 1996. "Perceptual Differentiation During Categorisation Learning by Pigeons". *JEP:ABP*, 22:43-50

Akamatsu, T. 1988. *The Theory of Neutralization and the Archiphoneme in Functional Phonology*. Amsterdam/Philadelphia: John Benjamins.

Alexander, T. M., & Enns, J. T. 1988. "Age Changes in the Boundaries of Fuzzy Categories". *CD*, 59:1372-1386.

Allen, V. L., & Wilder, D. A. 1975. "Categorisation, Belief Similarity, and Intergroup Discrimination". *JPSP*, 32:971-977.

- Ambruster, T. E. 1978. *The Psychological Reality of the Vowel Shift and Laxing Rules*. Ph.D. dissertation. University of California, Irvine.
- Andersen, S. M., & Klatzky, R. L. 1987. "Traits and Social Stereotypes: Levels of Categorisation in Person Perception". *JPSP*, 53:235-246.
- Anderson, J. R. 1991. "The Adaptive Nature of Human Categorisation". *PR*, 98:409-429.
- Anderson, J. R., & Finchman, J. M. 1996. "Categorisation and Sensitivity to Correlation". *JEP:LMC*, 22:259-277.
- Anderson, J. R., Kline, P. J., & Beasley, C. M. Jr. 1979. "A General Learning Theory and its Application to Schema Abstraction". In Bower, G. H. (ed.). 1979. *The Psychology of Learning and Motivation*. Vol. 13. New York: Academic Press, 227-318.
- Anderson, S. R. 1985. *Phonology in the Twentieth Century: Theories of Rules and Theories of Representation*. Chicago: University of Chicago Press.
- Anglin, J. M. 1977. *Word, Object, and Conceptual Development*. New York: W. W. Norton & Company.
- Anglin, J. M. 1985. "The Child's Expressible Knowledge of Word Concepts: What Preschoolers Can Say About the Meaning of Some Nouns and Verbs". In Nelson, K. E. (ed.). 1985. *Children's Language*. Vol. 5. Hillsdale, NJ: LEA, 77-127.
- Armstrong, S. L., Gleitman, L. R., & Gleitman, H. 1983. "What Some Concepts Might Not Be". *CG*, 13:263-308.
- Arnold, G. F. 1967. "Concerning the Theory of Plosives". *Le Maître Phonétique*, no 125.
- Ashby, F. G., & Maddox, T. W. 1993. "Relations between Prototype, Exemplar, and Decision Bound Models of Categorisation". *Journal of Mathematical Psychology*, 37:372-400.
- Ashcraft, M. H. 1978. "Property Norms for Typical and Atypical Items from 17 Categories: A Description and Discussion". *MC*, 6:227-232.
- Astley, S. L., & Wasserman, E. A. 1992. "Categorical Discrimination and Generalization in Pigeons: All Negative Stimuli Are Not Created Equal". *JEP:ABP*, 18:193-207.
- Aydin, A., & Pearce, J. M. 1994. "Prototype Effects in Categorisation by Pigeons". *JEP:ABP*, 20:264-277.
- Bailey, P. J., & Summerfield, Q. 1980. "Information in Speech: Observation on the Perception of [s]-Stop Clusters". *JEP:HPP*, 6:536-563.
- Baker, W. J., Prideaux, G. D., & Derwing, B. L. 1973. "Grammatical Properties of Sentences as a Basis for Concept Formation". *JPR*, 2:201-220.
- Barclay, J. R. 1972. "Non-categorical Perception of a Voiced Stop Consonant: A Replication". *PP*, 11:269-273.
- Barr, R. A., & Caplan, L. J. 1987. "Category Representations and their Implications for Category Structure". *MC*, 15:397-418.



- Barsalou, L. W. 1981. *Determinants of Graded Structure in Categories*. PhD. Diss. Stanford University, Stanford, California. 8201961.
- Barsalou, L. W. 1982. "Context-Independent and Context-Dependent Information in Concepts". *MC*, 10:82-93.
- Barsalou, L. W. 1983. "Ad Hoc Categories". *MC*, 11:211-227.
- Barsalou, L. W. 1985. "Ideals, Central Tendency, and Frequency of Instantiation as Determinants of Graded Structure in Categories". *JEP:LMC*, 11:629-654.
- Barsalou, L. W. 1987. "The Instability of Graded Structure: Implications for the Nature of Concepts". In Neisser, U. (ed.). 1987. *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorisation*. New York: CUP, 101-140.
- Barsalou, L. W. 1989. "Intraconcept Similarity and Its Implications for Interconcept Similarity". In Vosniadou, S., & Ortony, A. (eds.). 1989. *Similarity and Analogical Reasoning*. New York: CUP, 76-121.
- Barsalou, L. W. 1990a. "Access and Inference in Categorisation". *BPS*, 28:268-271.
- Barsalou, L. W. 1990b. "On the Indistinguishability of Exemplar Memory and Abstraction in Category Representation". In Srull, T. K., & Wyer, R. S. Jr. (eds.). 1990. *Advances in Social Cognition. Vol. 3.: Content and Process Specificity in the Effects of Prior Experiences*. Hillsdale, NJ: LEA, 61-88.
- Barsalou, L. W. 1992a. *Cognitive Psychology: An Overview for Cognitive Scientists*. Hillsdale, NJ: LEA.
- Barsalou, L. W. 1992b. "Frames, Concepts, and Conceptual Fields". In Kittay, E., & Lehrer, A. (eds.). 1992. *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization*. Hillsdale, NJ: LEA, 21-74.
- Barsalou, L. W. 1993. "Challenging Assumptions About Concepts". *CGD*, 8:169-180.
- Barsalou, L. W., & Hale, C. R. 1993. "Components of Conceptual Representation: From Feature Lists to Recursive Frames". In Van Mechelen, I., Hampton, J. A., Michalski, R. S., & Theuns, P. (eds.). 1993. *Categories and Concepts: Theoretical Views and Inductive Data Analysis*. San Diego, CA: Academic Press, 97-144.
- Barsalou, L. W., & Medin, D. L. 1986. "Concepts: Static Definitions or Context-Dependent Representations?". *CPC*, 6:187-202.
- Barsalou, L. W., & Sewell, D. R. 1985. "Contrasting the Representation of Scripts and Categories". *JML*, 24:646-665.
- Barsalou, L. W., Sewell, D. R., & Ballato, S. M. 1986. "The Instability of Category Representations: I. As Measured by Graded Structure". Unpublished manuscript.
- Barsalou, L. W., Spindler, J. L., Sewell, D. R., Ballato, S. M., & Gendel, E. M. 1987. "The Instability of Category Representations: II. As Measured by Property Generation". Unpublished Manuscript.
- Battig, W. F., & Montague, W. E. 1969. "Category Norms for Verbal Items in 56 Categories: A Replication and Extension of the Connecticut Category Norms". *JEP*, 80:1-45.

- Baudouin de Courtenay, J. 1895. "An Attempt at a Theory of Phonetic Alternations". In Stankiewicz, E. (ed.). 1972. *A Baudouin de Courtenay Anthology*. Bloomington: Indiana University Press, 144-212.
- Bauer, P. J., Dow, G. A., & Hertsgaard, L. A. 1995. "Effects of Prototypicality on Categorisation in 1- to 2-Year-Olds: Getting Down to Basics". *CGD*, 10:43-68.
- Bauer, P. J., & Mandler, J. M. 1989. "Taxonomies and Triads: Conceptual Organization in One- to Two-Year-Olds". *CP*, 21:156-184.
- Behl-Chadha, G. 1994. *Perceptually-Driven Superordinate-Like Categorical Representations in Early Infancy*. Ph.D. Diss. Brown University.
- Behl-Chadha, G. 1996. "Basic-Level and Superordinate-Like Categorical Representations in Early Infancy". *CG*, 60:105-141.
- Behl-Chadha, G., & Eimas, P. D. 1995. "Infant Categorisation of Left-Right Spatial Relations". *BJDP*, 13:69-79.
- Beletsky, L. D., Chao, S., & Smith, D. G. 1980. "An Investigation of Sound-Based Species Recognition in the Red-Winged Blackbird (*Agelaius Phoeniceus*)". *Behavior*, 73:189-203.
- Bellezza, F. S. 1984. "Reliability of Retrieval from Semantic Memory: Common Categories". *BPS*, 22:324-326.
- Berlin, B. 1978. "Ethnobiological Classification". In Rosch, E. H., & Lloyd, B. B. (eds.). 1978. *Cognition and Categorisation*. Hillsdale, NJ: LEA, 9-26.
- Berlin, B. 1992. *Ethnobiological Classification: Principles of Categorisation of Plants and Animals in Traditional Societies*. Princetown, NJ: Princetown University Press.
- Berlin, B., Breedlove, D. E., & Raven, P. H. 1973. "General Principles of Classification and Nomenclature in Folk Biology". *American Anthropologist*, 75:214-242.
- Bertinetto, P. M. 1992. "The Use and Misuse of External Evidence in Phonology". In Dressler, W. U., Luschützky, H. C., Pfeiffer, O. E., & Rennison, J. R. (eds.). 1992. *Phonologica 1988*. Cambridge: Cambridge University Press, 33-47.
- Bhatt, R. S. 1988. *Categorisation in Pigeons: Effects of Category Size, Congruity with Human Categories, Selective Attention, and Secondary Generalization*. Doctoral Dissertation. University of Iowa, Iowa City.
- Bhatt, R. S., Wasserman, E. A., Reynolds, W. F., & Knauss, K. S. 1988. "Conceptual Behavior in Pigeons: Categorisation of Both Familiar and Novel Examples from Four Classes of Natural and Artificial Stimuli". *JEP:ABP*, 14:219-234.
- Biederman, I., Subramaniam, S., Bar, M., Kalocsai, P., & Fiser, J. 1999. "Subordinate-Level Object Classification Reexamined". *Psychological Research*, 62:131-153.
- Billig, M., & Tajfel, H. 1973. "Social Categorisation and Similarity in Intergroup Behavior". *EJSP*, 3:27-52.
- Bjorklund, D. F., Smith, S. C., & Ornstein, P. A. 1982. "Young Children's Release from Proactive Interference: The Effects of Category Typicality". *BPS*, 20:211-213.



- Bjorklund, D. F., Thompson, B. E., & Ornstein, P. A. 1983. "Developmental Trends in Children's Typicality Judgments". *BRMI*, 15:350-356.
- Black, M., & Shulamuth, C. 1981. "Psycholinguistics Without 'Psychological Reality' ". *Linguistics*, 19:37-61.
- Blewitt, P. 1983. "Dog vs. Collie: Vocabulary in Speech to Young Children". *DP*, 19:602-609.
- Blewitt, P. 1994. "Understanding Categorical Hierarchies: The Earliest Levels of Skill". *CD*, 65:1279-1298.
- Blewitt, P., & Durkin, M. 1982. "Age, Typicality, and Task Effects on Categorisation of Objects". *Perception and Motor Skills*, 55:435-445.
- Bloch, B. 1941. "Phonemic Overlapping". *American Speech*, 16:278-284.
- Bloch, B., & Trager, G. L. 1942. *Outline of Linguistic Analysis*. Baltimore: Linguistic Society of America at the Waverly Press.
- Bloomfield, L. 1979 [1933]. *Language*. New York: Holt, Rinehart, & Winston.
- Bolton, N. 1977. *Concept Formation*. Oxford: Pergamon Press.
- Bomba, P. C. 1984. "The Development of Orientation Categories Between 2 and 4 Months of Age". *JECp*, 37:609-636.
- Bomba, P. C., & Siqueland, E. R. 1983. "The Nature and Structure of Infant Form Categories". *JECp*, 35:294-328.
- Bond, Z. S., & Garnes, S. 1980. "Misperceptions of Fluent Speech". In Cole, R. A. (ed.). 1980. *Perception and Production of Fluent Speech*. Hillsdale, NJ: LEA, 115-132.
- Borkenau, P. 1990. "Traits as Ideal-Based and Goal-Derived Social Categories". *JPSP*, 58:381-396.
- Bornstein, M. H. 1984. "A Descriptive Taxonomy and Psychological Categories Used by Infants". In Sophian, C. (ed.). 1984. *Origins of Cognitive Skills*. Hillsdale, NJ: LEA, 313-338.
- Bornstein, M. H. 1987. "Perceptual Categories in Vision and Audition". In Harnad, S. (ed.). 1987. *Categorical Perception: The Groundwork of Cognition*. New York: CUP, 287-300.
- Bornstein, M. H., Kessen, W., & Weiskopf, S. 1976. "Color Vision and Hue Categorisation in Young Human Infants". *JEP:HPP*, 2:115-129.
- Boster, J. S. 1988. "Natural Sources of Internal Category Structure: Typicality, Familiarity, and Similarity of Birds". *MC*, 16:258-270.
- Bourne, L. E. 1957. "Effects of Delay of Information Feedback and Task Complexity on the Identification of Concepts". *JEP*, 54:201-207.
- Bourne, L. E. 1966. *Human Conceptual Behavior*. Boston, NJ: Allyn & Bacon.
- Bourne, L. E. 1970. "Knowing and Using Concepts". *PR*, 77:546-556.

- Bourne, L. E. 1982. "Typicality Effects in Logically Defined Categories". *MC*, 10:3-9.
- Bourne, L. E., & Bunderson, C. V. 1963. "Effects of Delay of Informative Feedback and Length of Postfeedback Interval on Concept Identification". *JEP*, 65:1-7.
- Bourne, L. E., Guy, D. E., Dodd, D., & Justesen, D. R. 1965. "Concept Identification: The Effects of Varying Length and Informational Components of the Intertrial Interval". *JEP*, 69:624-629.
- Bower, G. H., Black, J. B., & Turner, T. J. 1979. "Scripts in Memory for Text". *CP*, 11:177-220.
- Bowerman, M. 1978. "The Acquisition of Word Meaning: An Investigation of Some Current Concepts". In Waterson, N., & Snow, C. (eds.). 1978. *The Development of Communications: Social and Pragmatic Factors in Language Acquisition*. New York: John Wiley & Sons, 263-287.
- Bowerman, M. 1980. "The Structure and Origin of Semantic Categories in the Language-Learning Child". In Foster, M. L., & Brandes, S. H. (eds.). 1980. *Symbol as Sense: New Approaches to the Analysis of Meaning*. New York: Academic Press, 277-299.
- Bowey, J. A., & Francis, J. 1991. "Phonological Analysis as a Function of Age and Exposure to Reading Instruction". *Applied Psycholinguistics*, 12: 91-121.
- Bradley, L., & Bryant, P. E. 1983. "Categorizing Sounds and Learning to Read -A Causal Connection". *Nature*, 301:419-421.
- Brewer, M. B., Dull, V., & Lui, L. 1981. "Perceptions of the Elderly: Stereotypes As Prototypes". *JPSP*, 41:656-670.
- Brooks, L. R. 1978. "Nonanalytic Concept Formation and Memory for Instances". In Rosch, E., & Lloyd, B. B. (eds.). 1978. *Cognition and Categorisation*. Hillsdale, NJ: LEA, 169-211.
- Brooks, L. R. 1987. "Decentralized Control of Categorisation: The Role of Prior Processing Episodes". In Neisser, U., (ed.). 1987. *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorisation*. New York: CUP, 141-174.
- Brooks, L. R., Norman, G. R., & Allen, S. W. 1991. "The Role of Specific Similarity in a Medical Diagnostic Task". *JEP:GE*, 126:279-287.
- Brown, C. H., Kolar, J., Torrey, B. J., Truong-Quang, T., & Volkman, P. 1976. "Some General Principles of Biological and Non-Biological Folk Classification". *AE*, 3:73-85.
- Brown, G. B. 1977. *Listening to Spoken English*. London: Longman.
- Brown, R. W. 1958. "How Shall a Thing Be Called?". *PR*, 65:14-21.
- Brown, R. W. 1976. "Reference: In Memorial Tribute to Eric Lenneberg". *CG*, 4:125-153.
- Brownell, H. H., & Caramazza, A. 1978. "Categorizing with Overlapping Categories". *MC*, 6:481-490.
- Bruck, M., & Treiman, R. 1990. "Phonological Awareness and Spelling in Normal Children and Dislexics: The Case of Initial Consonant Clusters". *JECP*, 50:156-178.



- Bruner, J. S. 1957. "On Perceptual Readiness". *PR*, 64:123-152. CHECK THE PAGES.
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. 1956. *A Study of Thinking*. New York: John Wiley & Sons.
- Bruner, J. S., Olver, R. R., & Greenfield, P. M. 1966. *Studies in Cognitive Growth*. New York: John Wiley & Sons.
- Burdick, C. K., & Miller, J. D. 1975. "Speech Perception by the Chinchilla: Discrimination of Sustained / a / and / i /". *JASA*, 58:415-427.
- Burton, M. W., Baum, S. R., & Blumstein, S. E. 1989. "Lexical Effects on the Phonetic Categorisation of Speech: The Role of Acoustic Structure". *JEP:HPP*, 15:567-575.
- Busemeyer, J. R., Dewey, G. I., & Medin, D. L. 1984. "Evaluation of Exemplar-Based Generalization and the Abstraction of Categorical Information". *JEP:LMC*, 10:638-648.
- Busemeyer, J. P., McDaniel, M. A., & Byun, E. 1993. "The Abstraction of Intervening Concepts from Experience with Multiple Input-Multiple Output Causal Environments". *CP*, 32:1-48.
- Buss, D. M., & Craik, K. H. 1983. "The Act Frequency Approach to Personality". *PR*, 90:105-126.
- Bybee, J. L. 1994. "A View of Phonology from a Cognitive and Functional Perspective". *CL*, 5:285-305.
- Bybee, J. L. 1999. "Usage-Based Phonology". In Darnell, M., Moravcsik, E., Newmeyer, F., Noonan, M., & Wheatley, K. (eds.). 1999. *Functionalism and Formalism in Linguistics. Volume I: General Papers*, 211-242.
- Bybee, J. L. & Moder, C. L. 1983. "Morphological Classes as Natural Categories". *Lg*, 59:251-270.
- Caisse, M. 1981. "Pitch Perturbations Induced by Voiceless Unaspirated Stops". Paper presented at the meeting of the Acoustical Society of America, Miami, December.
- Calfee, R., Chapman, R., & Venezky, R. 1972. "How a Child Needs to Think to Learn to Read". In Gregg, L. W. (ed.). 1972. *Cognition in Learning and Memory*. New York: John Wiley & Sons, 139-182.
- Callanan, M. A. 1983. *Parental Input and Young Children's Acquisition of Hierarchically Organized Concepts*. Ph.D. Diss. Stanford University, Palo Alto, CA.
- Callanan, M. A. 1985. "How Parents Label Objects for Young Children: The Role of Input in the Acquisition of Category Hierarchies". *CD*, 56:508-523.
- Callanan, M. A., & Markman, E. M. 1982. "Principles of Organization in Young Children's Natural Language Hierarchies". *CD*, 53:1093-1101.
- Campbell, L. 1979. "The Quest for Psychological Reality: External Evidence in Phonology". In Fischer-Jørgensen, E., Rischer, J., & Thorsen, N. (eds.). 1979. *Proceedings of the Ninth Congress of Phonetic Sciences*. Vol. 2. Copenhagen: University of Copenhagen, 74-78

- Campbell, L. 1986. "Testing Phonology in the Field". In Ohala, J. J., & Jaeger, J. J. (eds.). 1986. *Experimental Phonology*. Orlando: Academic Press, 163-173.
- Cantor, N., & Mischel, W. 1977. "Traits as Prototypes: Effects on Recognition Memory". *JPSP*, 35:38-48.
- Cantor, N., & Mischel, W. 1979a. "Prototypes in Person Perception". In Berkowitz, L. (ed.). 1979. *Advances in Experimental Social Psychology*. Vol. 12. New York: Academic Press, 3-52.
- Cantor, N., & Mischel, W. 1979b. "Prototypicality and Personality: Effects on Free Recall and Personality Impressions". *Journal of Research in Personality*, 13:187-205.
- Cantor, N., Mischel, W., & Schwartz, J. C. 1982. "A Prototype Analysis of Psychological Situations". *CP*, 14:45-77.
- Cantor, N., Smith, E. E., French, R. D., & Mezzich, J. 1980. "Psychiatric Diagnosis as Prototype Categorisation". *Journal of Abnormal Psychology*, 89:181-193.
- Capranica, R. R. 1966. "Vocal Response of the Bullfrog to Natural and Synthetic Mating Calls". *JASA*, 40:1131-1139.
- Carney, A. E., & Widin, G. P. 1976. "Acoustic Discrimination Within Phonetic Categories". *JASA*, 59. S24(A).
- Carney, A. E., Widin, G. P., & Viemeister, N. F. 1977. "Noncategorical Perception of Stop Consonants Differing in VOT". *JASA*, 62:961-970.
- Carroll, P., & Slowiczek, M. L. 1986. "Constraints on Semantic Priming in Reading: A Fixation Time Analysis". *MC*, 14:509-522.
- Carson, M. T., & Abrahamson, A. 1976. "Some Members Are More Equal than Others: The Effect of Semantic Typicality on Class-Inclusion Performance". *CD*, 47:1186-1190.
- Cena, R. M. 1978. *When is a Phonological Generalization Psychologically Real?*. Bloomington, Indiana: Indiana University Linguistics Club.
- Cerella, J. 1979. "Visual Classes and Natural Categories in the Pigeon". *JEP:HPP*, 5:68-77.
- Chaplin, W. F., John, O. P., & Goldberg, L. R. 1988. "Conceptions of States and Traits: Dimensional Attributes with Ideals as Prototypes". *JPSP*, 54:541-557.
- Cherniak, 1984. "Prototypicality and Deductive Reasoning". *JVLVB*, 23:625-642.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. 1981. "Categorisation and Representation of Physics Problems by Experts and Novices". *CS*, 5:121-152.
- Chomsky, N. 1964. *Current Issues in Linguistic Theory*. The Hague: Mouton.
- Chomsky, N. 1980. "Rules and Representations". *BBS*, 3:1-15.
- Chomsky, N., & Halle, M. 1968. *The Sound Pattern of English*. New York: Harper & Row.
- Christophersen, P. 1956. *An English Phonetics Course*. London: Longman

- Chumbley, J. I. 1986. "The Roles of Typicality, Instance Dominance, and Category Dominance in Verifying Category Membership". *JEP:LMC*, 12:257-267.
- Clark, E. V. 1973. "What's in a Word?. On the Child's Acquisition of Semantics in His First Language". In Moore, T. E. (ed.). 1973. *Cognitive Development and the Acquisition of Language*. New York: Academic Press, 65-110.
- Clark, J. and Yallop, C. 1995. [1990]. *An Introduction to Phonetics and Phonology*. (2nd edition). Oxford, UK: Blackwell.
- Cohen, A. 1952. *The Phonemes of English*. The Hague.
- Cohen, B., & Murphy, G. L. 1984. "Models of Concepts". *CS*, 8:27-58.
- Cohen, C. E. 1981. "Person Categories and Social Perception: Testing some Boundaries of the Processing Effects of Prior Knowledge". *JPSP*, 40:441-452.
- Cohen, L. B., & Strauss, M. S. 1979. "Concept Acquisition in the Human Infant". *CD*, 50:419-424.
- Cohen, L. B., & Younger, B. A. 1983. "Perceptual Categorisation in the Infant". In Scholnick, E. K. (ed.). 1983. *New Trends in Conceptual Representation: Challenges to Piaget's Theory?*. Hillsdale, NJ: LEA, 197-220.
- Coleman, L., & Kay, P. 1981. "Prototype Semantics: the English Word *Lie*". *Lg*, 57:26-44.
- Collins, A. M., & Loftus, E. F. 1975. "A Spreading-Activation Theory of Semantic Processing". *PR*, 82:407-428.
- Colombo, J., McCollam, K., Coldren, J. T., Mitchell, D. W., & Rash, S. J. 1990. "Form Categorisation in 10-Month-Olds". *JECp*, 49:173-188.
- Colombo, J., O'Brien, M., Mitchell, D. W., Roberts, K., & Horowitz, F. D. 1987. "A Lower Boundary for Category Formation in Preverbal Infants". *JCL*, 14:383-385.
- Connine, C. M., & Clifton, C. 1987. "Interactive Use of Lexical Information in Speech Perception". *JEP:HPP*, 13:291-299.
- Connine, C. M., & Titone, D. 1996. "Phoneme Monitoring". *Language and Cognitive Processes*, 11:635-645.
- Connine, C. M., Titone, D., & Wang, J. 1993. "Auditory Word Recognition: Extrinsic and Intrinsic Effects of Word Frequency". *JEP:LMC*, 19:81-94.
- Cook, R. G., Cavoto, K. K., & Cavoto, B. R. 1995. "Same-Different Texture Discrimination and Concept Learning in Pigeons". *JEP:ABP*, 21:253-260.
- Cook, R. G., Katz, J. S., & Cavoto, B. R. 1997. "Pigeon Same-Difference Concept Learning With Multiple Stimulus Classes". *JEP:ABP*, 23:417-433.
- Cook, R. G., Wright, A. A., & Kendrick, D. F. 1990. "Visual Categorisation by Pigeons". In Commons, M. L., Herrnstein, R. J., Kosslyn, S. M., Mumford, D. B. (eds.). 1990. *Quantitative Analyses of Behavior: Behavioral Approaches to Pattern Recognition and Concept Formation*. Vol. 8. Hillsdale, NJ. LEA, 187-214.

- Corrigan, R. 1986. "The Internal Structure of English Transitive Sentences". *MC*, 14:420-431.
- Corrigan, R., 1989. "Introduction: Linguistic and Non-Linguistic Categorisation: Structure and Process". In Corrigan, R., Eckman, F., & Noonan, M. (eds.). 1989. *Linguistic Categorisation*. Amsterdam/Philadelphia: John Benjamins, 1-28.
- Corrigan, R. 1991. "Sentences as Categories: Is There a Basic-Level Sentence?". *CL*, 2:339-356.
- Corter, J. E., & Gluck, M. A. 1992. "Explaining Basic Categories: Feature Predictability and Information". *PB*, 111:291-303.
- Cruse, D. A. 1977. "The Pragmatics of Lexical Specificity". *JL*, 13:153-164.
- Cutler, A. 1979. "The Psychological Reality of Word Formation and Lexical Stress Rules". In Fischer-Jørgensen, E., Rischer, J., & Thorsen, N. (eds.). 1979. *Proceedings of the Ninth Congress of Phonetic Sciences*. Vol. 2. Copenhagen: University of Copenhagen, 79-85.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. 1987. "Phoneme Identification and the Lexicon". *CP*, 19:141-177.
- Cutler, A. Norris, D., & van Ooyen, B. 1990. "Vowels as Phoneme Detection Targets". *Proceedings of the International Conference of Spoken Language Processing*. Kobe, Japan. Vol. 1, 581-584.
- Cutler, A., van Ooijen, B., Norris, D., & Sánchez-Casas, R. 1996. "Speeded Detection of Vowels: A Cross-Linguistic Study". *PP*, 58:807-822.
- Cutting, J. E. 1982. "Plucks and Bows Are Categorically Perceived, Sometimes". *PP*, 31:462-476.
- Cutting, J. E., & Rosner, B. S. 1974. "Categories and Boundaries in Speech and Music". *PP*, 16:564-570.
- Daehler, M. W., Lonardo, R., & Bukatko, D. 1979. "Matching and Equivalence Judgments in Very Young Children". *CD*, 50:170-179.
- Dahlgren, K. 1985. "The Cognitive Structure of Social Categories". *CS*, 9:379-398.
- D'Amato, M. R., & Van Sant, P. 1988. "The Person Concept in Monkeys (*Cebus Apella*)". *JEP:ABP*, 14:43-55.
- Davidsen-Nielsen, N. 1969. "English Stops After Initial / s /". *English Studies*, 4:321-339.
- Davidsen-Nielsen, N. 1974. "Syllabification in English Words with Medial *sp*, *st*, *sk*". *JP*, 2:15-45.
- Davidsen-Nielsen, N. 1975. "A Phonological Analysis of English *sp*, *st*, *sk*, with Special Reference to Speech Error Evidence". *Journal of the International Phonetic Association*, 5:3-25.
- Davidsen-Nielsen, Niels. 1978. *Neutralization and Archiphoneme: Two phonological Concepts and their History*. Copenhagen: Akademisk Forlag.



- Davis, K. 1995. "Phonetic and Phonological Contrasts in the Acquisition of Voicing: Voice Onset Time Production in Hindi and English". *JCL*, 22:275-305.
- Davis, K., & Kuhl, P. K. 1992. "Best Exemplars of English Velar Stops: A First Report". In Ohala, J. J., Nearey, T. M., Derwing, B. L., Hodge, M. M., & Weibe, G. E. (eds.). 1992. *Proceedings of the International Conference on Spoken Language Processing*. Alberta, Canada: University of Alberta, 495-498.
- Davis, K., & Kuhl, P. K. 1994. "Test of the Perceptual Magnet Effect for American English /k/ and /g/". *JASA*, 95. 2976.
- Déak, G. O., & Bauer, P. J. 1995. "The Effects of Task Comprehension on Preschoolers' and Adults' Categorisation Choices". *JECP*, 60:393-427.
- Déak, G. O., & Bauer, P. J. 1996. "The Dynamics of Preschoolers' Categorisation Choices". *CD*, 67:740-767.
- Dell, G. S., & Newman, J. E. 1980. "Detecting Phonemes in Fluent Speech". *JVLVB*, 19:608-623.
- Dennis, I., Hampton, J. A., & Lea, S. E. G. 1973. "New Problem in Concept Formation". *Nature*, 243:101-102.
- Derwing, B. L. 1979. "Psycholinguistic Evidence and Linguistic Theory". In Prideaux, G. D. (ed.). 1979. *Perspectives in Experimental Linguistics*. Amsterdam: John Benjamins, 113-138.
- Derwing, B. L. 1980. "Against Autonomous Linguistics". In Perry, T. A. (ed.). 1980. *Evidence and Argumentation in Linguistics*. Berlin: Walter de Gruyter, 163-189.
- Derwing, B. L., & Baker, W. J. 1977. "The Psychological Basis for Morphological rules". In Macnamara, J. (ed.). 1977. *Language Teaching and Thought*. New York: Academic Press, 85-110.
- Derwing, B. L., & Nearey, T. M. 1986. "Experimental Phonology at the University of Alberta". In Ohala, J. J., & Jaeger, J. J. (eds.). 1986. *Experimental Phonology*. Orlando: Academic Press, 187-209.
- Derwing, B. L., & Wang, H. S. 1995. "Concept Formation as a Tool for the Investigation of Phonological Units in Taiwanese". *Proceedings of the XIIIth International Congress of Phonetic Sciences*. Vol. 3. Stockholm: KTH and Stockholm University, 362-365.
- Diehl, R. L., & Kluender, K. R. 1987. "On the Categorisation of Speech Sounds". In Harnad, S. (ed.). 1987. *Categorical Perception: The Groundwork of Cognition*. New York: CUP, 226-253.
- Diehl, R. L., Kluender, K. R., Foss, D. J., Parker, E. M., & Gernsbacher, M. A. 1987. "Vowels as Islands of Reliability". *JML*, 26:564-573.
- Dirven, R., & Taylor, J. 1988. "The Conceptualisation of Vertical Space in English: the Case of *Tall*". In Rudzka-Ostyn, B. (ed.). 1988. *Topics in Cognitive Linguistics*. Amsterdam: John Benjamins, 379-402.
- Dominowski, R. L. 1970. "Concept Attainment". In Marx, M. H. (ed.). 1970. *Learning: Interactions*. New York: Macmillan, 152-191.

- Donegan, P. J., & Stampe, D. 1979. "The Study of Natural Phonology". In Dinnsen, D. A. (ed.). 1979. *Current Approaches to Phonological Theory*. Bloomington: Indiana University Press, 126-173.
- Dorman, M. F., Raphael, L. J., & Isenberg, D. 1980. "Acoustic Cues for a Fricative-Affricate Contrast in Word-Final Position". *JP*, 8:397-405
- Dougherty, J. W. D. 1978. "Salience and Relativity in Classification". *AE*, 5:66-80.
- Downing, P. 1977. "On 'Basic Levels' and the Categorisation of Objects in English Discourse". *BLS*, 3:475-87.
- Duncan, E. M., & Kellas, G. 1978. "Developmental Changes in the Internal Structure of Semantic Categories". *JECp*, 26:328-340.
- Eddington, D. 1996. "The Psychological Status of Phonological Analyses". *Linguistica*, 36:17-37.
- Eddington, D. 1999. "Role of Experiments in Phonological Investigation". *Journal of Quantitative Linguistics* 6, 14-28.
- Eddington, D. 2001. "Surface Analogy and Spelling Rules in English Vowel Alternations". *Southwest Journal of Linguistics* FALTAN DATOS
- Eddington, D. "The Psychological Relevance of Phonological Generalizations in Spanish: An Experiment".
- Ehret, G., & Haack, B. 1982. "Ultrasound Recognition in House Mice: Key-Stimulus Configurations and Recognition Mechanisms". *JCP*, 148:245-251.
- Edwards, C. A., & Honig, W. K. 1987. "Memorization and 'Feature Selection' in the Acquisition of Natural Concepts in Pigeons". *LM*, 18:235-260.
- Eimas, P. D. 1974. "Auditory and Linguistic Processing of Cues for Place of Articulation by Infants". *PP*, 16:539-541.
- Eimas, P. D. 1994. "Categorisation in Early Infancy and the Continuity of Development". *CG*, 50:88-93.
- Eimas, P. D., & Nygaard, L. C. 1992. "Contextual Coherence and Attention in Phoneme Monitoring". *JML*, 31:375-395.
- Eimas, P. D., & Quinn, P. C. 1994. "Studies on the Formation of Perceptually Based Basic-Level Categories in Young Infants". *CD*, 65:903-917.
- Eimas, P. D., Quinn, P. C., & Cowan, P. 1994. "Development of Exclusivity in Perceptually Based Categories of Young Infants". *JECp*, 58:418-431.
- Eimas, P. D., Miller, J. L., & Jusczyk, P. W. 1987. "On Infant Speech Perception and the Acquisition of Language". In Harnad, S. (ed.). 1987. *Categorical Perception: The Groundwork of Cognition*. New York: CUP, 161-195.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. 1971. "Speech Perception in Infants". *Science*, 171:303-306.



- Elio, R., & Anderson, J. R. 1981. "The Effects of Category Generalizations and Instance Similarity on Schema Abstraction". *JEP:HLM*, 7:397-417.
- Elio, R., & Anderson, J. R. 1984. "The Effects of Information Order and Learning Mode on Schema Abstraction". *MC*, 12:20-30.
- Epstein, R. 1982. "A Note on the Mythological Character of Categorisation Research in Psychology". *JMB*, 3:161-169.
- Erreich, A., & Valian, V. 1979. "Children's Internal Organization of Locative Categories". *CD*, 50:1071-1077.
- Estes, W. K. 1986a. "Memory Storage and Retrieval Processes in Category Learning". *JEP:GE*, 115:155-174.
- Estes, W. K. 1986b. "Array Models for Category Learning". *CP*, 18:500-549.
- Fagan, J. F. 1976. "Infants' Recognition of Invariant Features of Faces". *CD*, 47:627-638.
- Fehr, B. 1988. "Prototype Analysis of the Concepts of Love and Commitment". *JPSP*, 55:557-579.
- Fehr, B., & Russell, J. A. 1984. "Concept of Emotion Viewed from a Prototype Perspective". *JEP:GE*, 113:464-486.
- Fenson, L., Vella, D., & Kennedy, M. 1989. "Children's Knowledge of Thematic and Taxonomic Relations at Two Years of Age". *CD*, 60:911-919.
- Fink, R. 1974. "Orthography and the Perception of Stops After s". *LS*, 17:152-159.
- Fivush, R. 1987. "Scripts and Categories: Interrelationships in Development". In Neisser, U. (ed.). 1987. *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorisation*. Cambridge: CUP, 234-254.
- Flanagan, M. J., Fried, L. S., & Holyoak, K. J. 1986. "Distributional Expectations and the Induction of Category Structure". *JEP:LMC*, 12:241-256.
- Flavell, J. H. 1985. [1977]. *Cognitive Development*. (2nd edition). Englewood Cliffs, NJ: Prentice-Hall.
- Fodor, J. A., Garrett, M. F., & Brill, S. L. 1975. "Pi-Ka-Pu: The Perception of Speech Sounds by Pre-Linguistic Infants". *PP*, 18:74-78.
- Forestell, P. H., & Herman, L. M. 1988. "Delayed Matching of Visual Materials by a Bottlenosed Dolphin Aided by Auditory Symbols". *ALB*, 16:137-146.
- Foss, D. J. 1969. "Decision processes during sentence comprehension: Effects of Lexical Item Difficulty and Position Upon Decision Times". *JVLVB*, 8:457-462.
- Foss, D. J. 1998. "Two Strands of Scholarship on Language Comprehension: Phoneme Monitoring and Discourse Context". *JPR*, 27:191-202.
- Foss, D. J. & Blank, M. A. 1980. "Identifying the Speech Codes". *CP*, 12:1-31.

- Foss, D. J., & Gernsbacher, M. A. 1983. "Cracking the Dual Code: Toward a Unitary Model of Phoneme Identification". *JVLVB*, 22:609-632.
- Foss, D. J., & Lynch, R. H. 1969. "Decision Processes During Sentence Comprehension: Effects of Surface Structure on Decision Time". *PP*, 5:145-148.
- Foss, D. J., & Swinney, D. A. 1973. "On the Psychological Reality of the Phoneme: Perception, Identification, and Consciousness". *JVLVB*, 12:246-257.
- Fox, R. A. 1984. "Effect of Lexical Status on Phonetic Categorisation". *JEP:HPP*, 10:526-540.
- Frauenfelder, U. H., & Segui, J. 1989. "Phoneme Monitoring and Lexical Processing: Evidence for Associative Context Effects". *MC*, 17, 134-140.
- Frazier, L. 1976. "On the Perception of /w, l, y/ and Categorical Perception". *JASA*, 59. S24(A).
- Fried, L. S., & Holyoak, K. J. 1984. "Induction of Category Distributions: A Framework for Classification Learning". *JEP:LMC*, 10:234-257.
- Fromkin, V. A. 1973. *Speech Errors as Linguistic Evidence*. The Hague: Mouton.
- Fromkin, V. A. 1980. "The Psychological Reality of Phonological Descriptions". In Fischer-Jørgensen, E., Rischel, J., & Thorsen, N. (eds.). 1980. *Proceedings of the Ninth International Congress of Phonetic Sciences*. Vol. 3. Copenhagen: University of Copenhagen, 195-216.
- Fudge, E. C. 1969. "Syllables". *JL*, 5:253-286.
- Fudge, E. C. 1987. "Branching Structure within the Syllable". *JL*, 23:359-377.
- Fudge, E. C. 1989. "Syllable Structure: A Reply to Davis". *JL*, 25:219-220.
- Ganong, W. F. 1980. "Phonetic Categorisation in Auditory Word Perception". *JEP:HPP* 6, 110-125.
- Gardner, R. A., & Gardner, B. T. 1984. "A Vocabulary Test for Chimpanzees (*Pan Troglodites*)". *Journal of Comparative Psychology* 98, 381-404.
- Geeraerts, D. 1988. "Where Does Prototypicality Come From?". In Rudzka-Ostyn, B. (ed.). 1988. *Topics in Cognitive Linguistics*. Amsterdam: John Benjamins, 207-229.
- Gelman, S. A. 1984. *Children's Inductive Inferences from Natural Kind and Artefact Categories*. Ph.D. diss. Stanford University. DAB 45, 3351B.
- Gelman, S. A. 1988. "The Development of Induction within Natural Kind and Artefact Categories". *CP*, 20:65-95.
- Gelman, S. A., & Coley, J. D. 1990. "The Importance of Knowing a Dodo Is a Bird: Categories and Inferences in 2-Year-Old Children". *DP*, 26:796-804.
- Gelman, S. A., & Markman, E. M. 1986. "Categories and Induction in Young Children". *CG*, 23:183-209.
- Gelman, S. A., & Markman, E. M. 1987. "Young Children's Inductions from Natural Kinds: The Role of Categories and Appearances". *CD*, 58:1532-1541.



- Gelman, S. A., & O'Reilly, A. W. 1988. "Children's Inductive Inferences within Superordinate Categories: The Role of Language and Category Structure". *CD*, 59:876-887.
- Gelman, S. A., Collman, P., & Maccoby, E. E. 1986. "Inferring Properties from Categories versus Inferring Categories from Properties: The Case of Gender". *CD*, 57:396-404.
- Gelman, S. A., Coley, J. D., & Gottfried, G. M. 1994. "Essentialist Beliefs in Children: The Acquisition of Concepts and Theories". In Hirschfeld, A., & Gelman, S. A. (eds.). 1994. *Mapping the Mind: Domain Specificity in Cognition and Culture*. Cambridge: CUP, 341-367.
- Gerhardt, H. C. 1978. "Discrimination of Intermediate Sounds in a Synthetic Call Continuum by Female Green Treefrogs". *Science*, 199:1089-1091.
- Gimson, A. C. 1980. [1962]. *An Introduction to the Pronunciation of English*. (3rd edition). London: Edward Arnold.
- Givón, T. 1986. "Prototypes: Between Plato and Wittgenstein". In Craig, C. (ed.). 1986. *Noun Classes and Categorisation*. Amsterdam: John Benjamins, 77-102.
- Glass, A. L., & Meany, P. J. 1978. "Evidence for Two Kinds of Low-Typical Instances in a Categorisation Task". *MC*, 6:622-628.
- Gleason, H. A. 1955. *An Introduction to Descriptive Linguistics*. New York: Holt, Rinehart & Winston.
- Gleitman, L. R., Armstrong, S. L., & Gleitman, H. 1983. "On Doubting the Concept 'Concept' ". In Scholnick, E. K. (ed.). 1983. *New Trends in Conceptual Representation: Challenges to Piaget's Theory?*. Hillsdale, NJ: LEA, 87-110.
- Gluck, M. A. 1991. "Stimulus Generalization and Representation in Adaptive Network Models of Category Learning". *PSS*, 2:50-55.
- Gluck, M. A., & Bower, G. H. 1988. "From Conditioning to Category Learning: An Adaptive Network Model". *JEP:GE*, 117:227-247.
- Goldman, D., & Homa, D. 1977. "Integrative and Metric Properties of Abstracted Information as a Function of Category Discriminability, Instance Variability, and Experience". *JEP:HLM*, 3:375-385.
- Goldsmith, J. A. 1995. "Phonological Theory". In Goldsmith, J. A. (ed.). 1995. *The Handbook of Phonological Theory*. Oxford, UK: Blackwell, 1-23.
- Goldstone, R. L. 1994a. "Influences of Categorisation on Perceptual Discrimination". *JEP:GE*, 123:178-200.
- Goldstone, R. L. 1994b. "The Role of Similarity in Categorisation: Providing a Groundwork". *CG*, 52:125-157.
- Goldstone, R. L. 1995. "Effects of Categorisation on Color Perception". *PSS*, 6:298-304.
- Gopnik, A., & Meltzoff, A. 1987. "The Development of Categorisation in the Second Year and its Relation to Other Cognitive and Linguistic Developments". *CD*, 58:1523-1531.
- Gopnik, A., & Meltzoff, A. 1992. "Categorisation and Naming: Basic-Level Sorting in Eighteen-Month-Olds and Its Relation to Language". *CD*, 63:1091-1103.

- Gottlieb, G. 1980. "Development of Species Identification in Ducklings: VI. Specific Embryonic Experience Required to Maintain Species-Typical Perception in Peking Ducklings". *JCPP*, 94:579-587.
- Graesser, A. C., Gordon, S. E., & Sawyer, J. 1979. "Recognition Memory for Typical and Atypical Actions in Scripted Activities: Tests of a Script Pointer+Tag Hypothesis". *JVLVB*, 18:319-332.
- Greenberg, J. H., & Jenkins, J. J. 1964. "Studies in the Psychological Correlates of the Sound System of American English". *Word*, 20:157-177.
- Greenberg, M. S., & Bjorklund, D. F. 1981. "Category Typicality in Free Recall: Effects of Feature Overlap on Differential Category Encoding?". *JEP:HLM*, 7:145-147.
- Grieser, D. L., & Kuhl, P. K. 1983. "The Internal Structure of Vowel Categories in Infancy: Effects of Stimulus Goodness". *JASA*, 74 (Suppl.1). S102(A).
- Grieser, D. L., & Kuhl, P. K. 1989. "Categorisation of Speech by Infants: Support for Speech-Sound Prototypes". *DP*, 25:577-588.
- Hahn, U., & Chater, N. 1997. "Concepts and Similarity". In Lamberts, K., & Shanks, D. (eds.). 1997. *Knowledge, Concepts, and Categories*. Cambridge, MA: MIT Press, 49-92.
- Hampson, S. E. 1982. "Person Memory: A Semantic Category Model of Personality Traits". *BJP*, 73:1-11.
- Hampton, J. A. 1979. "Polymorphous Concepts in Semantic Memory". *JVLVB*, 18:441-461.
- Hampton, J. A. 1981. "An Investigation of the Nature of Abstract Concepts". *MC*, 9:149-156.
- Hampton, J. A. 1982. "A Demonstration of Intransitivity in Natural Concepts". *CG*, 12:151-164.
- Hampton, J. A. 1987. "Inheritance of Attributes in Natural Concept Conjunctions". *MC*, 15:55-71.
- Hampton, J. A. 1988a. "Overextension of Conjunctive Concepts: Evidence for a Unitary Model of Concept Typicality and Class Inclusion". *JEP:LMC*, 14:12-32.
- Hampton, J. A. 1988b. "Disjunction of Natural Concepts". *MC*, 16:579-591.
- Hampton, J. A. 1991. "The Combination of Prototype Concepts". In Schwanenflugel, P. (ed.). 1991. *The Psychology of Word Meanings*. Hillsdale, NJ: LEA, 91-116.
- Hampton, J. A. 1993. "Prototype Models of Concept Representation". In Van Mechelen, I., Hampton, J. A., Michalski, R. S., & Theuns, P. (eds.). 1993. *Categories and Concepts: Theoretical Views and Inductive Data Analysis*. New York: Academic Press, 67-95.
- Hampton, J. A. 1995. "Testing the Prototype Theory of Concepts". *JML*, 34:686-708.
- Hampton, J. A. 1997. "Conceptual Combination". In Lamberts, K., & Shanks, D. (eds.). 1997. *Knowledge, Concepts, and Categories*. Cambridge, MA: MIT Press, 133-160.



- Hampton, J. A., & Dubois, D. 1993. "Psychological Models of Concepts: Introduction". In Van Mechelen, L., Hampton, J., Michalski, R. S., & Theuns, P. (eds.). 1993. *Categories and Concepts: Theoretical Views and Inductive Data Analyses*. New York: Academic Press, 11-32.
- Hampton, J. A., & Gardiner, M. M. 1983. "Measures of Internal Category Structure: A Correlational Analysis of Normative Data". *BJP*, 74:491-516.
- Harris, Z. S. 1944. "Simultaneous Components in Phonology". *Lg*, 20:181-205
- Hartley, J., & Homa, D. 1981. "Abstraction of Stylistic Concepts". *JEP:HLM*, 7:33-46.
- Hary, J. M., & Massaro, D. W. 1982. "Categorical Results do not Imply Categorical Perception". *PP*, 32: 409-418.
- Hastie, R., Schroeder, C., & Weber, R. 1990. "Creating Complex Social Conjunction Categories from Simple Categories". *BPS*, 28:242-247.
- Hawkins, B. 1988. "The Natural Category MEDIUM: An Alternative to Selection Restrictions and Similar Constructs". In Rudzka-Ostyn, B. (ed.). 1988. *Topics in Cognitive Linguistics*. Amsterdam: John Benjamins, 231-270.
- Hayes, B. K., & Taplin, J. E. 1992. "Developmental Changes in Categorisation Processes: Knowledge and Similarity-Based Modes of Categorisation". *JECp*, 54:188-212.
- Hayes-Roth, B., & Hayes-Roth, F. 1977. "Concept Learning and the Recognition and Classification of Exemplars". *JVLVB*, 16:321-338.
- Haygood, R. C., & Bourne, L. E. 1965. "Attribute and Rule-Learning Aspects of Conceptual Behavior". *Psychological Review*, 72:175-195.
- Healy, A. F., & Cutting, J. E. 1976. "Units of Speech Perception: Phoneme and Syllable". *JVLVB*, 15:73-83.
- Healy, A. F., & Repp, B. H. 1982. "Context Independent and Phonetic Mediation in Categorical Perception". *JEP:HPP*, 8:68-80.
- Heidbreder, E. 1946. "The Attainment of Concepts I. Terminology and Methodology". *Journal of General Psychology*, 35:173-189.
- Heider, E. R. 1971. " 'Focal' Color Areas and the Development of Color Names". *DP*, 4:447-455.
- Heider, E. R. 1972. "Universals in Color Naming and Memory". *JEP*, 93:10-20.
- Heit, E. 1992. "Categorisation Using Chains of Examples". *CP*, 24:341-380.
- Heit, E. 1994. "Models of the Effects of Prior Knowledge on Category Learning". *JEP:LMC*, 20:1264-1282.
- Heit, E. 1997. "Knowledge and Concept Learning". In Lamberts, K., & Shanks, D. (eds.). 1997. *Knowledge, Concepts, and Categories*. Cambridge, MA: MIT Press, 7-40.
- Hemenway, K. 1981. *The Role of Perceived Parts in Categorisation*. Ph.D. Diss. Stanford University.

- Herman, L. M., Hovancik, J. R., Gory, J. D., & Bradshaw, G. L. 1989. "Generalization of Visual Matching by a Bottlenosed Dolphin (*Tursiops Truncatus*): Evidence for Invariance of Cognitive Performance with Visual and Auditory Materials". *JEP:ABP*, 15:124-136.
- Herrnstein, R. J. 1979. "Acquisition, Generalization, and Discrimination Reversal of a Natural Concept". *JEP:ABP*, 5:116-129.
- Herrnstein, R. J. 1985. "Riddles of Natural Categorisation". *Philosophical Transactions of the Royal Society*, 308:129-144.
- Herrnstein, R. J., & de Villiers, P. A. 1980. "Fish As a Natural Category for People and Pigeons". In Bower, G. H. (ed.). 1980. *The Psychology of Learning and Motivation*. Vol. 14. New York: Academic Press, 59-95.
- Herrnstein, R. J., Loveland, D. H., & Cable, C. 1976. "Natural Concepts in Pigeons". *JEP:ABP*, 2:285-302.
- Hillenbrand, J. 1984. "Speech Perception by Infants: Categorisation Based on Nasal Consonant Place of Articulation". *JASA*, 75:1613-1622.
- Hintzman, D. L. 1986. " 'Schema Abstraction' in a Multiple-Trace Memory Model". *PR*, 93:411-428.
- Hintzman, D. L., & Ludlam, G. 1980. "Differential Forgetting of Prototypes and Old Instances: Simulation by an Exemplar-Based Classification Model". *MC*, 8:378-382.
- Hoard, J. E. 1971. "Aspiration, Tenseness, and Syllabification in English". *Lg*, 47:133-140.
- Hockett, C. F. 1942. "A System of Descriptive Phonology". *Lg*, 20, 181-205
- Hockett, C. F. 1955. *A Manual of Phonology*. *International Journal of American Linguistics*. Vol. 21. Baltimore: Waverly Press.
- Hockett, C. F. 1958. *A Course in Modern Linguistics*. New York: Macmillan.
- Hodgson, P. 1993. *Internal Structure of Phonetic Categories: The Role of Multiple Acoustic Properties*. Ph.D. Thesis. Northeastern University.
- Hodgson, P., & Miller, J. L. 1991. "The Role of Multiple Acoustic Properties in Specifying the Internal Structure of Phonetic Categories". *JASA*, 89, 1997.
- Hodgson, P., & Miller, J. L. 1992. "Internal Phonetic Category Structure Depends on Multiple Acoustic Properties: Evidence for Within-Category Trading Relations". *JASA*, 92, 2464.
- Hodgson, P., & Miller, J. L. 1996. "Internal Structure of Phonetic Categories: Evidence for Within-Category Trading Relations". *JASA*, 100:565-576.
- Hoffman, C., Mischel, W., & Mazze, K. 1981. "The Role of Purpose in the Organization of Information About Behavior: Trait-Based Versus Goal-Derived Categories in Person Cognition". *JPSP*, 40:211-224.
- Homa, D. 1984. "On the Nature of Categories". In Bower, G. H. (ed.). 1984. *The Psychology of Learning and Motivation*. Vol. 18. New York: Academic Press, 49-94.



- Homa, D., & Chambliss, D. 1975. "The Relative Contribution of Common and Distinctive Information on the Abstraction from Ill-Defined Categories". *JEP:HLM*, 104:351-359.
- Homa, D., & Cultice, J. 1984. "Role of Feedback, Category Size, and Stimulus Distortion on the Acquisition and Utilization of Ill-Defined Categories". *JEP:LMC*, 10:83-94.
- Homa, D., Burrell, L., & Field, D. 1987. "The Changing Composition of Abstracted Categories under Manipulations of Decisional Change, Choice Difficulty, and Category Size". *JEP:LMC*, 13:401-412.
- Homa, D., Cross, J., Cornell, D., Goldman, D., & Schwartz, S. 1973. "Prototype Abstraction and Classification of New Instances As a Function of the Number of Instances Defining the Prototype". *JEP*, 101:116-122.
- Homa, D., Sterling, S., & Trepel, L. 1981. "Limitations of Exemplar-Based Generalization and the Abstraction of Categorical Information". *JEP:HLM*, 7:418-439.
- Homa, D., & Vosburgh, R. 1976. "Category Breadth and the Abstraction of Prototypical Information". *JEP:HLM*, 2:322-230.
- Hombert, J.-M. 1986. "Word Games: Some Implications for Analysis of Tone and Other Phonological Constructs". In Ohala, J. J., & Jaeger, J. J. (eds.). 1986. *Experimental Phonology*. Orlando: Academic Press, 175-186.
- Honeck, R. P., Firment, M., & Case, T. J. 1987. "Expertise and Categorisation". *BPS*, 25:431-434.
- Honeck, R. P., Kibler, C. T., & Sugar, J. 1985. "The Conceptual Base View of Categorisation". *JPR*, 14:155-174.
- Hooper, J. B. 1972. "The Syllable in Phonological Theory". *Lg*, 48:525-540.
- Hopper, P. J., & Thompson, S. A. 1984. "The Discourse Basis for Lexical Categories in Universal Grammar". *Lg*, 60:703-752.
- Horowitz, L. M., Wright, J. C., Lowenstein, E., & Parad, H. W. 1981. "The Prototype as a Construct in Abnormal Psychology: 1. A Method for Deriving Prototypes". *Journal of Abnormal Psychology*, 90:568-574.
- Horton, M. S. 1982. *Category Familiarity and Taxonomic Organization in Young Children*. Phd. Diss. Stanford University, Palo Alto, CA.
- Horton, M. S., & Markman, E. M. 1980. "Developmental Differences in the Acquisition of Basic and Superordinate Categories". *CD*, 51:708-719.
- Hovland, C. I., & Weiss, W. 1953. "Transmission of Information Concerning Concepts Through Positive and Negative Instances". *JEP*, 45:175-182.
- Howard, J. W., & Rothbart, M. 1980. "Social Categorisation and Memory for In-Group and Out-Group Behavior". *JPSP*, 38:301-310.
- Hubbell, A. F. 1950. *The Pronunciation of English in New York City: Consonants and Vowels*. New York: Kings Crown Press.

- Huber, L., & Lenz, R. 1993. "A Test of the Linear Feature Model of Polymorphous Concept Discrimination with Pigeons". *QJEP:CPP*, 46:1-18.
- Huber, L., & Lenz, R. 1996. "Categorisation of Prototypical Stimulus Classes by Pigeons". *QJEP:CPP*, 49:111-133.
- Hull, C. L. 1920. "Quantitative Aspects of the Evolution of Concepts: An Experimental Study". *Psychological Monographs*, 28. No. 1. (Whole No. 123).
- Hulse, S. H., Egeth, H., & Deese, J. 1980. [1952]. *The Psychology of Learning*. (5th edition). New York: McGraw-Hill Book Company.
- Hunt, E. B. 1962. *Concept Learning: An Information Processing Problem*. New York: John Wiley & Sons.
- Hupp, S. C., & Mervis, C. B. 1982. "Acquisition of Basic Object Categories by Severely Handicapped Children". *CD*, 53:760-767.
- Hurch, B. 1988. *Über Aspiration. Ein Kapitel aus der natürlichen Phonologie*. Tübingen: Narr.
- Hyman, L. 1975. *Phonology: Theory and Analysis*. New York: Holt, Rinehart, & Winston.
- Isen, A. M., & Daubman, K. A. 1984. "The Influence of Affect on Categorisation". *JPSP*, 47:1206-1217.
- Isen, A. M., Niedenthal, P. M., & Cantor, N. 1992. "An Influence of Positive Affect on Social Categorisation". *Motivation and Emotion*, 16:65-78.
- Jacoby, L. L., & Brooks, L. R. 1984. "Nonanalytic Cognition: Memory, Perception, and Concept Learning". In Bower, G. H. (ed.). 1984. *The Psychology of Learning and Motivation*. Vol. 18. New York: Academic Press, 1-47.
- Jaeger, J. J. 1980a. *Categorisation in Phonology: An Experimental Approach*. Ph.D. diss., University of California, Berkeley.
- Jaeger, J. J. 1980b. "Testing the Psychological Reality of Phonemes". *LS*, 23:233-253.
- Jaeger, J. J. 1984. "Assessing the Psychological Status of the Vowel Shift Rule". *JPR*, 13:13-56.
- Jaeger, J. J. 1986a. "Concept Formation as a Tool for Linguistic Research". In Ohala, J. J., & Jaeger, J. J. (eds.). 1986. *Experimental Phonology*. Orlando: Academic Press, 211-238.
- Jaeger, J. J. 1986b. "On the Acquisition of Abstract Representations for English Vowels". *PHY*, 3:71-97.
- Jaeger, J. J., & Ohala, J. J. 1984. "On the Structure of Phonetic Categories". *BLS*, 10:15-26.
- Jakobson, R. 1968. *Child Language, Aphasia, and Phonological Universals*. The Hague: Mouton.
- Jakobson, R., Fant, G., & Halle, M. 1952. *Preliminaries to Speech Analysis*. Cambridge, MA: MIT Press.
- Jitsumori, M. 1993. "Category Discrimination of Artificial Polymorphous Stimuli Based on Feature Learning". *JEP: ABP*, 19:244-254.



- Jitsumori, M. 1994. "Discrimination of Artificial Polymorphous Categories by Rhesus Monkeys (*Macaca Mulatta*)". *QJEP:CPP*, 47:371-386.
- Jitsumori, M. 1996. "A Prototype Effect and Categorisation of Artificial Polymorphous Stimuli in Pigeons". *JEP:ABP*, 22:405-419.
- Johansson, S., & Hofland, K. 1989. *Frequency Analysis of English Vocabulary and Grammar. Based on the LOB Corpus. Volume 1. Tag Frequencies and Word Frequencies*. Oxford: Clarendon Press.
- John, O. P., Hampson, S. E., & Goldberg, L. R. 1991. "The Basic Level in Personality-Trait Hierarchies: Studies of Trait Use and Accessibility in Different Contexts". *JPSP*, 60:348-361.
- Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. 1984. "Pictures and Names: Making the Connection". *CP*, 16:243-275.
- Jones, D. 1984. [1909]. *The Pronunciation of English*. (4th edition). Cambridge University Press.
- Jones, D. 1989. [1918]. *An Outline of English Phonetics*. (9th edition). Cambridge: CUP.
- Jones, D. 1973. "The History and Meaning of the Term Phoneme" In Jones, W. E., & Laver, J. (eds). 1973. *Phonetics in Linguistics: A Book Readings*. London: Longman, 187-204.
- Jones, D. 1950. *The Phoneme: Its Nature and Use*. Heffer: Cambridge.
- Jones, G. V. 1982. "Stacks Not Fuzzy Sets: An Ordinal Basis for Prototype Theory of Concepts". *CG*, 12:281-290.
- Jones, G. V. 1983. "Identifying Basic Categories". *PB*, 94:423-428.
- Kalish, C. W., & Gelman, S. A. 1992. "On Wooden Pillows: Multiple Classification and Children's Category-Based Inductions". *CD*, 63:1536-1557.
- Kaplan, A. S., & Murphy, G. L. 1999. "The Acquisition of Category Structure in Unsupervised Learning". *MC*, 27:699-712.
- Katz, J. J., & Postal, P. M. 1964. *An Integrated Theory of Linguistic Descriptions*. Cambridge, MA: MIT Press.
- Kawasaki, H. 1986. "Phonetic Explanation for Phonological Universals: The Case of Distinctive Vowel Nasalization". In Ohala, J. J., & Jaeger, J. J. (eds.). 1986. *Experimental Phonology*. Orlando: Academic Press, 81-103.
- Kay, P., & McDaniel, C. K. 1978. "The Linguistic Significance of the Meanings of Basic Color Terms". *Lg*, 54:610-646.
- Kaye, J. 1989. *Phonology: A Cognitive View*. Hillsdale, NJ: LEA.
- Keil, F. C. 1987. "Conceptual Development in Category Structure". In Neisser, U. (ed.). 1987. *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorisation*. New York: CUP, 175-200.

- Keil, F. C., & Kelly, M. H. 1987. "Developmental Changes in Category Structure". In Harnad, S. (ed.). 1987. *Categorical Perception: The Groundwork of Cognition*. New York: CUP, 491-510.
- Keller, D. 1982. "Developmental Effects of Typicality and Superordinate Property Dominance on Sentence Verification". *JECP*, 33:288-297.
- Keller, D., & Kellas, G. 1978. "Typicality as a Dimension of Encoding". *JEP:HLM*, 4:78-85.
- Kellogg, R. T. 1981. "Feature Frequency in Concept Learning: What Is Counted?". *MC*, 9:157-163.
- Kelly, M. H., Bock, J. K., & Keil, F. C. 1986. "Prototypicality in a Linguistic Context: Effects on Sentence Structure". *JML*, 25:59-74.
- Kemler Nelson, D. G. 1984. "The Effect of Intention on What Concepts Are Acquired". *JVLVB*, 23:734-759.
- Kemler Nelson, D. G. 1988. "When Category Learning Is Holistic: A Reply to Ward and Scott". *MC*, 16:79-84.
- Kempton, W. 1978. "Category Grading and Taxonomic Relations: A Mug Is a Sort of a Cup". *AE*, 5:44-65.
- Kempton, W. 1981. *The Folk Classification of Ceramics: A Study of Cognitive Prototypes*. New York: Academic Press.
- Kendler, T. S. 1961. "Concept Formation". *ARP*, 12:447-472.
- Kendrick, D. F. 1992. "Pigeon's Concept of Experienced and Non-Experienced Real-World Locations: Discrimination and Generalization Across Seasonal Variations". In Honig, W. K., & Fetterman, J. G. (eds.). 1992. *Cognitive Aspects of Stimulus Control*. Hillsdale, NJ: LEA, 113-134.
- Kenstowicz, M., & Kisseberth, C. 1979. *Generative Phonology*. San Diego: Academic Press.
- Kidd, G. R. 1989. "Articulatory-Rate Context Effects in Phoneme Identification". *JEP:HPP*, 15:736-748.
- Kim, C. W. 1970. "A Theory of Aspiration". *Phonetica*, 21:107-116.
- King, J. E. 1973. "Learning and Generalizations of a Two-Dimensional Sameness-Difference Concept by Chimpanzees and Orangutans". *JCPP*, 84:140-148.
- Kiparsky, P. 1975. "What are Phonological Theories About?". In Cohen, D., & Wirth, J. R. (eds.). 1975. *Testing Linguistic Hypotheses*. Washington DC: Hemisphere.
- Kirtley, C., Bryant, P., MacLean, M., & Bradley, L. 1989. "Rhyme, Rime, and the Onset of Reading". *JECP*, 48:224-245.
- Klatt, D. H. 1975. "Voice Onset Time, Frication, and Aspiration in Word-Initial Consonant Clusters". *Journal of Speech and Hearing Research*, 18:686-706.
- Kluender, K. R., Diehl, R. L., & Killeen, P. R. 1987. "Japanese Quail Can Learn Phonetic Categories". *Science*, 237:1195-1197.



- Knapp, A. G., & Anderson, J. A. 1984. "Theory of Categorisation Based on Distributed Memory Storage". *JEP:LMC*, 10:616-637.
- Komatsu, L. K. 1992. "Recent Reviews of Conceptual Structure". *PB*, 112:500-526.
- Kossan, N. E. 1981. "Developmental Differences in Concept Acquisition Strategies". *CD*, 52:290-298.
- Krámsky, J. 1974. *The Phoneme: Introduction to the History and Theories of a Concept*. München: Wilhelm fink Verlag.
- Kreidler, C. W. 1989. *The Pronunciation of English: A Course Book in Phonology*. Oxford, UK: Blackwell.
- Kruschke, J. K. 1992. "ALCOVE: An Exemplar-Based Connectionist Model of Category Learning". *PR*, 99:22-44.
- Kuhl, P. K. 1977. "Speech Perception in Early Infancy: Perceptual Constancy for Vowel Categories /a/ and /ɔ/". *JASA*, 61. S39(A).
- Kuhl, P. K. 1979. "Speech Perception in Early Infancy: Perceptual Constancy for Spectrally Dissimilar Vowel Categories". *JASA*, 66:1668-1679.
- Kuhl, P. K. 1980. "Perceptual Constancy for Speech-Sound Categories in Early Infancy". In Yeni-Komshian, G. H., Kavanagh, J. F., & Ferguson, C. A. (eds.). 1980. *Child Phonology. Vol. 2.: Perception*. New York : Academic Press, 41-66.
- Kuhl, P. K. 1981. "Discrimination of Speech by Nonhuman Animals: Basic Auditory Sensitivities Conducive to the Perception of Speech-Sound Categories". *JASA*, 70:340-349.
- Kuhl, P. K. 1985. "Categorisation of Speech by Infants". In Mehler, J., & Fox, R. (eds.). 1985. *Neonate Cognition: Beyond the Blooming, Buzzing Confusion*. Hillsdale, NJ: LEA, 231-262.
- Kuhl, P. K. 1987. "The Special-Mechanisms Debate in Speech Research: Categorisation Tests on Animals and Infants". In Harnad, S. (ed.). 1987. *Categorical Perception: The Groundwork of Cognition*. New York: CUP, 355-386.
- Kuhl, P. K. 1991. "Human Adults and Human Infants Show a 'Perceptual Magnet Effect' for the Prototypes of Speech Categories, Monkeys Do Not". *PP*, 50:93-107.
- Kuhl, P. K., & Meltzoff, A. 1982. "The Bimodal Perception of Speech in Infancy". *Science*, 218:1138-1141.
- Kuhl, P. K., & Meltzoff, A. 1984. "The Intermodal Representation of Speech in Infants". *IBD*, 7:361-381.
- Kuhl, P. K., & Miller, J. D. 1975a. "Speech Perception by the Chinchilla: Voiced-Voiceless Distinction in Alveolar Plosive Consonants". *Science*, 190:69-72.
- Kuhl, P. K., & Miller, J. D. 1975b. "Speech Perception in Early Infancy: Discrimination of Speech-Sound Categories". *JASA*, 58. S56(A).
- Kuhl, P. K., & Miller, J. D. 1978. "Speech Perception by the Chinchilla: Identification Functions for Synthetic VOT Stimuli". *JASA*, 63:905-917.

- Kuhl, P. K., & Padden, D. M. 1982. "Enhanced Discriminability at the Phonetic Boundaries for the Voicing Feature in Macaques". *PP*, 32:542-550.
- Kuhl, P. K., & Padden, D. M. 1983. "Enhanced Discriminability at the Phonetic Boundaries for the Place Feature in Macaques". *JASA*, 73:1003-1010.
- Labov, W. 1973. "The Boundaries of Words and Their Meanings". In Bailey, C.-J. N., & Shuy, R. W. (eds.). 1973. *New Ways of Analyzing Variation in English*. Vol. 1. Washington, DC: Georgetown University Press, 340-373.
- Ladefoged, P. 1993. [1975]. *A Course in Phonetics*. (3rd edition). New York: Harcourt, Brace, Jovanovich.
- Lakoff, G. 1973. "Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts". *Journal of Philosophical Logic*, 2:458-508.
- Lakoff, G. 1982. "Categories and Cognitive Models". *Berkeley Cognitive Science Report no 2*. Berkeley, CA: Institute for Human Learning, University of California.
- Lakoff, G. 1986. "Classifiers As a Reflection of Mind". In Craig, C. (ed.). 1986. *Noun Classes and Categorisation*. Amsterdam: John Benjamins, 13-51.
- Lakoff, G. 1987a. *Women, Fire and Dangerous Things: What Categories Reveal About the Mind*. Chicago, IL: University of Chicago Press.
- Lakoff, G. 1987b. "Cognitive Models and Prototype Theory". In Neisser, U. (ed.). 1987. *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorisation*. New York: CUP, 63-100.
- Lamberts, K. 1994. "Flexible Tuning of Similarity in Exemplar-Based Categorisation". *JEP:LMC*, 20:1003-1021.
- Lamberts, K. 1995. "Categorisation under Time Pressure". *JEP:GE*, 124:161-180.
- Lamberts, K. 1998. "The Time Course of Categorisation". *JEP:LMC*, 24:695-711.
- Lamberts, K., & Brockdorff, N. 1997. "Fast Categorisation of Stimuli with Multivalued Dimensions". *MC*, 25:296-304.
- Lamberts, K., & Freeman, R. J. 1999. "Categorisation of Briefly Presented Objects". *Psychological Research*, 62:107-117.
- Langacker, R. 1988. "A Usage-Based Model". In Rudka-Ostyn, B. (ed.). 1988. *Topics in Cognitive Linguistics*. Amsterdam: John Benjamins, 127-161.
- Lass, R. P. 1984. *Phonology: An Introduction to Basic Concepts*. Cambridge: Cambridge University Press.
- Lassaline, M. E., Wisniewski, E. J., & Medin, D. L. 1992. "The Basic Level in Artificial and Natural Categories: Are All Levels Created Equal?". In Burns, B. (ed.). 1992. *Percepts, Concepts, and Categories: The Representation and Processing of Information*. Amsterdam: Elsevier, 327-278.



- Lea, S. E. G. 1984. "In What Sense Do Pigeons Learn Concepts?". In Roitblat, H. L., Bever, T. G., & Terrace, H. S. (eds.). 1984. *Animal Cognition*. Hillsdale, NJ: LEA, 263-277.
- Lea, S. E. G., & Harrison, S. N. 1978. "Discrimination of Polymorphous Stimulus Sets by Pigeons". *QJEP*, 30:521-537.
- Lea, S. E. G., & Ryan, C. M. E. 1990. "Unnatural Concepts and the Theory of Concept Discrimination in Birds". In Commons, M. L., Herrnstein, R. J., Kosslyn, S., & Mumford, D. (eds.). 1990. *Quantitative Analyses of Behavior: Behavioral Approaches to Pattern Recognition and Concept Formation*. Vol. 8. Cambridge, MA: Ballinger, 165-185.
- Lécuyer, R. 1991. "La Catégorisation de Formes Géométriques Continues chez des Enfants de Cinq Mois". *Archives de Psychologie*, 59:143-155.
- Lécuyer, R., & Poirier, C. 1994. "Categorisation of Geometric Figures Composed of Four Elements by 5.5-Month-Old Infants". *CPC*, 13:493-509.
- Levine, S. C., & Carey, S. 1982. "Up Front: The Acquisition of a Concept and a Word". *JCL*, 9:645-657.
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. 1957. "The Discrimination of Speech Sounds within and across Phoneme Boundaries". *JEP*, 54:358-368.
- Liberman, A. M., Cooper, F. S., & Shankweiler, D. P., & Studdert-Kennedy, M. 1967. "Perception of the Speech Code". *PR*, 74:431-461.
- Lin, E. L., & Murphy, G. L. 1997. "Effects of Background Knowledge on Object Categorisation and Part Detection". *JEP:HPP*, 23:1153-1169.
- Lin, E. L., Murphy, G. L., & Shoben, E. J. 1997. "The Effects of Prior Processing Episodes on Basic-Level Superiority". *QJEP:HEP*, 50:25-48.
- Lin, P. J., Schwanenflugel, P. J., & Wisenbaker, J. M. 1990. "Category Typicality, Cultural Familiarity, and the Development of Category Knowledge". *DP*, 26:805-813.
- Lingle, J. H., Altom, M. W., & Medin, D. L. 1984 "Of Cabbages and Kings: Assessing the Extendibility of Natural Object Concept Models to Social Things". In Wyer, R. S., & Srull, T. K. (eds.). 1984. *Handbook of Social Cognition*. Vol. 1. Hillsdale, NJ: LEA, 71-117.
- Linell, P. 1979. *Psychological Reality in Phonology*. Cambridge: Cambridge University Press.
- Lisker, L., & Abramson, A. S. 1964. "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements". *Word*, 20:384-422.
- Loken, B., & Ward, J. 1990. "Alternative Approaches to Understanding the Determinants of Typicality". *Journal of Consumer Research*, 17:111-126.
- Lotto, A. J. 2000. "Language Acquisition as Complex Category Formation". *PHN*, 57:189-196.
- Lotz, J., Abramson, A. S., Gerstman, L. J., Ingemann, F., & Nemser, W. 1960. "The Perception of English Stops by Speakers of English, Spanish, Hungarian, and Thai : A Tape-Cutting Experiment". *LS*, 3:71-77.
- Lubow, R. E. 1974. "High-Order Concept Formation in the Pigeon". *JEAB*, 21:475-483.

- MacKay, D. G. 1972. "The Structure of Words and Syllables: Evidence from Errors in Speech". *CP*, 3:210-227.
- MacKintosh, N. J. 1995. "Categorisation by People and Pigeons: The Twenty-Second Bartlett Memorial Lecture". *QJEP:CPP*, 48:193-214.
- MacPhail, E. M., & Reilly, S. 1989. "Rapid Acquisition of a Novelty Versus Familiarity Concept by Pigeons (*Columba Livia*)". *JEP:ABP*, 15:242-252.
- MacWhinney, B. 1984. "Commentary: Where do Categories Come From?". In Sophian, C. (ed.). 1984. *Origins of Cognitive Skills*. Hillsdale, NJ: LEA, 407-415.
- Maddieson, J. 1985. *Patterns of Sounds*. Cambridge: CUP.
- Malot, R. W., & Siddall, J. W. 1972. "Acquisition of the People Concept in Pigeons". *Psychological Reports*, 31:3-13.
- Malt, B. C. 1989. "An On-Line Investigation of Prototype and Exemplar Strategies in Classification". *JEP:LMC*, 15:539-555.
- Malt, B. C. 1990. "Features and Beliefs in the Mental Representation of Categories". *JML*, 29:289-315.
- Malt, B. C. 1993. "Concept Structure and Category Boundaries". In Nakamura, G. V., Taraban, R. M., & Medin, D. L. (eds.). 1993. *The Psychology of Learning and Motivation. Vol. 29.: Categorisation by Humans and Machines*. New York: Academic Press, 363-390.
- Malt, B. C. 1994. "Water Is not H₂O". *CP*, 27:41-70.
- Malt, B. C. 1995. "Category Coherence in Cross-Cultural Perspective". *CP*, 29:85-148.
- Malt, B. C., & Johnson, E. C. 1992. "Do Artefact Concepts Have Cores?". *JML*, 31:195-217.
- Malt, B. C., & Smith, E. E. 1982. "The Role of Familiarity in Determining Typicality". *MC*, 10:69-75.
- Malt, B. C., & Smith, E. E. 1984. "Correlated Properties in Natural Categories". *JVLVB*, 23:250-269.
- Malt, B. C., Ross, B. H., & Murphy, G. L. 1995. "Predicting Features for Members of Natural Categories when Categorisation Is Uncertain". *JEP:LMC*, 21:646-661.
- Mandler, J. M. 1997. "Development of Categorisation: Perceptual and Conceptual Categories". In Bremner, G., Slater, A., & Butterworth, G. (eds.). 1997. *Infant Development: Recent Advances*. East Sussex, UK: Psychology Press, 163-189.
- Mandler, J. M., & Bauer, P. J. 1988. "The Cradle of Categorisation: Is the Basic Level Basic?". *CGD*, 3:247-264.
- Mandler, J. M., & McDonough, L. 1993. "Concept Formation in Infancy". *CGD*, 8:291-318.
- Mandler, J. M., Bauer, P. J., & McDonough, L. 1991. "Separating the Sheep from the Goats: Differentiating Global Categories". *CP*, 23:263-298.



- Mandler, R. C. 1976. "Categorical Perception Along an Oral-Nasal Continuum". *JASA*, 59. S24(A).
- Maratsos, M. P. 1982. "The Child's Construction of Grammatical Categories". In Wanner, E., & Gleitman, L. R. (eds.). 1982. *Language Acquisition: The State of the Art*. New York: CUP, 240-266.
- Maratsos, M. P., & Chalkley, M. A. 1980. "The Internal Language of Children's Syntax: The Ontogenesis and Representation of Syntactic Categories". In Nelson, K. E. (ed.). 1980. *Children's Language. Vol 2*. New York: Gardner Press, 127-213.
- Margard, N. A. 1977. "Perceptual and Semantic Features in Children's Use of the Animate Concept". *DP*, 13:572-576.
- Margolis, E. 1994. "A Reassessment of the Shift from the Classical Theory of Concepts to Prototype Theory". *CG*, 51:73-89.
- Markman, E. M. 1978. "Empirical Versus Logical Solutions to Part-Whole Comparison Problems Concerning Classes and Collections". *CD*, 49:168-177.
- Markman, E. M. 1979. "Classes and Collections: Conceptual Organization and Numerical Abilities". *CP*, 11:395-411.
- Markman, E. M. 1983. "Two Different Kinds of Hierarchical Organization". In Scholnick, E. K. (ed.). 1983. *New Trends in Conceptual Representation: Challenges to Piaget's Theory?*. Hillsdale, NJ: LEA, 165-184.
- Markman, E. M. 1984. "The Acquisition and Hierarchical Organization of Categories by Children". In Sophian, C. (ed.). 1984. *Origins of Cognitive Skills*. Hillsdale, NJ: LEA, 371-406.
- Markman, E. M. 1985. "Why Superordinate Category Terms Can Be Mass Nouns". *CG*, 19:31-53.
- Markman, E. M. 1987. "How Children Constrain the Possible Meanings of Words". In Neisser, U. (ed.). 1987. *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorisation*. Cambridge, England: CUP, 256-287.
- Markman, E. M. 1989. *Categorisation and Naming in Children: Problems of Induction*. Cambridge, MA: MIT Press.
- Markman, E. M., & Hutchinson, J. E. 1984. "Children's Sensitivity to Constraints on Word Meaning: Taxonomic versus Thematic Relations". *CP*, 16:1-27.
- Markman, E. M., & Seibert, J. 1976. "Classes and Collections: Internal Organization and Resulting Holistic Properties". *CP*, 8:561-577.
- Markman, E. M., Horton, M. S., & McLanahan, A. G. 1980. "Classes and Collections: Principles of Organization in the Learning of Hierarchical Relations". *CG*, 8:227-241.
- Martin, R. C., & Caramazza, A. 1980. "Classification in Well-Defined and Ill-Defined Categories: Evidence for Common Processing Strategies". *JEP:GE*, 109:320-353.
- Martínez Celdrán, E. 1991. *Fonética Experimental: Teoría y Práctica*. Madrid: Síntesis

- Masataka, N. 1983. "Categorical Responses to Natural and Synthesized Alarm Calls in Goeldi's Monkeys (*Callimico Goeldi*)". *Primates*, 24:40-51.
- Massaro, D. W. 1987. "Categorical Partition: A Fuzzy-Logical Model of Categorisation Behavior". In Harnad, S. (ed.). 1987. *Categorical Perception: The Groundwork for Cognition*. New York: Academic Press, 254-283.
- Massaro, D. W., & Cohen, M. M. 1983. "Categorical or Continuous Speech Perception: A New Test". *Speech Communication*, 2:15-35.
- Mattingley, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. 1971. "Discrimination in Speech and Nonspeech Modes". *CP*, 2:131-137.
- McCauley, R. N. 1987. "The Role of Theories in a Theory of Concepts". In Neisser, U. (ed.). 1987. *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorisation*. New York: CUP, 288-309.
- McCawley, J. D. 1986. "Today the World, Tomorrow Phonology". *PHY*, 3:27-43.
- McClelland, J. L. 1981. "Retrieving General and Specific Information from Stored Knowledge of Specifics". *PACSS*, 3:170-172.
- McClelland, J. L., & Rumelhart, D. E. 1985. "Distributed Memory and the Representation of General and Specific Information". *JEP:GE*, 114:159-188.
- McCloskey, M. E. 1980. "The Stimulus Familiarity Problem in Semantic Memory Research". *JVLVB*, 19:485-502.
- McCloskey, M. E., & Glucksberg, S. 1978. "Natural Categories: Well-Defined or Fuzzy Sets?". *MC*, 6:462-472.
- McCloskey, M. E., & Glucksberg, S. 1979. "Decision Processes in Verifying Category Membership Statements: Implications for Models of Semantic Memory". *CP*, 11:1-37.
- McFarland, C. E., Duncan, E. M., & Kellas, G. 1978. "Isolating the Typicality Effect in Semantic Memory". *QJEP*, 30:251-262.
- McLaren, I. P. L., Leever, H. J., & Mackintosh, N. J. 1994. "Recognition, Categorisation, and Perceptual Learning (or, How Learning to Classify Things Together Helps One to Tell them Apart)". In Umiltà, C., & Moscovitch, M. (eds.). 1994. *Attention and Performance XV: Conscious and Unconscious Information Processing*. Cambridge, MA: MIT Press, 889-909.
- McNamara, T., & Sternberg, R. 1983. "Mental Models of Word Meaning". *JVLVB*, 22:449-474.
- McNeill, D., & Lindig, K. 1973. "The Perceptual Reality of Phonemes, Syllables, Words, and Sentences". *JVLVB*, 12:419-430.
- McQueen, J. M. 1991. "The Influence of the Lexicon on Phonetic Categorisation: Stimulus Quality in Word-Final Ambiguity". *JEP:HPP*, 17:433-443.
- Medin, D. L. 1983. "Structural Principles in Categorisation". In Tighe, T. J., & Shepp, B. E. (eds.). 1983. *Perception, Cognition, and Development: Interactional Analyses*. Hillsdale, NJ: LEA, 203-230.

- Medin, D. L. 1986. "Comment on 'Memory Storage and Retrieval Processes in Category Learning' ". *JEP:GE*, 115:373-381.
- Medin, D. L. 1988. "Social Categorisation: Structures, Processes, and Purposes". In Srull, T. K., & Wyer, R. S. (eds.). 1988. *Advances in Social Cognition*. Vol. 1. Hillsdale, NJ: LEA, 119-126.
- Medin, D. L. 1989. "Concepts and Conceptual Structure". *American Psychologist*, 44:1469-1481.
- Medin, D. L., & Barsalou, L. W. 1987. "Categorisation Processes and Categorical Perception". In Harnad, S. (ed.). 1987. *Categorical Perception: The Groundwork of Cognition*. New York: CUP, 455-490.
- Medin, D. L., & Heit, E. 1999. "Categorisation". In Bly, B. M., & Rumelhart, D. (eds.). 1999. *Handbook of Perception and Cognition: Cognitive Science*. Academic Press, 99-143.
- Medin, D. L., & Ortony, A. 1989. "Psychological Essentialism". In Vosniadou, S., & Ortony, A. (eds.). 1989. *Similarity and Analogical Reasoning*. New York: CUP, 179-195.
- Medin, D. L., & Schaffer, M. M. 1978. "Context Theory of Classification Learning". *PR*, 85:207-238.
- Medin, D. L., & Schwanenflugel, P. J. 1981. "Linear Separability in Classification Learning". *JEP:HLM*, 7:355-368.
- Medin, D. L., & Shoben, E. J. 1988. "Context and Structure in Conceptual Combination". *CP*, 20:158-190.
- Medin, D. L., & Smith, E. E. 1981. "Strategies and Classification Learning". *JEP:HLM*, 7:241-253.
- Medin, D. L., & Smith, E. E. 1984. "Concepts and Concept Formation". *ARP*, 35:113-138.
- Medin, D. L., & Wattenmaker, W. D. 1987. "Category Cohesiveness, Theories, and Cognitive Archaeology". In Neisser, U. (ed.). 1987. *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorisation*. Cambridge, England: CUP, 25-62.
- Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. 1982. "Correlated Symptoms and Simulated Medical Classification". *JEP:LMC*, 8:37-50.
- Medin, D. L., Altom, M. W., & Murphy, T. D. 1984. "Given Versus Induced Category Representations: Use of Prototype and Exemplar Information in Classification". *JEP:LMC*, 10:333-352.
- Medin, D. L., Dewey, G. I., & Murphy, T. D. 1983. "Relationships between Item and Category Learning: Evidence that Abstraction Is not Automatic". *JEP:LMC*, 9:607-625.
- Medin, D. L., Goldstone, R. L., & Gentner, D. 1993. "Respects for Similarity". *PR*, 100:254-278.
- Medin, D. L., Lynch, E. B., & Coley, J. D. 1997. "Categorisation and Reasoning Among Tree Experts: Do All Roads Lead to Rome?". *CP*, 32:49-96.
- Medin, D. L., Ross, B. H., & Markman, A. B. 2001. *Cognitive Psychology*. Harcourt College

- Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. 1987. "Family Resemblance, Conceptual Cohesiveness, and Category Construction". *CP*, 19:242-279.
- Mehler, J., Dommergues, J.-Y., Frauenfelder, U. H., & Segui, J. 1981. "The Syllable's Role in Speech Segmentation". *JVLVB*, 20:298-305.
- Melkman, R., Tversky, B., & Baratz, D. 1981. "Developmental Trends in the Use of Perceptual and Conceptual Attributes in Grouping, Clustering, and Retrieval". *JEC*, 31:470-486.
- Merriman, W. E., Schuster, M., & Hager, L. 1991. "Are Names Ever Mapped onto Preexisting Categories". *JEP:GE*, 120:288-300.
- Mervis, C. B. 1980. "Category Structure and the Development of Categorisation". In Spiro, R. J., Bruce, B. C., & Brewer, W. F. (eds.). 1980. *Theoretical Issues in Reading Comprehension*. Hillsdale, NJ: LEA, 279-307.
- Mervis, C. B. 1984. "Early Lexical Development: The Contributions of Mother and Child". In Sophian, C. (ed.). 1984. *Origins of Cognitive Skills*. Hillsdale, NJ: LEA, 339-370.
- Mervis, C. B. 1985. "On the Existence of Prelinguistic Categories: A Case Study". *IBD*, 8:293-300.
- Mervis, C. B. 1987. "Child-Basic Object Categories and Early Lexical Development". In Neisser, U. (ed.), 1987. *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorisation*. New York: CUP, 201-233.
- Mervis, C. B., Catlin, J., & Rosch, E. 1975. "Development of the Structure of Color Categories". *DP*, 11:54-60.
- Mervis, C. B., Catlin, J., & Rosch, E. 1976. "Relationships Among Goodness-of-Example, Category Norms, and Word Frequency". *BPS*, 7:283-284.
- Mervis, C. B., & Crisafi, M. A. 1982. "Order of Acquisition of Subordinate-, Basic- and Superordinate-Level Categories". *CD*, 53:258-266.
- Mervis, C. B., & Greco, C. 1984. "Parts and Early Conceptual Development: Comment on Tversky and Hemenway". *JEP:GE*, 113:194-197.
- Mervis, C. B., & Mervis, C. A. 1982. "Leopards Are Kitty-Cats: Object Labeling by Mothers for Their Thirteen-Month-Olds". *CD*, 53:267-273.
- Mervis, C. B., & Pani, J. R. 1980. "Acquisition of Basic Object Categories". *CP*, 12:496-522.
- Mervis, C. B., & Rosch, E. 1981. "Categorisation of Natural Objects". *ARP*, 32:89-115.
- Mervis, C. B., & Roth, E. M. 1981. "The Internal Structure of Basic and Non-Basic Color Categories". *Lg*, 57:384-405.
- Michalski, R. S., 1993. "Beyond Prototypes and Frames: The Two-Tiered Concept Representation". In Van Mechelen, L., Hampton, J., Michalski, R. S., & Theuns, P. (eds.). 1993. *Categories and Concepts: Theoretical Views and Inductive Data Analysis*. New York: Academic Press, 145-172.



- Miller, G. A., & Nicely, P. E. 1955. "An Analysis of Perceptual Confusions Among Some English Consonants". *JASA*, 27:338-352.
- Miller, J. D., Henderson, B. C., Sullivan, H. T., & Ridgen, G. K. 1978. "Speech Perception by the Chinchilla: Learning Functions for Pairs of Synthetic Stimuli from Various Points Along a VOT Continuum". *JASA*, Suppl. 1. 64, S18. Kuhl81
- Miller, J. L. 1977. "Properties of Feature Detectors for VOT: The Voiceless Channel of Analysis". *JASA*, 62:641-648.
- Miller, J. L. 1994. "On the Internal Structure of Phonetic Categories: a Progress Report". *CG*, 50:271-285.
- Miller, J. L., Connine, C. M., Schermer, T. M., & Kluender, K. R. 1983. "A Possible Auditory Basis for Internal Structure of Phonetic Categories". *JASA*, 73:2124-2133.
- Miller, J. L., & Eimas, P. D. 1983. "Studies on the Categorisation of Speech by Infants". *CG*, 13:135-165.
- Miller, J. L., & Eimas, P. D. 1996. "Internal Structure of Voicing Categories in Early Infancy". *PP*, 58:1157-1167.
- Miller, J. L., & Volaitis, L. E. 1989. "Effect of Speaking Rate on the Perceptual Structure of a Phonetic Category". *PP*, 46:505-512.
- Miller, J. L., O'Rourke, T. B., & Volaitis, L. E. 1997. "Internal Structure of Phonetic Categories: Effects of Speaking Rate". *PHN*, 54:121-137.
- ?Miller, R. 1973. "The Use of Concrete and Abstract Concepts by Children and Adults". *CG*, 2:49-58.
- Mills, C. B. 1980. "Effects of Context on Reaction Time to Phonemes". *JVLVB*, 19:75-83.
- Mohanan, K. P. 1986. *The Theory of Lexical Phonology*. Dordrecht: D. Reidel.
- Moiseff, A., Pollack, G. S., & Hoy, R. R. 1978. "Steering Responses to Flying Crickets to Sound and Ultrasound: Mate Attraction and Predator Avoidance". *Proceedings of the National Academy of Sciences*, 75:4052-4056.

- Mompeán-González, J. A. 1999. "A Cognitive Account of the Concept of the Phoneme". Paper delivered at the VI International Cognitive Linguistics Conference. University of Stockholm, July 1999.
- Mompeán-González, J. A. 2001a. "A Comparison Between English and Spanish Subjects' Typicality Ratings in Phoneme Categories: A First Report". *International Journal of English Studies*, 115-155.
- Mompeán-González, J. A. 2001b. "The Need for Psycholinguistic Evidence in Cognitive Phonology". Paper delivered at the VII International Cognitive Linguistics Conference. University of California, Santa Barbara. July 2001.
- Mompeán-González, J. A. In press. "Category Overlap and Neutralisation: The Importance of Speakers' Classification in Phonology". To appear in *Cognitive Linguistics*.
- Mompeán-González, J. A., & Hernández-Campoy, J. M. 2001. "Advantages and Disadvantages of RP as an EFL Model of Pronunciation" In de la Cruz, I., Santamaría, C., Tejedor, C., & Valero, C. (eds.). 2001. *La Lingüística Aplicada a Finales del Siglo XX. Ensayos y Propuestas*. Vol. 2. Alcalá de Henares: Universidad de Alcalá, 707-713
- Monroy-Casas, R. 1998. *Sistemas de Transcripción Fonética del Inglés: Teoría y Textos*. Granada: Grupo Editorial Universitario.
- Morgan, M. J., Fitch, M. D., Holman, J. G., & Lea, S. E. G. 1976. "Pigeons Learn the Concept of an 'A' ". *Perception*, 5:57-66.
- Morris, M. W., & Murphy, G. L. 1990. "Converging Operations on a Basic Level in Event Taxonomies". *MC*, 18:407-418.
- Morse, P. A., & Snowdown, C. T. 1975. "An Investigation of Categorical Speech Discrimination by Rhesus Monkeys". *PP*, 17:9-16.
- Morton, J., & Long, J. 1976. "Effect of Word Transitional Probability on Phoneme Identification". *JVLVB*, 15:43-51.
- Moskowitz, B. A. 1973. "On the Status of the Vowel Shift in English". In Moore, T. E. (ed.) 1973. *Cognitive Development and the Acquisition of Language*. New York: Academic Press, 223-260.
- Mulford, R. 1979. "Prototypicality and the Development of Categorisation". *Papers and Reports on Child Language Development*, 16:13-25.
- Murphy, G. L. 1982. "Cue Validity and Levels of Categorisation". *PB*, 91:174-177.
- Murphy, G. L. 1988. "Comprehending Complex Concepts". *CS*, 12:529-562.
- Murphy, G. L. 1990. "Noun Phrase Interpretation and Conceptual Combination". *JML*, 29:259-288.
- Murphy, G. L. 1991a. "Parts in Object Concepts: Experiments with Artificial Categories". *MC*, 19:423-438.
- Murphy, G. L. 1991b. "More on Parts in Object Concepts: Response to Tversky and Hemenway". *MC*, 19:443-447.



- Murphy, G. L. 1993a. "A Rational Theory of Concepts". In Nakamura, G. V., Taraban, R. M., & Medin, D. L. (eds.). 1993. *The Psychology of Learning and Motivation. Vol. 29: Categorisation by Humans and Machines*. New York: Academic Press, 327-259.
- Murphy, G. L. 1993b. "Theories and Concept Formation". In Van Mechelen, L., Hampton, J., Michalski, R. S., & Theuns, P. (eds.). 1993. *Categories and Concepts: Theoretical Views and Inductive Data Analysis*. New York: Academic Press, 173-200.
- Murphy, G. L., & Allopenna, P. D. 1994. "The Locus of Knowledge Effects in Concept Learning". *JEP:LMC*, 20:904-919.
- Murphy, G. L., & Brownell, H. H. 1985. "Category Differentiation in Object Recognition: Typicality Constraints on the Basic Category Advantage". *JEP:LMC*, 11:70-84.
- Murphy, G. L., & Lassaline, M. E. 1997. "Hierarchical Structure in Concepts and the Basic Level of Categorisation". In Lamberts, K., & Shanks, D. (eds.). 1997. *Knowledge, Concepts, and Categories*. Cambridge, MA: MIT Press, 93-131.
- Murphy, G. L., & Medin, D. L. 1985. "The Role of Theories in Conceptual Coherence". *PR*, 92:289-316.
- Murphy, G. L., & Ross, B. H. 1994. "Predictions from Uncertain Categorisations". *CP*, 27:148-193.
- Murphy, G. L., & Ross, B. H. 1999. "Induction with Cross-Classified Categories". *MC*, 27:1024-1041.
- Murphy, G. L., & Smith, E. E. 1982. "Basic-Level Superiority in Picture Categorisation". *JVLVB*, 21:1-20.
- Murphy, G. L., & Wisniewski, E. J. 1989a. "Categorizing Objects in Isolation and in Scenes: What a Superordinate Is Good For". *JEP:LMC*, 15:572-586.
- Murphy, G. L., & Wisniewski, E. J. 1989b. "Feature Correlations in Conceptual Representations". In Tiberghien, G. (ed.). 1989. *Advances in Cognitive Science. Vol. 2: Theory and Applications*. Chichester, England: Ellis Horwood, 23-45.
- Murphy, G. L., & Wright, J. C. 1984. "Changes in Conceptual Structure with Expertise: Differences between Real-World Experts and Novices". *JEP:LMC*, 10:144-155.
- Myerson, R. F. 1976. "Children's Knowledge of Selected Aspects of 'Sound Pattern of English'". In Campbell, R. N., & Smith, P. T. (eds.). *Recent Advances in the Psychology of Language: Formal and Experimental Approaches*. New York: Plenum Press, 377-402.
- Nakamura, G. V. 1985. "Knowledge-Based Classification of Ill-Defined Categories". *MC*, 13:377-384.
- Nathan, G. S. 1986. "Phonemes as Mental Categories". *BLS*, 12:212-223.
- Nathan, G. S. 1989. "Preliminaries to a Theory of Phonological Substance: the Substance of Sonority". In Corrigan, R., Eckman, F., & Noonan, M. (eds.) 1989. *Linguistic Categorisation*. Amsterdam/Philadelphia: John Benjamins, 55-67.
- Nathan, G. S. 1994. "How the Phoneme Inventory Gets Its Shape: Cognitive Grammar's View of Phonological Systems". *Rivista di Linguistica*, 6:275-287.

- Nathan, G. S. 1996. "Steps Towards a Cognitive Phonology". In Hurch, B., & Rhodes, R. (eds.). 1996. *Natural Phonology: The State of the Art*. Berlin: Mouton de Gruyter, 107-120.
- Nathan, G. S. 1999. "What Functionalists Can Learn from Formalists in Phonology". In Darnell, M., Moravcsik, E., Newmeyer, F., Noonan, M., & Wheatley, K. (eds.). 1999. *Functionalism and Formalism in Linguistics: Volume I: General Papers*. Amsterdam/Philadelphia: John Benjamins, 305-327.
- Nearey, T. M. 1981. "The Psychological Reality of Phonological Representations: Experimental Evidence". In Myers, T., Laver, J., & Anderson, J. (eds.). 1981. *The Cognitive Representation of Speech*. Amsterdam: North-Holland, 359-369.
- Neisser, U. 1979. "The Concept of Intelligence". *Intelligence*, 3:217-227.
- Neisser, U. 1987a. "Introduction: The Ecological and Intellectual Bases of Categorisation". In Neisser, U. (ed.). 1987. *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorisation*. New York: CUP, 1-10.
- Neisser, U. 1987b. "From Direct Perception to Conceptual Structure". In Neisser, U. (ed.). 1987. *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorisation*. New York: CUP, 11-24.
- Neisser, U., & Weene, P. 1962. "Hierarchies in Concept Attainment". *JEP*, 64:640-645.
- Nelson, K. 1973. "Some Evidence for the Cognitive Primacy of Categorisation and Its Functional Basis". *MPQBD*, 19:21-39.
- Nelson, K. 1974a. "Concept, Word, and Sentence: Interrelations in Acquisition and Development". *PR*, 81:267-285.
- Nelson, K. 1974b. "Variations in Children's Concepts by Age and Category". *CD*, 45:577-584.
- Nelson, K. 1983. "The Derivation of Concepts and Categories from Event Representations". In Scholnick, E. K. (ed.). 1983. *New Trends in Conceptual Representation: Challenges to Piaget's Theory?*. Hillsdale, NJ: LEA, 129-150.
- Nelson, K., Rescolda, L., Gruendel, J., & Benedict, H. 1978. "Early Lexicons: What Do They Mean?". *CD*, 49:960-968.
- Nelson, L. J., & Miller, D. T. 1995. "The Distinctiveness Effect in Social Categorisation: You Are What Makes You Unusual". *PSS*, 6:246-249.
- Neumann, P. G. 1974. "An Attribute Frequency Model for the Abstraction of Prototypes". *MC*, 2:241-248.
- Neumann, P. G. 1977. "Visual Prototype Formation with Discontinuous Representation of Dimensions of Variability". *MC*, 5:187-197.
- Newman, J. E., & Dell, G. S. 1978. "The Phonological Nature of Phoneme Monitoring: A Critique of Some Ambiguity Studies". *JVLVB*, 17:359-374.
- Newman, R. S., Sawusch, J. R., Luce, P. A. 1997. "Lexical Neighborhood Effects in Phonetic Processing". *JEP:HPP*, 23:873-889.



- Newport, E. L., & Bellugi, U. 1978. "Linguistic Expression of Category Levels in a Visual-Gestural Language: A Flower Is a Flower Is a Flower". In Rosch, E., & Lloyd, B. B. (eds.). 1978. *Cognition and Categorisation*. Hillsdale, NJ: LEA, 49-71.
- Niedenthal, P. M., & Halberstadt, J. B. 1995. "The Acquisition and Structure of Emotional Response Categories". In Medin, D. L. (ed.). 1995. *The Psychology of Learning and Motivation*. Vol. 33. New York: Academic Press, 23-64.
- Niedenthal, P. M., Halberstadt, J. B., & Innes-Ker, A. H. 1999. "Emotional Response Categorisation". *PR*, 106:337-361.
- Nosofsky, R. M. 1984. "Choice, Similarity, and the Context Theory of Classification". *JEP:LMC*, 10:104-114.
- Nosofsky, R. M. 1986. "Attention, Similarity, and the Identification-Categorisation Relationship". *JEP:GE*, 115:39-57.
- Nosofsky, R. M. 1987. "Attention and Learning Processes in the Identification and Categorisation of Integral Stimuli". *JEP:LMC*, 13:87-108.
- Nosofsky, R. M. 1988a. "Exemplar-Based Accounts of Relations Between Classification, Recognition, and Typicality". *JEP:LMC*, 14:700-708.
- Nosofsky, R. M. 1988b. "Similarity, Frequency, and Category Representations". *JEP:LMC*, 14:54-65.
- Nosofsky, R. M. 1991a. "Tests of an Exemplar Model for Relating Perceptual Classification and Recognition Memory". *JEP:HPP*, 17:3-27.
- Nosofsky, R. M. 1991b. "Relation between the Rational Model and the Context Model of Categorisation". *PSS*, 2:416-421.
- Nosofsky, R. M. 1991c. "Exemplars, Prototypes, and Similarity Rules". In Healy, A. M., Kosslyn, S. F., & Shiffrin, R. M. (eds.). 1991. *From Learning Theory to Connectionist Theory: Essays in Honor of W. K. Estes*. Vol. 1. Hillsdale, NJ: LEA, 149-167.
- Nosofsky, R. M. 1991d. "Typicality in Logically Defined Categories: Exemplar-Similarity versus Rule Instantiations". *MC*, 19:131-150.
- Nosofsky, R. M., Clark, S. E., & Shin, H. J. 1989. "Rules and Exemplars in Categorisation, Identification, and Recognition". *JEP:LMC*, 15:282-304.
- Oakes, L. M., Madole, K. L., & Cohen, L. B. 1991. "Infants' Object Examining: Habituation and Categorisation". *CGD*, 6:377-392.
- O'Connor, J. D. 1973. *Phonetics*. Harmondsworth, Penguin Books.
- O'Connor, J. D., & Trim, J. L. M. "Vowel, Consonant and Syllable -A Phonological Definition". *Word*, 9:103-122.
- Oden, G. C. 1977. "Fuzziness in Semantic Memory: Choosing Exemplars of Subjective Categories". *MC*, 5:198-204.
- Oden, G. C. 1987. "Concept, Knowledge, and Thought". *ARP*, 38:203-227.

- Oden, G. C., & Massaro, D. W. 1978. "Integration of Featural Information in Speech Perception". *PR*, 85:172-191.
- Ohala, J. J. 1974, "Experimental Historical Phonology". In Anderson, J. M., & Jones, C. (eds.). 1974. *Historical Linguistics*. Vol. 2. Amsterdam: North-Holland, 353-389.
- Ohala, J. J. 1983. "The Phonological End Justifies Any Means". In Hattori, S., & Inoue, K. (eds.). *Proceedings of the XIIIth International Congress of Linguists*. Tokyo: Tokyo Press, 232-243.
- Ohala, J. J. 1986. "Consumer's Guide to Evidence in Phonology". *PHY* 3:3-26.
- Ohala, J. J. 1987. "Explanation in Phonology: Opinions and Examples". In Dressler, W. U., Luschützky, H.C., Pfeiffer, O.E., & Rennison, J.R. (eds.). 1987. *Phonologica 1984*. London: CUP, 215-225.
- Ohala, J. J. 1990. "There is No Interface Between Phonology and Phonetics: A Personal View". *JP*, 18:153-171.
- Ohala, J. J. 1995. "Experimental Phonology". In Goldsmith, J. A. (ed.). 1995. *The Handbook of Phonological Theory*. Oxford, UK: Blackwell, 713-722.
- Ohala, J. J., & Jaeger, J. J. 1986a. "Introduction". In Ohala, J. J., & Jaeger, J. J. (eds.). 1986. *Experimental Phonology*. Orlando: Academic Press, 1-12.
- Ohala, J. J., & Jaeger, J. J. 1986b. (eds.). *Experimental Phonology*. Orlando, FL: Academic Press.
- Ohala, J. J., & Ohala, M. 1986. "Testing Hypotheses Regarding the Psychological Manifestation of Morpheme Structure Constraints". In Ohala, J. J., & Jaeger, J. J. (eds.). 1986. *Experimental Phonology*. Orlando: Academic Press, 239-252.
- Ohala, M., & Ohala, J. J. 1987. "Psycholinguistic Probes of Native Speakers' Phonological Knowledge". In Dressler, W. U., Luschützky, H. C., Pfeiffer, O. E., & Rennison, J. R. (eds.). 1987. *Phonologica 1984*. Cambridge: Cambridge University Press, 227-233.
- Ohde, R. N. 1984. "Fundamental Frequency As an Acoustic Correlate of Stop Consonant Voicing". *JASA*, 75:224-230.
- Omohundro, J. 1981. "Recognition vs. Classification of Ill-Defined Category Exemplars". *MC*, 9:324-331.
- Osherson, D. N., & Smith, E. E. 1981. "On the Adequacy of Prototype Theory As a Theory of Concepts". *CG*, 9:35-58.
- Osherson, D. N., & Smith, E. E. 1982. "Gradedness and Conceptual Combination". *CG*, 12:299-318.
- Osherson, D. N., & Smith, E. E. 1997. "On Typicality and Vagueness". *CG*, 64:189-206.
- Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. B. 1990. "Category-Based Induction". *PR*, 97:185-200.
- Palmer, S. E. 1975. "The Effects of Contextual Scenes on the Identification of Objects". *MC*, 3:519-526.



- Pazzani, M. J. 1991. "Influence of Prior Knowledge on Concept Acquisition: Experimental and Computational Results". *JEP:LMC*, 17:416-432.
- Pearce, J. M. 1988. "Stimulus Generalization and the Acquisition of Categories by Pigeons". In Weiskrantz, L. (ed.). 1988. *Thought Without Language*. Oxford, England: Clarendon Press, 132-155.
- Pearce, J. M. 1989. "The Acquisition of an Artificial Category by Pigeons". *QJEP:CPP*, 41:381-406.
- Peppenberg, I. M. 1987. "Acquisition of the Same/Difference Concept by an African Grey Parrot (*Psittacus Erithacus*): Learning with Respect to Categories of Color, Shape, and Material". *ALB*, 15:423-432.
- Pierrehumbert, J. B. 2000. "What People Known About Sounds of Language". *Studies in the Linguistic Sciences*, 29:111-120.
- Pierrehumbert, J. B., Beckman, M. E., & Ladd, D. R. 1996 "Laboratory Phonology". In Durand, J., & Laks, B. (eds.). 1996. *Current Trends in Phonology: Models and Methods*. CNRS, Paris-X and University of Salford: University of Salford Press, 535-548.
- Pierrehumbert, J. B., Beckman, M. E., & Ladd, D. R. 2000. "Conceptual Foundations of Phonology as a Laboratory Science". In Burton-Roberts, N., Carr, P., & Docherty, G. (eds.). 2000. *Phonological Knowledge: Its Nature and Status*. Oxford, UK: Oxford University Press, 273-304.
- Pietrewicz, A. T., & Kamil, A. C. 1977. "Visual Detection of Cryptic Prey by Blue Jays (*Cyanocitta Cristata*)". *Science*, 195:580-582.
- Pikas, A. 1966. *Abstraction and Concept Formation*. Cambridge, MA: Harvard University Press.
- Pike, K. L. 1943. *Phonetics: A Critical Analysis of Theory and a Technic for the Practical Description of Sounds*. Ann Arbor, MI: The University of Michigan Press.
- Pike, K. L. 1947. *Phonemics: A Technique for Reducing Language to Writing*. Ann Arbor, MI: The University of Michigan Press.
- Pisoni, D. B., & Tash, J. 1974. "Reaction Times to Comparisons Within and Across Phonetic Categories". *PP*, 15:285-290.
- Pitt, M. A., & Samuel, A. G. 1993. "An Empirical and Meta-Analytic Evaluation of the Phoneme Identification Task". *JEP:HPP*, 19:699-725.
- Poole, J., & Lander, D. G. 1971. "The Pigeon's Concept of Pigeon". *PS*, 25:157-158.
- Posnansky, C. J., & Neumann, P. G. 1976. "The Abstraction of Visual Prototypes by Children". *JECP*, 21:367-379.
- Posner, M. I. 1969. "Abstraction and the Process of Recognition". In Bower, G. H., & Spence, J. T. (eds.). 1969. *The Psychology of Learning and Motivation*. Vol. 3. New York: Academic Press, 43-100.

- Posner, M. I. 1986. "Empirical Studies of Prototypes". In Craig, C. (ed.). 1986. *Noun Classes and Categorisation*. Amsterdam: John Benjamins, 53-61.
- Posner, M. I., & Keele, S. W. 1968. "On the Genesis of Abstract Ideas". *JEP*, 77:353-363.
- Posner, M. I., & Keele, S. W. 1970. "Retention of Abstract Ideas". *JEP*, 83:304-308
- Posner, M. I., Goldsmith, R., & Welton, K. E., Jr. 1967. "Perceived Distance and the Classification of Distorted Patterns". *JEP*, 73:28-38.
- Poulin-Dubois, D., Graham, S. & Sippola, L. 1995. "Early Lexical Development: The Contribution of Parental Labelling and Infants' Categorisation Abilities". *JCL*, 22:325-343.
- Quinn, P. C. 1987. "The Categorical Representation of Visual Pattern Information by Young Infants". *CG*, 27:147-179.
- Quinn, P. C. 1994. "The Categorisation of *Above* and *Below* Spatial Relations by Young Infants". *CD*, 65:58-69.
- Quinn, P. C., & Bomba, P. C. 1986. "Evidence for a General Category of Oblique Orientations in Four-Month-Old Infants". *JECP*, 42:345-354.
- Quinn, P. C., & Eimas, P. D. 1996. "Perceptual Cues that Permit Categorical Differentiation of Animal Species by Infants". *JECP*, 63:189-211.
- Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. 1993. "Evidence for Representations of Perceptually Similar Natural Categories by 3-Month-Old and 4-Month-Old Infants". *Perception*, 22:463-475.
- Quinn, P. C., & Johnson, M. H. 1997. "The Emergence of Perceptual Category Representations in Young Infants: A Connectionist Analysis". *JECP*, 66:236-263.
- Rakison, D. H., & Butterworth, G. E. 1998. "Infants' Use of Object Parts in Early Categorisation". *DP*, 34:49-62.
- Read, C. 1971. "Pre-School Children's Knowledge of English Phonology". *Harvard Educational Review*, 41:1-34.
- Read, C. 1975. "Children's Categorisation of Speech Sounds in English". NCTE Research Report No. 17. Urbana, IL: National Council of Teachers of English.
- Read, C. 1986. *Children's Creative Spelling*. London: Routledge & Kegan Paul.
- Read, S. J., Jones, D. K., & Miller, L. C. 1990. "Traits as Goal-Based Categories: The Importance of Goals in the Coherence of Dispositional Categories". *JPSP*, 58:1048-1061.
- Read, S. J., Miller, L. C., & Jones, D. K. 1990. "Goals in the Conceptual Coherence of Social Categories". *BPS*, 28:261-267.
- Reber, A. S. 1989. "Implicit Learning and Tacit Knowledge". *JEP:GE*, 3:219-235.
- Reber, A. S., & Allen, R. 1978. "Analogic and Abstraction Strategies in Synthetic Grammar Learning: A Functionalist Interpretation". *CG*, 6:189-221.
- Reed, S. K. 1972. "Pattern Recognition and Categorisation". *CP*, 3:382-407.

- Reed, S. K. 1978. "Category vs. Item Learning: Implications for Categorisation Models". *MC*, 6:612-621.
- Reed, S. K., & Friedman, M. P. 1973. "Perceptual vs. Conceptual Categorisation". *MC*, 1:157-163.
- Reeds, J. A., & Wang, W. S-Y. 1961. "The Perception of Stops after s". *PHN*, 6:78-81.
- Repp, B. H. 1977. "Dichotic Competition of Speech Sounds: The Role of Acoustic Stimulus Structure". *JEP:HPP*, 3:37-50.
- Repp, B. H. 1984. "Categorical Perception: Issues, Methods, Findings". In Lass, N. J. (ed.). 1984. *Speech and Language: Advances in Basic Research and Practice*. Vol. 10. New York: Academic Press, 243-335.
- Repp, B. H., & Liberman, A. M. 1987. "Phonetic Category Boundaries Are Flexible". In Harnad, S. (ed.). 1987. *Categorical Perception: The Groundwork of Cognition*. New York: CUP, 89-112.
- Rescolda, L. A. 1980. "Overextension in Early Language Development". *JCL*, 7:321-335.
- Rey, G. 1983. "Concepts and Stereotypes". *CG*, 15:237-262.
- Rey, G. 1985. "Concepts and Conceptions: A Reply to Smith, Medin, and Rips". *CG*, 19:297-303.
- Rifkin, A. 1985. "Evidence for a Basic Level in Event Taxonomies". *MC*, 13:538-556.
- Rips, L. J. 1975. "Inductive Judgments About Natural Categories". *JVLVB*, 14:665-681.
- Rips, L. J. 1989. "Similarity, Typicality, and Categorisation". In Vosniadou, S., & Ortony, A. (eds.). 1989. *Similarity and Analogical Reasoning*. New York: CUP, 19-59.
- Rips, L. J. 1991. "Similarity and the Structure of Categories". In Napoli, D. J., & Kegl, J. A. (eds.). 1991. *Bridges Between Psychology and Linguistics: A Swarthmore Festschrift for Lila Gleitman*. Hillsdale, NJ: LEA, 35-53.
- Rips, L. J. 1995. "The Current State of Research on Concept Combination". *Mind and Language*, 10:72-104.
- Rips, L. J., & Collins, A. 1993. "Categories and Resemblance". *JEP:GE*, 122:468-486.
- Rips, L. J., Shoben, E. J., & Smith, E. E. 1973. "Semantic Distance and the Verification of Semantic Relations". *JVLVB*, 12:1-20.
- Roach, P. 1983. *English Phonetics and Phonology*. Cambridge, England: CUP.
- Robbins, D., Barresi, J., Compton, P., Furst, A., Russo, M., & Smith, M. A. 1978. "The Genesis and Use of Exemplar vs. Prototype Knowledge in Abstract Category Learning". *MC*, 6:473-480.
- Roberts, K. 1988. "Retrieval of a Basic-Level Category in Prelinguistic Infants". *DP*, 24:21-27.
- Roberts, K., & Cuff, M. D. 1989. "Categorisation Studies of 9- to 15-Month-Old Infants: Evidence for Superordinate Categorisation". *IBD*, 12:265-288.

- Roberts, K., & Horowitz, F. D. 1986. "Basic-Level Categorisation in Seven- and Nine-Month-Old Infants". *JCL*, 13:191-208.
- Roberts, W. A., & Mazmanian, D. S. 1988. "Concept Learning at Different Levels of Abstraction by Pigeons, Monkeys, and People". *JEP:ABP*, 14:247-260.
- Rosch, E. H. 1973a. "Natural Categories". *CP*, 4:328-350.
- Rosch, E. H. 1973b. "On the Internal Structure of Perceptual and Semantic Categories". In Moore, T. E. (ed.). 1973. *Cognitive Development and the Acquisition of Language*. New York: Academic Press, 111-144.
- Rosch, E. H. 1974. "Linguistic Relativity". In Silverstein, A. (ed.). 1974. *Human Communication: Theoretical Perspectives*. Hillsdale, NJ: LEA, 94-121.
- Rosch, E. H. 1975a. "Cognitive Reference Points". *CP*, 7:532-547.
- Rosch, E. H. 1975b. "Cognitive Representations of Semantic Categories". *JEP:GE*, 104:192-233.
- Rosch, E. H. 1975c. "The Nature of Mental Codes for Color Categories". *JEP:HPP*, 1:303-322.
- Rosch, E. H. 1975d. "Universals and Cultural Specifics in Human Categorisation". In Brislin, R. W., Bochner, S., & Lonner, W. J. (eds.). 1975. *Cross-Cultural Perspectives on Learning*. New York: Halsted Press, 177-206.
- Rosch, E. H. 1976. "Classifications of Real-World Objects: Origins and Representations in Cognition". *Bulletin de Psychologie*. Special Annual 242-250.
- Rosch, E. H. 1977. "Human Categorisation". In Warren, N. (ed.). 1977. *Advances (Studies) in Cross-Cultural Psychology*. Vol. 1. London: Academic Press, 3-49.
- Rosch, E. H. 1978. "Principles of Categorisation". In Rosch, E. H., & Lloyd, B. B. (eds.). 1978. *Cognition and Categorisation*. Hillsdale, NJ: LEA, 27-48.
- Rosch, E. H. 1983. "Prototype Classification and Logical Classification: The Two Systems". In Scholnick, E. K. (ed.). 1983. *New Trends in Conceptual Representation: Challenges to Piaget's Theory?*. Hillsdale, NJ: LEA, 73-86.
- Rosch, E. H. 1988. "Coherences and Categorisation: A Historical View". In Kessel, F. S. (ed.). 1988. *The Development of Language and Language Researchers: Essays in Honor of Roger Brown*. Hillsdale, NJ: LEA, 373-392.
- Rosch, E. H., & Mervis, C. B. 1975. "Family Resemblances: Studies in the Internal Structure of Categories". *CP*, 7:573-605.
- Rosch, E. H., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. 1976a. "Basic Objects in Natural Categories". *CP*, 8:382-439.
- Rosch, E. H., Simpson, C., & Miller, R. S. 1976b. "Structural Bases of Typicality Effects". *JEP:HPP*, 2:491-502.
- Rosen, S., & Howell, P. 1981. "Plucks and Bows Are Not Categorically Perceived". *PP*, 30:156-168.



- Rosen, S., & Howell, P. 1983. "Sinusoidal Plucks and Bows Are not Categorically Perceived, Either". *PP*, 34:233-236.
- Rosner, S. R., & Hayes, D. S. 1977. "A Developmental Study of Category Item Production". *CD*, 48:1062-1065.
- Ross, B. H., & Murphy, G. L. 1996. "Category-Based Predictions: The Influence of Uncertainty and Feature Associations". *JEP:LMC*, 22:736-753.
- Ross, B. H., & Murphy, G. L. 1999. "Food for Thought: Cross-Classification and Category Organization in a Complex Real-World Category". *CP*, 38:495-553.
- Ross, G. S. 1980. "Categorisation in 1- to 2-Year-Olds". *DP*, 16:391-396.
- Roth, E. M., & Mervis, C. B. 1983. "Fuzzy-Set Theory and Class Inclusion Relations in Semantic Categories". *JVLVB*, 22:509-525.
- Roth, E. M., & Shoben, E. J. 1983. "The Effect of Context on the Structure of Categories". *CP*, 15:346-378.
- Rubin, P., Turvey, M. T., & Van Gelder, P. 1976. "Initial Phonemes are Detected Faster in Spoken Words than in Spoken Nonwords". *PP*, 19:394-398.
- Ryan, C. M. E. 1982. "Concept Formation and Individual Recognition in the Domestic Chicken (*Gallus gallus*)". *Behavior Analysis Letters*, 2:213-220.
- Ryan, C. M. E., & Lea, S. E. G. 1994. "Images of Conspecifics As Categories to be Discriminated by Pigeons and Chickens: Slides, Video Tapes, Stuffed Birds and Live Birds". *Behavioural Processes*, 33:155-176. Cook97.
- Ryle, G. 1949. *The Concept of Mind*. London: Hutchinson's University Library.
- Ryle, G. 1951. "Thinking and Language". *Proceedings of the Aristotelian Society. Supplementary Series*. 25:65-82.
- Saltz, E., Dixon, D., Klein, S., & Becker, G. 1977. "Studies of Natural Language Concepts. III. Concept Overdiscrimination in Comprehension between Two and Four Years of Age". *CD*, 48:1682-1685.
- Saltz, E., Soller, E., & Sigel, I. E. 1972. "The Development of Natural Language Concepts". *CD*, 43:1192-1202.
- Samuel, A. G. 1982. "Phonetic Prototypes". *PP*, 31:307-314.
- Sandell, J. H., Gross, C. G., & Bornstein, M. H. 1979. "Color Categories in Macaques". *JCPP*, 93:626-635.
- Sands, S. F., Lincoln, C. E., & Wright, A. A. 1982. "Pictorial Similarity Judgments and the Organization of Visual Memory in the Rhesus Monkey". *JEP:GE*, 111:369-389.
- Santiago, H. C., & Wright, A. C. 1984. "Pigeon Memory: *Same/Different* Concept Learning, Serial Probe Recognition Acquisition, and Probe Delay Effects on the Serial-Position Function". *JEP:ABP*, 10:498-512.

- Sapir, E. 1921. *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace & World. Inc.
- Sapir, E. 1933. "La Réalité Psychologique Des Phonèmes". *Journal de Psychologie Normale et Pathologique*, 30:247-265.
- Savin, H. B., & Bever, T. G. 1970. "The Nonperceptual Reality of the Phoneme". *JVLVB*, 9:295-302.
- Sawusch, J. R., & Jusczyk, P. 1981. "Adaptation and Contrast in the Perception of Voicing". *JEP:HPP*, 7:408-421.
- Schane, S. A. 1968. "On the Non-Uniqueness of Phonological Representations". *Lg*, 44:709-716.
- Schane, S. A. 1971. "The Phoneme Revisited". *Lg*, 47:503-521
- Schank, R. C., Collins, G. C., & Hunter, L. E. 1986. "Transcending Inductive Category Formation in Learning". *BBS*, 9:639-651.
- Schlesinger, I. M. 1982. *Steps to Language. Towards a Theory of Native Language Acquisition*. Hillsdale, NJ: LEA.
- Scholnick, E. K. 1983. "Classes, Collections, and other Connections". In Scholnick, E. K. (ed.). 1983. *New Trends in Conceptual Representation: Challenges to Piaget's Theory?*. Hillsdale, NJ: LEA, 185-197.
- Schrier, A. M., Angarella, R., & Povar, M. L. 1984. "Studies of Concept Formation by Stumptailed Monkeys: Concepts Humans, Monkeys and Letter A". *JEP:ABP*, 4:564-584.
- Schrier, A. M., & Brady, P. M. 1987. "Categorisation of Natural Stimuli by Monkeys (*Macaca Mulatta*): Effects of Stimulus Set Size and Modification of Exemplars". *JEP:ABP*, 13:136-143.
- Schwanenflugel, P. J., & Rey, M. 1986. "The Relationship Between Category Typicality and Concept Familiarity: Evidence from Spanish- and English-Speaking Monolinguals". *MC*, 14:150-163.
- Segalowitz, N., & Poulin-Dubois, D. 1990. "The Structure of Categories: Typicality Gradients, Perceived Linguistic Familiarity and Cross-Linguistic Comparisons". *CPC*, 10:491-512.
- Segui, J., Frauenfelder, U., & Mehler, J. 1981. "Phoneme Monitoring and Lexical Access". *BJP*, 72:471-477.
- Selkirk, E. O. 1982. "The Syllable". In Van der Hulst, H., & Smith, N. (eds.). 1982. *The Structure of Phonological Representations, Part II*. Dordrecht, Holland: Foris, 337-383.
- Shafir, E. B., Smith, E. E., & Osherson, D. N. 1990. "Typicality and Reasoning Fallacies". *MC*, 18:229-239.
- Shanks, D. R. 1991. "Categorisation by a Connectionist Network". *JEP:LMC*, 17:433-443.
- Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. 1987. "Emotion Knowledge: Further Exploration of a Prototype Approach". *JPSP*, 52:1061-1086.
- Sherman, T. L. 1985. "Categorisation Skills in Infants". *CD*, 56:1561-1573.



Shipley, E. F. 1993. "Categories, Hierarchies, and Induction". In Medin, D. L. (ed.). 1993. *The Psychology of Learning and Motivation*. Vol. 30. San Diego: Academic Press, 265-301.

Shipley, E. F., Kuhn, I. F., & Madden, E. C. 1983. "Mothers' Use of Superordinate Category Terms". *JCL*, 10:571-588.

Shyan, M. R., Wright, A. A., Cook, R. G., & Jitsumori, M. 1987. "Acquisition of the Auditory Same/Different Task in a Rhesus Monkey". *PS*, 25: 1-4.

Siegel, R. K., & Honig, W. K. 1970. "Pigeon Concept Formation: Successive and Simultaneous Acquisition". *JEAB*, 13:385-390.

Sinnott, J. M., Beecher, M. D., Moody, D. B., & Stebbins, W. C. 1976. "Speech Sound Discrimination by Monkeys and Humans". *JASA*, 60:687-695.

Skousen, R. 1975. *Substantive Evidence In Phonology*. Jauna Linguarum, Series minor 217. The Hague: Mouton.

Skousen, R. 1982. "English Spelling and Phonemic Representation". *Visual Language*, 16:28-38.

Sloman, S. A. 1997. "Categorical Inference Is Not a Tree: The Myth of Inheritance Hierarchies". *CP*, 35:1-13.

Smith, E. E. 1990. "Categorisation". In Osherson, D. N., & Smith, E. E. (eds.). 1990. *An Invitation to Cognitive Science. Vol 3.: Thinking*. Cambridge, MA: MIT Press, 33-53.

Smith, E. E., & Medin, D. L. 1981. *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Smith, E. E., Medin, D. L., & Rips, L. J. 1984. "A Psychological Approach to Concepts: Comments on Rey's 'Concepts and Stereotypes' ". *CG*, 17:265-274.

Smith, E. E., & Osherson, D. N. 1984. "Conceptual Combination with Prototype Concepts". *CS*, 8:337-361.

Smith, E. E., & Osherson, D. N. 1987. "Compositionality and Typicality". In Schifter, S., & Steele, S. (eds.). 1987. *The 2nd Arizona Colloquium on Cognitive Science*. Tucson, AZ: University of Tucson Press, 37-52.

Smith, E. E., Osherson, D. N., Rips, L. J., & Keane, M. 1988. "Combining Prototypes: A Selective Modification Model". *CS*, 12:485-527.

Smith, E. E., Shoben, E. J., & Rips, L. J. 1974. "Structure and Process in Semantic Memory: A Featural Model for Semantic Decisions". *PR*, 81:214-241.

Smith, E. E., & Sloman, S. A. 1994. "Similarity- Versus Rule-Based Categorisation". *MC*, 22:377-386.

Smith, E. E., Balzano, G. J., & Walker, J. 1978. "Nominal, Perceptual, and Semantic Codes in Picture Categorisation". In Cotton, J. W., & Klatzky, R. L. (eds.). 1978. *Semantic Factors in Cognition*. Hillsdale, NJ: LEA, 137-168.

- Smith, E. R., & Zárate, M. A. 1990. "Exemplar and Prototype Use in Social Categorisation". *Social Cognition*, 8:243-262.
- Smith, J. D. 1989. "Analytic and Holistic Processes in Categorization". In Shepp, B. E., & Ballesteros, S. (eds.). 1989. *Object Perception: Structure and Processes*. Hillsdale, NJ: LEA, 135-148.
- Smith, J. D., & Kemler Nelson, D. G. 1984. "Overall Similarity in Adults' Classification: The Child in All of Us". *JEP:GE*, 113:137-159.
- Smith, J. D., & Shapiro, J. H. 1989. "The Occurrence of Holistic Categorisation". *JML*, 28:386-399.
- Smith, L. B., & Jones, S. S. 1993. "Cognition Without Concepts". *CGD*, 8:181-188.
- Smith, L. B., & Samuelson, L. K. 1997. "Perceiving and Remembering: Category Stability, Variability and Development". In Lamberts, K., & Shanks, D. R. (eds.). 1997. *Knowledge, Concepts, and Categories*. Cambridge, MA: MIT Press, 161-195.
- Smoke, K. L. 1932. "An Objective Study of Concept Formation". *Psychological Monographs*, 42. No 4. (Whole No. 191).
- Snowdown, C. T., & Pola, Y. V. 1978. "Interspecific and Intraspecific Responses to Synthesized Marmoset Vocalizations". *Animal Behaviour*, 26:192-206.
- Spalding, T. L., & Murphy, G. L. 1996. "Effects of Background Knowledge on Category Construction". *JEP:LMC*, 22:525-538.
- Stampe, David. 1979. *A Dissertation on Natural Phonology*. Bloomington: IULC.
- Stampe, D. 1987. "On Phonological Representations". In Dressler, W. U., Luschützky, H.C., Pfeiffer, O.E., & Rennison, J.R. (eds.). 1987. *Phonologica 1984*. London: CUP, 287-299.
- Stanovich, K. E., Cunningham, A. E., & Cramer, B. R. 1984. "Assessing Phonological Awareness in Kindergarten Children: Issues of Task Comparability". *JECP*, 38:175-190.
- Starkey, D. 1981. "The Origins of Concept Formation: Object Sorting and Object Preferences in Early Infancy". *CD*, 52:489-497.
- Steinberg, D. D., & Krohn, R. K. 1975. "The Psychological Reality of Chomsky and Halle's Vowel Shift Rule". In Koerner, E. F. K. (ed.). 1975. *The Transformational-Generative Paradigm and Modern Linguistic Theory*. Amsterdam: John Benjamins, 233-259.
- Stemberger, J. P. 1983. *Speech Errors and Theoretical Phonology: A Review*. Bloomington: Indiana University Linguistics Club.
- Stemberger, J. P., Elman, J. L., & Haden, P. 1985. "Interference between Phonemes During Phoneme Monitoring: Evidence for an Interactive Activation Model of Speech Perception". *JEP:HPP*, 11:475-489.
- Strange, W., & Jenkins, J. J. 1978. "Role of Linguistic Experience in the Perception of Speech". In Walk, R. D., & Pick, H. L. Jr., (eds.). 1978. *Perception and Experience*. New York: Plenum Press, 125-169.



- Strange, W., Keeney, T., Kessel, F. S., & Jenkins, J. J. 1970. "Abstraction over Time of Prototypes from Distortions of Random Dot Patterns: A Replication". *JEP*, 83:508-510.
- Strauss, M. S. 1979. "Abstraction of Prototypical Information in Adults and 10-Month-Old Infants". *JEP:HLM*, 5:618-632.
- Stross, B. 1973. "Acquisition of Botanical Terminology by Tzeltal Children". In Edmonson, M. S. (ed.). 1973. *Meaning in Mayan Languages*. The Hague: Mouton, 107-141.
- Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. 1970. "Motor Theory of Speech Perception: A Reply to Lane's Critical Review". *PR*, 77:234-249.
- Summerfield, Q. 1981. "Articulatory Rate and Perceptual Constancy in Phonetic Perception". *JEP:HPP*, 7:1074-1095.
- Sutcliffe, J. P. 1993. "Concept, Class, and Category in the Tradition of Aristotle". In Van Mechelen, I., Hampton, J., Michalski, R. S., & Theuns, P. (eds.). 1993. *Categories and Concepts: Theoretical Views and Inductive Data*. New York: Academic Press, 35-65.
- Swadesh, M. 1934. "The Phonemic Principle". *Lg*, 10:117-129.
- Swinney, D. A., & Prather, P. 1980. "Phonemic Identification in a Phoneme Monitoring Experiment: The Variable Role of Uncertainty about Vowel Contexts". *PP*, 27:104-110.
- Tajfel, H., & Billig, M. 1974. "Familiarity and Categorisation in Intergroup Behavior". *JESP*, 10:159-170.
- Tajfel, H., Billig, M., Bundy, R. P., & Flament, C. 1971. "Social Categorisation and Intergroup Behavior". *EJSP*, 1:149-178.
- Tanaka, J. W., & Taylor, M. 1991. "Object Categories and Expertise: Is the Basic Level in the Eye of the Beholder?". *CP*, 23:457-482.
- Taylor, J. R. 1995. [1989]. *Linguistic Categorisation: Prototypes in Linguistic Theory*. (2nd edition). Oxford, England: Clarendon Press.
- Taylor, J. R. 1990. "Schemas, Prototypes, and Models: In Search of the Unity of the Sign". In Tsohatzidis, S. L. (ed.). 1990. *Meanings and Prototypes. Studies in Linguistic Categorisation*. London: Routledge, 521-534.
- Taylor, S. E. 1981. "A Categorisation Approach to Stereotyping". In Hamilton, D. L. (ed.). 1981. *Cognitive Processes in Stereotyping and Intergroup Behavior*. Hillsdale, NJ: LEA, 83-114.
- Thorson, J. Weber, T., & Huber, F. 1982. "Auditory Behavior of the Cricket. II. Simplicity of Calling-Song Recognition in *Gryllus*, and Anomalous Phonotaxis at Abnormal Carrier Frequency". *JCP*, 146:361-378.
- Trager, G. L. 1942. "The Phonemic Treatment of Semivowels". *Lg*, 18:220-223.
- Trager, G. L., & Smith, H. L. 1951. *An Outline of English Structure*. Norman, Oklahoma: Battenburg Press.
- Treiman, R. 1983. "The Structure of Spoken Syllables: Evidence from Novel Word Games". *CG*, 15:49-74.

- Treiman, R. 1984. "On the Status of Final Consonant Clusters in English Syllables". *JVLVB*, 23:343-356.
- Treiman, R. 1985a. "Onsets and Rimes as Units of Spoken Syllables: Evidence from Children". *JECP*, 39:161-181.
- Treiman, R. 1985b. "Spelling of Stop Consonants After / s / by Children and Adults". *Applied Psycholinguistics*, 6:261-282.
- Treiman, R. 1985c. "Phonemic Awareness and Spelling: Children's Judgments Do Not Always Agree with Adults". *JECP*, 39:182-201.
- Treiman, R. 1986. "The Division Between Onsets and Rimes in English Syllables". *JML*, 25:476-491.
- Treiman, R. 1987. "On the Relationship between Phonological Awareness and Literacy". *CPC*, 7:524-529.
- Treiman, R. 1988. "The Internal Structure of the Syllable". In Carlson, G. N., & Tanenhaus, M. K. (eds.). 1988. *Linguistic Structure in Language Processing*. Dordrecht, Holland: Kluwer Academic Publishers, 27-52.
- Treiman, R. 1992. "The Role of Intrasyllabic Units in Learning to Read and Spell". In Gough, P. B., Ehri, L. C., & Treiman, R. (eds.). 1992. *Reading Acquisition*. Hillsdale, NJ: LEA, 65-106.
- Treiman, R. 1993. *Beginning to Spell: A Study of First-Grade Children*. New York: Oxford University Press.
- Treiman, R., & Baron, J. 1981. "Segmental Analysis Ability: Development and Relation to Reading Ability". In Walker, T. G., & MacKinnon, G. E. G. (eds.) 1981. *Reading Research: Advances in Theory and Practice*. Vol. 3. New York: Academic Press, 159-198.
- Treiman, R., & Breaux, A. M. 1982. "Common Phoneme and Overall Similarity Relations Among Spoken Syllables: Their Use by Children and Adults". *JPR*, 11:569-598.
- Treiman, R., Cassar, M., & Zukowski, A. 1994. "What Types of Linguistic Information Do Children Use in Spelling? The Case of Flaps". *CD*, 65:1318-1337.
- Treiman, R., Salasoo, A., Slowiczek, L. M., & Pisoni, D. B. 1982. "Effects of Syllable Structure on Adults' Phoneme Monitoring Performance". *Research on Speech Perception Progress Report No 8*. Bloomington: Indiana University, Speech Research Laboratory, 63-81.
- Trentin, R., & Salmaso, P. 1990. "The Influence of Affect and Knowledge on the Work Category Prototype". *CPC*, 10:413-432.
- Trillmich, F. 1976. "Learning Experiments on Individual Recognition in Budgerigars". *Zeitschrift für Tierpsychologie*, 41:372-395.
- Trubetzkoy, N. 1969. [1939]. *Principles of Phonology*. Translated by C. A. M. Baltaxe. Berkeley and Los Angeles: University of California Press.
- Tversky, A. 1977. "Features of Similarity". *PR*, 84:327-352.



- Tversky, A., & Gati, I. 1978. "Studies of Similarity". In Rosch, E., & Lloyd, B. B. (eds.). 1978. *Cognition and Categorisation*. Hillsdale, NJ: LEA, 79-98.
- Tversky, B. 1985. "Development of Taxonomic Organization of Named and Pictured Categories". *DP*, 21:1111-1119.
- Tversky, B. 1986. "Components and Categorisation". In Craig, C. (ed.). 1986. *Noun Classes and Categorisation*. Amsterdam: John Benjamins, 63-75.
- Tversky, B. 1989. "Parts, Partonomies, and Taxonomies". *DP*, 25:983-995.
- Tversky, B. 1990. "Where Partonomies and Taxonomies Meet". In Tsohatzidis, S. L. (ed.). 1990. *Meanings and Prototypes: Studies in Linguistic Categorisation*. London: Routledge, 334-344.
- Tversky, B., & Hemenway, K. 1983. "Categories of Environmental Scenes". *CP*, 15:121-149.
- Tversky, B., & Hemenway, K. 1984. "Objects, Parts, and Categories". *JEP:GE*, 113:169-193.
- Tversky, B., & Hemenway, K. 1991. "Parts and the Basic Level in Natural Categories and Artificial Stimuli: Comments on Murphy (1991)". *MC*, 19:439-442.
- Twaddell, W. F. 1935. "On Defining the Phoneme". *Language Monographs*, Vol. 16. Baltimore: Linguistic Society of America.
- Ungerer, F. 1994. "Basic Level Concepts and Parasitic Categorisation: A Cognitive Alternative to Conventional Semantic Hierarchies". *Zeitschrift für Anglistik und Amerikanistik*, 42:148-162.
- Ungerer, F. & Schmid, H. J. *An Introduction to Cognitive Linguistics*. New York: Addison Wesley Longman.
- Van de Walle, G. A. 1997. *Perceptual Foundations of Categorisation in Infancy: The Animal/Vehicle Distinction*. Phd. Dissertation. Cornell University, Ithaca, NY.
- Van Ooyen, B., Cutler, A., & Norris, D. 1991. "Detection Times for Vowels versus Consonants". *Proceedings of EuroSpeech 91*, Vol. 3. Genoa, Italy, 1451-1454.
- Van Ooyen, B., Cutler, A., & Norris, D. 1992. "Detection of Vowels and Consonants with Minimal Acoustic Variation". *Speech Communication*, 11:101-108.
- Vaughan, W., Jr. 1988. "Formation of Equivalence Sets in Pigeons". *JEP:ABP*, 14:36-42.
- Volaitis, L. E., & Miller, J. L. 1991. "Influence of a Syllable's Form on the Perceived Internal Structure of Voicing Categories". *JASA*, 89, 1998.
- Volaitis, L. E., & Miller, J. L. 1992. "Phonetic Prototypes: Influence of Place of Articulation and Speaking Rate on the Internal Structure of Voicing Categories". *JASA*, 92:723-735.
- von Fersen, L., & Lea, S. E. G. 1990. "Category Discrimination by Pigeons Using Five Polymorphous Features". *JEAB*, 54:69-84.
- von Gelder, T. J. 1993. "Is Cognition Categorisation?". In Nakamura, G. V., Taraban, R. M., & Medin, D. L. (eds.). 1993. *The Psychology of Learning and Motivation. Vol. 29.: Categorisation by Humans and Machines*. San Diego: Academic Press, 469-494.

- Walker, J. H. 1975. "Real-World Variability, Reasonableness Judgments, and Memory Representations for Concepts". *JVLVB*, 14:241-252.
- Wang, H. S. 1985. *On the Productivity of Vowel Shift Alternations in English: An Experimental Study*. Ph.D. dissertation. University of Alberta.
- Wang, H. S., & Derwing, B. L. 1986. "More on English Vowel Shift: The Back Vowel Question". *PHY*, 99-116.
- Ward, T. B., & Scott, J. 1987. "Analytic and Holistic Modes of Learning Family-Resemblance Concepts". *MC*, 15:85-89.
- Ward, T. B., Vela, E., & Hass, S. D. 1990. "Children and Adults Learn Family-Resemblance Categories Analytically". *CD*, 61:593-605.
- Wasserman, E. A., DeVolder, C. L., & Coppage, D. J. 1992. "Non-Similarity-Based Conceptualization in Pigeons via Secondary or Mediated Generalization". *PSS*, 3:374-379.
- Wasserman, E. A., Kiedinger, R. E., & Bhatt, R. S. 1988. "Conceptual Behavior in Pigeons: Categories, Subcategories, and Pseudocategories". *JEP:ABP*, 14:235-246.
- Watanabe, S. 1988. "Failure of Visual Prototype Learning in the Pigeon". *ALB*, 16:147-152.
- Waters, R. S., & Wilson, W. A., Jr. 1976. "Speech Perception by Rhesus Monkeys: The Voicing Distinctions in Synthesized Labial and Velar Stop Consonants". *PP*, 19:285-289.
- Wattenmaker, W. D. 1991. "Learning Modes, Feature Correlations, and Memory-Based Categorisation". *JEP:LMC*, 17:908-923.
- Wattenmaker, W. D. 1993. "Incidental Concept Learning, Feature Frequency, and Correlated Properties". *JEP:LMC*, 19:203-222.
- Wattenmaker, W. D. 1995. "Knowledge Structures and Linear Separability: Integrating Information in Object and Social Categorisation". *CP*, 28:274-328.
- Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. 1986. "Linear Separability and Concept Learning: Context, Relational Properties, and Concept Naturalness". *CP*, 18:158-194.
- Wattenmaker, W. D., Nakamura, G. V., & Medin, D. L. 1988. "Relationships between Similarity-Based and Explanation-Based Categorisation". In Hilton, D. J. (ed.). 1988. *Contemporary Science and Natural Explanation: Commonsense Conceptions of Causality*. Brighton, England: Harvester Press, 204-240.
- Waxman, S. R. 1985. *Hierarchies in Classification and Language: Evidence from Preschool Children*. Unpublished doctoral dissertation, University of Pennsylvania.
- Waxman, S. R. 1990. "Linguistic Biases and the Establishment of Conceptual Hierarchies: Evidence for Preschool Children". *CGD*, 5:123-150.
- Waxman, S. R., & Hall, D. G. 1993. "The Development of a Linkage between Count Nouns and Object Categories: Evidence from Fifteen- to Twenty-One-Month-Old Infants". *CD*, 64:1224-1241.



- Waxman, S. R., Shipley, E. F., & Shepperson, B. 1991. "Establishing New Subcategories: The Role of Category Labels and Existing Knowledge". *CD*, 62:127-138.
- Wayland, S. C., Miller, J. L., & Volaitis, L. E. 1994. "The Influence of Sentential Speaking Rate on the Internal Structure of Phonetic Categories". *JASA*, 95:2694-2701.
- Weber, T., Thorson, J., & Huber, F. 1981. "Auditory Behavior of the Cricket. I. Dynamics of Compensated Walking and Discrimination Paradigm on the Kramer Treadmill". *JCP*, 141:214-232.
- Weitzman, R. 1990a. "The Role of the Critical Band in Learning Vowel Categorizations". *JASA*, 87, S159.
- Weitzman, R. 1990b. "Further Studies of the Critical Band's Role in Learning Vowel Categories". *JASA*, 88, S56.
- Weitzman, R. 1992. "Vowel Categorisation and the Critical Band". *LS*, 35:115-125.
- Weitzman, R. 1993. "How to Get the Horse to Open Its Mouth: Using the Concept Formation Paradigm in Speech Perception Research". In Nevis, Joel A., et al. (eds.). *Papers in Honor of Frederick Brengelman on the Occasion of the 25th Anniversary of the Linguistics Department at California State University*. Fresno, CA: California State University, 141-149.
- Welker, R. 1982. "Abstraction of Themes from Melodic Variations". *JEP:HPP*, 8:435-447.
- Wells-Jensen, S. 1999. *Cognitive Correlates of Linguistic Complexity: A Cross-Linguistic Comparison of Errors in Speech*. Ph.D. dissertation. University of New York at Buffalo.
- Wheeler, C. J. 1980. "On the Relationship Between Phonology and Psychology". *Papers in Linguistics: International Journal of Human Communication*, 13:51-100.
- White, K. G., Alsop, B., & Williams, L. 1993. "Prototype Identification and Categorisation of Incomplete Figures by Pigeons". *Behavioural Processes*, 30:253-258.
- White, T. G. 1982. "Naming Practices, Typicality, and Underextension in Child Language". *JECp*, 33:324-346.
- Whitfield, T. W. A., & Slatter, P. E. 1979. "The Effects of Categorisation and Prototypicality on Aesthetic Choice in a Furniture Selection Task". *BJP*, 70:65-75.
- Whittlesea, B. W. A. 1987. "Preservation of Specific Experiences in the Representation of General Knowledge". *JEP:LMC*, 13:3-17.
- Whittlesea, B. W. A. 1997. "The Representation of General and Particular Knowledge". In Lamberts, K., & Shanks, D. (eds.). 1997. *Knowledge, Concepts, and Categories*. Cambridge, MA: MIT Press, 335-370.
- Wierzbicka, A. 1990. "'Prototypes saves': On the Uses and Abuses of the Notion of 'Prototype' in Linguistics and Related Fields". In Tsohatzidis, S. L. (ed.). 1990. *Meanings and Prototypes: Studies in Linguistic Categorisation*. London: Routledge, 347-367.
- Wilder, D. A. 1981. "Perceiving Persons as a Group: Categorisation and Intergroup Relations". In Hamilton, D. L. (ed.). 1981. *Cognitive Processes in Stereotyping and Intergroup Behavior*. Hillsdale, NJ: LEA, 213-257.

- Wilkins, A. J. 1971. "Conjoint Frequency, Category Size and Categorisation Time". *JVLVB*, 10:382-385.
- Williams, T. M., Fryer, M. L., & Aiken, L. S. 1977. "Development of Visual Pattern Classification in Pre-School Children: Prototypes and Distinctive Features". *DP*, 13:577-584.
- Willis, C. 1971. "Synthetic Vowel Categorization and Dialectology". *LS*, 14:213-228.
- Windsor Lewis, J. 1990. "Happy land Reconnoitred: The Unstressed Word-Final -y Vowel in General British Pronunciation". In Ramsaran, S. (ed.). 1990. *Studies in the Pronunciation of English*. Londong: Routledge, 159-167.
- Winitz, H., LaRiviere, C., & Herriman, E. 1975. "Variations in VOT for English Initial Stops". *JP*, 3:41-52.
- Wisniewski, E. J. 1995. "Prior Knowledge and Functionally Relevant Features in Concept Learning". *JEP:LMC*, 21:449-468.
- Wisniewski, E. J. 1996. "Construal and Similarity in Conceptual Combination". *JML*, 35:434-453.
- Wisniewski, E. J., Imai, M., & Casey, L. 1996. "On the Equivalence of Superordinate Concepts". *CG*, 60:269-298.
- Wisniewski, E. J., & Medin, D. L. 1994. "On the Interaction of Theory and Data in Concept Learning". *CS*, 18:221-281.
- Wisniewski, E. J., & Medin, D. L. 1991. "Harpoons and Long Sticks: The Interaction of Theory and Similarity in Rule Induction". In Fisher, D. H., Pazzani, M. J., & Langley, P. (eds.). *Concept Formation: Knowledge and Experience in Unsupervised Learning*. San Mateo, CA: Morgan Kaufman, 237-278.
- Wittgenstein, L. 1953. *Philosophical Investigations*. New York: Macmillan.
- Wojciszke, B., & Pienkowski, R. 1991. "Prototypical Structure of Personality Trait Concepts and Person Perception". *CPC*, 11:213-228.
- Wolf, C. G. 1973. "The Perception of Stop Consonants by Children". *JECp*, 16:318-331.
- Wright, A. A., Santiago, H. C., & Sands, S. F. 1984. "Monkey Memory: Same/Different Concept Learning, Serial Probe Acquisition, and Probe Delay Effects". *JEP:ABP*, 10:513-529.
- Wright, A. A., Shyan, M. R., & Jitsumori, M. 1990. "Auditory Same/Different Concept Learning by Monkeys". *ALB*, 18:287-294.
- Yoshikubo, S. 1985. "Species Discrimination and Concept Formation by Rhesus Monkeys (*Macaca Mulatta*)". *Primates*, 26:285-299.
- Younger, B. A. 1984. *The Co-Variation of Dimensional Values as a Determinant of the Segregation of Items into Categories by 10-Month-Old Infants*. Doctoral Dissertation. University of Texas, Austin.
- Younger, B. A. 1985. "The Segregation of Items into Categories by Ten-Month-Old Infants". *CD*, 56:1574-1583.

Younger, B. A. 1990a. "Infants' Detection of Correlations Among Feature Categories". *CD*, 61:614-620.

Younger, B. A. 1990b. "Infant Categorisation: Memory for Category-Level and Specific Item Information". *JECp*, 50:131-155.

Younger, B. A. 1992. "Developmental Change in Infant Categorisation: The Perception of Correlations Among Facial Features". *CD*, 63:1526-1535.

Younger, B. A., & Cohen, L. B. 1983. "Infant Perception of Correlations Among Attributes". *CD*, 54:858-867.

Younger, B. A., & Cohen, L. B. 1985. "How Infants Form Categories". In Bower, G. H. (ed.). 1985. *The Psychology of Learning and Motivation. Vol 9*. New York: Academic Press, 211-247.

Younger, B. A., & Cohen, L. B. 1986. "Developmental Changes in Infants' Perception of Correlations Among Attributes". *CD*, 57:803-815.

Younger, B. A., & Gotlieb, S. 1988. "Development of Categorisation Skills: Changes in the Nature or Structure of Infant Form Categories ?". *DP*, 24:611-619.

Zadeh, L. A. 1982. "A Note on Prototype Theory and Fuzzy Sets". *CG*, 12:291-297.

<http://www.linguistlist.org/issues/4/4-934.html>

<http://www.linguistlist.org/issues.4/4-952.html>

<http://www.linguistlist.org/~ask-ling/archive-most-recent/msg01429.html>

LA CATEGORIZACIÓN DE LOS SONIDOS DEL INGLÉS: EVIDENCIA EXPERIMENTAL EN FONOLOGÍA

TABLA DE CONTENIDOS.....	i
INTRODUCCIÓN:.....	iii
CAPÍTULO I: SOBRE LA CATEGORIZACIÓN.....	1
1.1. Categorización: Un Proceso Cognitivo Básico.....	1
1.1.1. Categorización y Percepción.....	2
1.1.2. Categorización, Cognición, y Memoria.....	2
1.1.3. Categorización y Conducta.....	4
1.1.4. Categorización y Lenguaje.....	4
1.1.5. Tipos de Categorías.....	5
1.2. Estudios de Categorización con Humanos Adultos.....	7
1.2.1. Modelos Clásicos de Categorización.....	7
1.2.1.1. Aprendizaje, Representación Mental y Almacenamiento.....	7
1.2.1.2. Clasificación.....	8
1.2.1.3. Decadencia del Modelo Clásico: Cuestionando el Paradigma.....	8
1.2.1.3.1. Imposibilidad de Especificar Rasgos Definitivos.....	8
1.2.1.3.2. Uso de Atributos Innecesarios en la Categorización.....	9
1.2.1.3.3. Miembros de Categorías Dudosos.....	9
1.2.1.3.4. Tipicidad.....	9
1.2.1.3.5. Fenómenos Taxonómicos.....	10
1.2.1.4. La Utilidad del Modelo Clásico.....	10
1.2.1.5. Modelos Alternativos de Categorización: Modelos Basados en la Similitud.....	11
1.2.2. Modelos Probabilísticos de Categorización.....	13
1.2.2.1. Aprendizaje, Representación y Almacenamiento Mentales.....	13
1.2.2.2. Clasificación.....	13
1.2.2.3. Ventajas y Desventajas de los Modelos Probabilísticos de la Categorización.....	14
1.2.3. Modelos de Ejemplares de Categorización.....	14
1.2.3.1. Aprendizaje, Representación y Almacenamiento Mentales.....	15
1.2.3.2. Clasificación.....	15
1.2.3.3. Ventajas y Desventajas de Modelos de Ejemplares de Categorización.....	15
1.2.4. Los Modelos Mixtos de Categorización.....	16
1.2.4.1. Aprendizaje, Representación Mental, Almacenamiento, y Clasificación.....	17
1.2.4.2. Ventajas y Desventajas de los Modelos Mixtos de Categorización.....	17
1.2.5. Insuficiencia de los Modelos Basados en la Similitud.....	17
1.2.5.1. La Emergencia de la Explicación Basada en “Teorías” de la Categorización.....	18
1.2.6. Modelos de Categorización: Comentarios Finales.....	19
1.3. La Investigación acerca de la Categorización con Otras Poblaciones.....	19
1.3.1. La Categorización en los Niños.....	20
1.3.2. La Categorización en los Bebés.....	20
1.3.3. La Categorización en las Especies no Humanas.....	21
1.3.3.1. Comparación entre Especies.....	22
CAPÍTULO II: SOBRE LA CATEGORIZACIÓN DE LOS SONIDOS.....	23
2.1. La categorización de los sonidos: Evidencia experimental.....	23
2.1.1. La categorización de los sonidos por los Adultos.....	23
2.1.2. La categorización de los sonidos por los Niños, Los Bebés, y las Especies No Humanas.....	25
2.2. La Categorización de los Sonidos del Inglés y su Relación con la Fonología.....	25
2.2.1. Fonología Descriptiva vs. Fonología Psicolingüística o Psicológica.....	26
2.2.2. Críticas al Uso de los Experimentos en Fonología.....	27
2.2.3. Ventajas del Uso de Experimentos en Fonología.....	28
2.2.4. La Complementariedad de la Fonología Descriptiva y La Fonología Psicolingüística.....	28
2.3. Formación de Conceptos como Paradigma Experimental.....	29
2.3.1. Partes de un Experimento de Formación de Conceptos.....	29
2.3.1.1. Las instrucciones.....	30
2.3.1.2. La Sesión de Aprendizaje.....	30
2.3.1.3. La Sesión de Test.....	31
2.3.1.4. La Entrevista Post-Experimental.....	31
2.3.2. Interpretación de los Datos.....	31

CAPÍTULO III. EVIDENCIA EXPERIMENTAL PROCEDENTE DE LA CATEGORIZACIÓN DE LOS SONIDOS: UNA INVESTIGACIÓN FONOLÓGICA PSICOLINGÜÍSTICA.32

3. 1. Introducción.	32
3.1.1. Aspectos Generales Relacionados con los Experimentos Descritos.	32
3.1.1.1. Propósitos Generales y Específicos de los Experimentos y Selección de los Fenómenos Fonológicos Investigados.	32
3.1.1.2. Selección de los Sujetos.	33
3.1.1.3. Asignación de las Tareas Experimentales a los Sujetos.	33
3.1.1.4. Estímulos, Materiales, y Aparatos Experimentales.	33
3.1.1.5. Aspectos Procedimentales Específicos de las Experimentos CF Descritos.	34
3.2. Experimento 1. La categoría “Sonidos Consonánticos” en Inglés.	34
3.2.1. Diseño.	34
3.2.2. Método.	35
3.2.2.1. Sujetos.	35
3.2.2.2. Estímulos.	35
3.2.2.3. Procedimiento.	36
3.2.2.3.1. Instrucciones.	36
3.2.2.3.2. Sesión de Aprendizaje, Sesión de Test, y Entrevista Post-Experimental.	37
3.2.3. Resultados.	37
3.2.3.1. Sesiones de Aprendizaje y de Test.	37
3.2.4. Discusión.	39
3.3. Experimento 2. La categoría “Oclusivas Orales” en Inglés.	39
3.3.1. Diseño.	39
3.3.2. Método.	41
3.3.2.1. Sujetos.	41
3.3.2.2. Estímulos.	41
3.3.2.3. Procedimiento.	43
3.3.2.3.1. Instrucciones.	43
3.3.2.3.2. Sesión de Aprendizaje, Sesión de Test, y Entrevista Post-Experimental.	43
3.3.3. Resultados.	43
3.3.3.1. Sesiones de Aprendizaje y Test.	43
3.3.4. Discusión.	44
3.4. Experimento 3. La categoría “Fonema / p /” en Inglés.	44
3.4.1. Diseño.	44
3.4.2. Método.	45
3.4.2.1. Sujetos.	45
3.4.2.2. Estímulos.	46
3.4.2.3. Procedimiento.	47
3.4.2.3.1. Instrucciones.	47
3.4.2.3.2. Sesión de Aprendizaje, Sesión de test, y Entrevista Post-Experimental.	48
3.4.3. Resultados.	48
3.4.3.1. Sesión de Aprendizaje y Sesión de Test.	48
3.4.4. Discusión.	49
3.5. Experimento 4. La categoría “Alófono [p ^h]” en Inglés.	49
3.5.1. Diseño.	49
3.5.2. Método.	50
3.5.2.1. Sujetos.	50
3.5.2.2. Estímulos.	51
3.5.2.3. Procedimiento.	52
3.5.2.3.1. Instrucciones.	52
3.5.2.3.2. Sesión de Aprendizaje, Sesión de test, y Entrevista Post-Experimental.	52
3.5.3. Resultados.	53
3.5.3.1. Sesión de Aprendizaje y Sesión de Test.	53
3.5.4. Discusión.	53

CAPÍTULO IV: CONCLUSIONES.55

4.1. Introducción.	55
4.2. Categorización, La Categorización de los Sonidos, la Fonología, y La Experimentación.	55
4.3. CF como un Paradigma Experimental.	55
4.3.1. Evidencia Obtenida de las Categorías Individuales.	56
4.3.2. Evidencia Obtenida de las Categorías: Comparación entre Categorías.	56
4.3.3. La Utilidad del Paradigma de “Formación de Conceptos”: Direcciones Futuras para la Aplicación de la Técnica CF.	57
REFERENCIAS.	57

INTRODUCCIÓN

El propósito general del presente trabajo es el de obtener evidencia experimental para proporcionar una explicación a ciertas cuestiones fonológicas que corresponda muy de cerca al conocimiento que los hablantes tienen acerca de los sonidos del lenguaje. En otras palabras, se pretende proponer descripciones fonológicas que representen el conocimiento que los hablantes tienen de los fenómenos fonológicos investigados. Para ello se recurre al estudio de la categorización que los hablantes hacen de los sonidos del lenguaje. Como la categorización es un habilidad cognitiva tan esencial para el ser humano y su utilidad para averiguar la concepción que los hablantes tienen de la estructura fónica de su lengua parece incuestionable, esta tesis revisa la literatura existente sobre la categorización en general, para pasar posteriormente a una revisión de la literatura sobre la categorización de los sonidos, y, después de haberse escogido una técnica experimental útil como la de “formación de conceptos”, finalmente describir cuatro estudios fonológicos experimentales con hablantes reales llevados a cabo para obtener evidencia experimental y poder así proporcionar una explicación satisfactoria desde un punto de vista psicológico a los problemas que tratan dichos experimentos y satisfacer así el objetivo del presente trabajo.

CAPÍTULO I. SOBRE LA CATEGORIZACIÓN.

1.1. Categorización: Un Proceso Cognitivo Básico.

Con frecuencia se ha dicho que la categorización es un proceso básico, esencial, o fundamental. Por ejemplo, “no hay nada más básico que la categorización para nuestro pensamiento, percepción, acción, y habla” (Lakoff 1987: 5), “la categorización es un incuestionablemente importante tipo de proceso cognitivo” (Busemeyer *et al.* 1997: 2), “la categorización es una de las funciones cognitivas más básicas” (Corter & Gluck 1992: 291), etc.

La categorización es básica en dos sentidos: en un sentido técnico, la categorización es un proceso cognitivo básico en el sentido del término de Flavell (1985), quien propuso dividir los fenómenos de la memoria en cuatro tipos distintos. Uno de estos tipos, llamado “procesos básicos”, se caracteriza por dos rasgos fundamentales: no sufrir modificación significativa con la edad (que no sea la debida al proceso de madurez) y ser procesos inconscientes ya que las personas no son conscientes del funcionamiento de dichos procesos. En principio, la categorización cumple con estas dos características ya que ésta comienza es innata y se presenta ya en los bebés de tan sólo 3 meses de edad y parece permanecer básicamente inalterada durante el resto de la vida del individuo. Por otro lado, la categorización es un proceso inconsciente que la gente realiza automáticamente y sin esfuerzo (Barsalou 1992a: 16) y cuyo funcionamiento es solamente sentido en casos muy especiales (Lakoff 1987: 6).

Sin embargo, las afirmaciones acerca de lo básica que es la categorización simplemente parecen transmitir la idea de que la categorización es simplemente una función particularmente importante. Tres son los aspectos que nos ayudan a clasificar el porqué de la fundamental de la categorización. En primer lugar, el hecho de que todo ser humano posee esta capacidad revela lo necesaria que es para el ser humano. En segundo lugar, el funcionamiento de la categorización es omnipresente, ocurriendo siempre que el organismo está expuesto a algún tipo de estímulo (Len & Murphy 1997: 1154; Medin *et al.* 2001: 366) y en cualquier modalidad sensorial. En tercer lugar, el papel jugado por la categorización en el funcionamiento cognitivo humano la destaca como una pieza fundamental del engranaje de las habilidades cognitivas ya que la categorización parece estar a su vez íntimamente unida a los procesos perceptivos, cognitivos, la memoria, la conducta, y el lenguaje (Bornstein 1984: 313-315).

1.1.1. Categorización y Percepción.

La categorización parece esencial para la percepción. Según Bruner, “la percepción implica un acto de categorización” (1957: 123). Para Hunt, “el acto de categorizar estímulos sensoriales es básico para nuestra percepción del mundo que nos rodea” (1962: 5). Los organismos se enfrentan a un mundo infinito de estímulos únicos y esto constituye, sin duda, un problema para la percepción ya que si los organismos tuvieran que percibir cada estímulo no encontrado previamente como exclusivamente único, aquellos se verían superados rápidamente por la extrema diversidad del entono y serían, por tanto, incapaces de encontrar sentido a su experiencia (Smith & Medin 1981: 1).

Sin embargo, los organismos simplifican este problema por medio de la habilidad de categorizar ya que la categorización permite a los individuos ordenar su experiencia al considerar algunos estímulos percibidos como similares o equivalentes y no como únicos (Smith & Medin 1981). La categorización permite a los organismos organizar el input recibido del entono en unidades cognitivamente manejables al dirigirlos a percibir grupos de experiencias como básicamente similares de tal manera que la complejidad del entono se reduce (Bornstein 1984: 314). En otras palabras, la categorización hace posible que un organismo reduzca la infinita variación en el mundo a proporciones manejables y así se consigue simplificar dicha variación hasta cierto punto. En este sentido, los organismos consiguen una de sus necesidades más esenciales: un sentido de estabilidad y orden en el medio en el que se desenvuelven (Mervis 1985: 293). La categorización permite a los individuos percibir estímulos en el entono como si estuvieran organizados en un sistema de estructuras y relaciones y así les permite conseguir un sentido de orden y continuidad.

Sin embargo, la categorización afecta a la percepción también en otro sentido. Los psicólogos han estado siempre intrigados por la posibilidad de que los conceptos que la gente aprende influyan en sus habilidades perceptuales. Por ejemplo, al tiempo que expertos como los traumatólogos aprenden a distinguir entre distintos tipos de fracturas existen, parecen asimismo adquirir nuevas formas de estructura perceptualmente los objetos que categorizan. En otras palabras, la categorización puede producir un aprendizaje perceptual que dura más allá del acto de categorización. Existe evidencia experimental que avala la hipótesis de que la experiencia en adquirir nuevas categorías puede alterar la sensibilidad perceptiva y la habilidad de realizar discriminaciones perceptuales (Goldstone 1994a; MacKintosh 1995: 196-1998).

1.1.2. Categorización, Cognición, y Memoria.

Los investigadores están de acuerdo en el papel facilitador de la categorización y las categorías para la cognición y la memoria. En primer lugar, a través de la categorización se

crean categorías conceptuales abstractas de la experiencia con los estímulos, lo cual constituye “uno de los logros de la inteligencia humana” (Busemeyer *et al.* 1997: 1). La categorización permite el aprendizaje conceptual, un aspecto fundamental del pensamiento enteligente (Wisniewski & Medin 1991: 237). En segundo lugar, las categorías conceptuales ayudan a reconocer información familiar al proveer “receptáculos” (es decir, las categorías conceptuales) en las que la información recibida queda codificada para su uso en subsiguientes actos de categorización. Los organismos están constantemente expuestos a estímulos que no han visto antes y sin embargo los categorizan como miembros de clases familiares. Por tanto, esos receptáculos convierte lo no familiar en familiar (Bornstein 1984: 314; Len & Murphy 1997: 1154; Mervis 1980: 279; Mervis & Pani 1980: 497; Spalding & Murphy 1996: 525).

En tercer lugar, las categorías no sólo permiten identificar nuevos estímulos como miembros de categorías familiares, sino también aplicar o generalizar el conocimiento previamente adquirido acerca de la categoría a nuevos miembros a través de las inferencias (ej. Bornstein 1984: 314; Malt *et al.* 1995; Mervis 1980: 279; Murphy & Ross 1994: 148; Ross & Murphy 1996: 736). Las categorías son útiles porque, una vez que se sabe a qué categoría un determinado estímulo es asignado, el individuo puede predecir los atributos internos o funciones de dicho estímulo (Mervis 1985: 293; Wisniewski & Medin 1991: 237). Por ejemplo, un organismo que busca comida necesita saber si un determinado tipo de planta es venenosa, nutritiva, dulce, etc. Habiendo identificado dicha planta como un determinado tipo de seta, el organismo puede entonces comprender si es peligrosa o no, si es nutritiva o no, etc.

Finalmente, otra ventaja de las categorías creadas a través de la categorización es que aquellas reducen la necesidad de un aprendizaje continuo ya que los estímulos nuevos pueden ser inmediatamente entendidos como ejemplos de algún tipo de estímulo previamente experimentado. En consecuencia, aquellos estímulos pueden ser tratados apropiadamente en base a un aprendizaje previo (Mervis & Rosch 1981: 89). Por tanto, muchos menos estímulos necesitan ser retenidos en la memoria tras la decodificación (Bornstein 1984: 314). Además, agrupar objetos en categorías permite un almacenamiento más eficiente de la información en la memoria de estos grupos de objetos (Bornstein 1984: 314; Corter & Gluck 1992: 291). Más aún, resulta más fácil recordar estímulos previamente encontrados. De hecho, sin categorías, los organismos probablemente serían incapaces de recordar poco más que una pequeña fracción de lo que experimentan. Los conceptos sirven para mejorar la capacidad memorística (Bornstein 1984: 314; Smith & Medin 1981: 1).

Debido a todas las formas en las que la categorización se relaciona con la cognición y la memoria, no es pues sorprendente encontrar afirmaciones en la literatura especializada que así lo reflejan: “la categorización es una parte central del pensamiento inteligente” (Murphy & Ross 1994: 148).

1.1.3. Categorización y Conducta.

La categorización y las categorías cognitivas son fenómenos internos no observables. Sin embargo, el funcionamiento de la categorización y la existencia de tales categorías se infiere de los patrones de comportamiento. En primer lugar, la categorización se relaciona con la conducta porque aquella hace que los organismos respondan de forma similar a los miembros de una clase de estímulos y de forma diferente a los miembros de otras clases de estímulos. En otras palabras, una categoría se entiende que existe en la mente cuyo se trata de forma equivalente a dos o más estímulos distinguibles y de forma diferente de los miembros de otras categorías.

Sin embargo, la categorización influye en el comportamiento también de otra forma. Clasificar una determinada entidad como un miembro de una categoría u otra determina qué tipo de comportamiento los organismos realizarán en el futuro hacia ese estímulo una vez que ha sido clasificado. Esto se relaciona con las inferencias que promueve la categorización. Sin embargo, un tratamiento igual hacia los miembros de una categoría puede también causar problemas ya que tratarlos exactamente igual en cada ocasión puede no ser justo. Para ser más precisos, categorizar una serie de objetos como iguales puede llevar a tratarlos como más similares de lo que realmente lo son simplemente por su pertenencia común a la misma categoría. Este riesgo es especialmente evidente en la categorización social, donde los estereotipos pueden a veces tener consecuencias peligrosas.

1.1.4. Categorización y Lenguaje.

Se ha sugerido que la categorización es básica para un entendimiento mayor de la naturaleza de la adquisición del lenguaje, su desarrollo y uso (ej. Anglin 1977; Bruner *et al.* 1956; Clark 1973; Nelson 1973; Schlesinger 1982; etc.). Según Nelson, “la categorización de los patrones sonoros y de los eventos en el mundo real es básica para el aprendizaje de una lengua” (Nelson 1973: 21). La categorización también ha sido considerada esencial para la comprensión del lenguaje (van Gelder 1993), la percepción o reconocimiento del habla (Hunt 1962: 5; Lotto 2000), el aprendizaje léxico más temprano (Poulin-Dubois *et al.* 1995), el uso léxico (Bowerman 1978, 1980; Roberts & Horowitz 1986), etc. Como dice Labov (1973: 342), en muchos aspectos, el lenguaje es “una actividad categorizadora”.

En breve, gracias a la categorización y a las categorías cognitivas, cada estímulo nuevo se percibe, procesa, recuerda, actúa y se habla sobre él no como único, sino como un ejemplo de una categoría ya existente en la mente. La categorización implica una estrategia cognitiva básica, económica (Bornstein 1984: 315). La categorización une simultáneamente la unidad y la diversidad, lo constante y lo variable, y es central para un gran número de funciones mentales significativas críticas para la supervivencia de un organismo (Wisniewski & Medin 1991: 237).

No es sorprendente pues que el hombre haya sido definido como “un animal categorizador” (Labov 1973: 342) o como “el hombre clasificador” (Berlin *et al.* 1973: 214).

Sin embargo, aunque es básica para la cognición, propuestas extremas que reducen toda la cognición a un acto de categorización no parecen muy apropiadas. La hipótesis de que todas las operaciones cognitivas son simplemente actividades mentales categorizadoras es demasiado extrema. Por ejemplo, aunque es importante para la comprensión del lenguaje y del control sensorimotor, la categorización no es el componente más esencial de esos procesos (van Gelder 1993). Debería pues situarse a la categorización *dentro* de la cognición, pero la cognición no debería verse como una clase de categorización. De igual forma, es inútil reducir todo los actos del lenguaje a una actividad categorizadora ya que muchos de aquellos sirven muchas otras funciones al igual que muchos actos de la conducta no tienen un fin categorizador.

1.1.5. Tipos de Categorías.

Los diferentes tipos de categorías estudiados en la literatura sobre la categorización reflejan la heterogeneidad de los estímulos con los que nuestros sistemas cognitivos trabajan.

La mayor parte de la investigación realizada ha tratado con categorías comunes y concretas, que se dividen normalmente en dos grupos: categorías *naturales* (ej. clases naturales como “pájaros”, “gatos”, “árboles”, “hojas”, “verduras”, etc.) y categorías de *artefactos* (artefactos humanos como “herramientas”, “juguetes”, “coches”, “sillas”, “prendas”, “vehículos”, etc.).

Estos dos tipos de categorías han sido estudiados intensivamente debido a la facilidad de su uso ya que los miembros de estas categorías son normalmente familiares, fácilmente describibles y relativamente bien entendidos por los experimentadores y los sujetos experimentales (Tversky 1990: 335). Esto los hace atractivos para los experimentos en el laboratorio (Busemeyer *et al.* 1997: 2).

Aparte de estos dos tipos de categorías, un gran número de estudios ha utilizado categorías artificiales, que son aquellas construidas para un determinado experimento. Este tipo de categorías son un poderoso instrumento científico ya que permiten explorar fenómenos previamente observados en las categorías naturales o de artefactos. Esto es así porque las categorías artificiales son más fáciles de manipular ya que los atributos que especifican la pertenencia a la categoría de un determinado miembro pueden ser formulados precisamente y su contribución puede ser evaluada de una forma más directa.

Existen dos tipos de categorías artificiales. El primer tipo incluye categorías que imitan de cerca de las categorías naturales. Este tipo de categorías incluye dibujos esquemáticos de caras (ej. Goldman & Homa 1977; Kellogg 1981; Kemler Nelson 1984; Lamberts 1994, 1995; Martin & Caramazza 1980; Medin & Schaffer 1978; Murphy & Ross 1994; Neumann 1977;



Reed 1972, 1978), dibujos de animales imaginarios (Malt 1989), etc. El segundo tipo incluye categorías completamente arbitrarias, muy usadas en los experimentos de formación de conceptos de las eras conductivas y del procesamiento de la información. Los miembros de dicho tipo de categorías se construyen normalmente de un número pequeño de dimensiones perceptuales ortogonales y separables (ej. forma, color, talla) que varían según un pequeño número de atributos (ej. rojo, azul, verde, para la dimensión color). Estudios de este segundo tipo incluye los famosos patrones de puntos (Homa & Chambliss 1975; Homa & Cultice 1984; Homa *et al.* 1973; Homa & Vosburgh 1976; Homa *et al.* 1987; Knapp & Anderson 1984; Omohundro 1981; Posner 1969; Posner & Keele 1968, 1970; Posner *et al.* 1967; Robbins *et al.* 1978), figuras geométricas de todo tipo (ej. Bourne 1982; Dennis *et al.* 1973; Goldstone 1994a; Lamberts & Brockhoff 1997; Medin & Schwanenflugel 1981; Medin *et al.* 1984; Murphy & Ross 1994; Neumann 1974), etc.

A pesar de que las categorías naturales, de artefactos, y artificiales han sido utilizadas en multitud de estudios, los investigadores también se han centrado en otro tipo de categorías, como las sociales. Éstas son categorías que hacen referencia a grupos de edades, ocupaciones, religiones, razas, nacionalidades, etc. Se distinguen aquí tres tipos de categorías: grupos de *personas* como “nosotros”, “ellos”, etc. (Allen & Wilder 1975; Billig & Tajfel 1973; Brewer *et al.* 1981; Howard & Rothbart 1980; Nelson & Miller 1995; Smith & Zárate 1990; Tajfel & Billig 1974; Tajfel *et al.* 1971; Wilder 1981), *rasgos de personalidad*, como “combativo”, “extrovertido”, “hablador”, etc. (Andersen & Klatzky 1987; Borkenau 1990; Buss & Craik 1983; Canto & Mischel 1977; Chaplen *et al.* 1988; Hampson 1982; Isen *et al.* 1992; John *et al.* 1991; Read *et al.* 1990a, 1990b; Wojciszke & Pienkowski 1991), y *estereotipos* como “librero”, “trabajador”, “político”, etc. (Brewer *et al.* 1981; Canto & Mischel 1979a, 1979b; Cohen 1981; Dahlgren 1985; Hastie *et al.* 1990; Trentin & Salmaso 1990).

Además, una gran proporción de categorías referidas a entidades abstractas han sido estudiadas. Estas incluyen conceptos como “inteligencia” (Neisser 1979), emociones (Fehr 1988; Fehr & Russell 1984; Shaver *et al.* 1987), diagnósticos psiquiátricos (Canto *et al.* 1980; Horowitz *et al.* 1981; Murphy & Wright 1984), enfermedades físicas (Brooks *et al.* 1991; Medin *et al.* 1982; Shanks 1991), etc.

La literatura especializada ha tratado a su vez las similitudes y diferencias entre diferentes tipos de categorías con frecuencia. En la mayor parte de los casos, las comparaciones han sido realizadas entre las categorías naturales o de artefactos y algún otro tipo de categoría (ej. Barsalou 1983, 1985, 1987; Barsalou *et al.* 1986; Barsalou *et al.* 1987; Borkenau 1990; Canto & Mischel 1979; Dahlgren 1985; Medin & Smith 1984; Trentin & Salmaso 1990; Wattenmaker 1995).

1.2. Estudios de Categorización con Humanos Adultos.

Durante décadas, diferentes modelos de categorización han intentado proporcionar una explicación de la habilidad de categorizar en los humanos adultos. Estos modelos difieren en lo relativo a la naturaleza del aprendizaje, la representación mental de las categorías y los procesos usados en la clasificación.

El propósito del resto de esta sección es revisar la extensa literatura para proporcionar al lector una vista panorámica de los estudios sobre la categorización en adultos humanos (véanse también Barsalou 1992a; Barsalou & Hale 1993; Homa 1984; Komatsu 1992; Lengle *et al.* 1984; Medin 1989, 1999; Medin & Barsalou 1987; Medin & Smith 1984; Mervis 1980; Mervis & Rosch 1981; Smith & Medin 1981). Aunque existen diferencias entre submodelos específicos dentro de cada tipo, en general se pueden distinguir cuatro grandes tipos de estos: clásicos, probabilísticos, de ejemplares, y mixtos. Los modelos de categorización probabilísticos, de ejemplares, y mixtos representan alternativas que emergieron en oposición a los problemas detectados en los modelos clásicos, a los que se dedica ahora nuestra atención.

1.2.1. Modelos Clásicos de Categorización.

El modelo clásico de los conceptos se remonta a Aristóteles. Este modelo fue hasta hace poco la posición dominante en filosofía (ej. Margolis 1994; Rey 1983, 1985), la psicología (Bourne 1966, 1970; Bruner *et al.* 1956; Hull 1920; Hunt 1962; Neisser & Weene 1962), la lingüística (Chomsky & Halle 1968; Katz & Postal 1964), etc.

1.2.1.1. Aprendizaje, Representación Mental y Almacenamiento.

Según el modelo clásico aprender conceptos consiste en dos aspectos: primero, abstraer los atributos relevantes que definen la categoría de sus miembros y segundo, la creación de información que contiene las relaciones lógicas o reglas entre esos atributos. Además, la representación de un concepto o categoría es, según el modelo clásico, una abstracción que consiste en un número pequeño de atributos o rasgos compartidos por todas las entidades clasificadas como miembros de la categoría. Además, esta abstracción es una especificación de los rasgos que son, en su conjunto, suficiente e individualmente necesarios para pertenecer a la categoría.

Aunque se considera que los modelos clásicos han de requerir una pequeña capacidad de almacenamiento en la mente, lo cierto es que requieren una capacidad amplia ya que un concepto clásico no posee atributos necesarios y suficientes si hay excepciones a estos atributos. Por consiguiente, si una entidad es clasificada como miembro de una determinada categoría por

una fuente externa a pesar de no poseer todos los atributos necesarios y suficientes, una nueva abstracción ha de ser computada. Para asegurarse de que no hay excepciones pues, memorias de los ejemplares deben ser almacenadas para permitir una recomputación de las reglas que no son confirmadas por una entidad clasificada como miembro de una determinada categoría por una fuente externa.

1.2.1.2. Clasificación.

Según el modelo clásico, una vez que los categorizadores han aprendido una categoría, estos deberían poder determinar si una determinada entidad es miembro de la categoría simplemente determinando si la entidad posee todos los atributos definitorios de la categoría conceptual o no. Los categorizadores simplemente tienen que testar los posibles candidatos para ver si son miembros de la categoría con la abstracción de la categoría. Esta forma de clasificación requiere que cada entidad clasificada como ejemplo de la categoría cumpla con los criterios definitorios asociados con esa categoría. Una entidad es asignada a una categoría si y solo si posee todos los atributos definitorios de la representación abstracta. Si esto no es así, no será clasificada como ejemplo de la representación abstracta.

1.2.1.3. Decadencia del Modelo Clásico: Cuestionando el Paradigma.

Los modelos clásicos de categorización proporcionan explicaciones lógicas e intuitivas de la categorización. Sin embargo, en las últimas décadas, la visión clásica de la categorización ha sido fuertemente cuestionada fundamentalmente por los resultados experimentales obtenidos desde los años 70 y que parecen contradecir las bases del modelo clásico.

La incompatibilidad del modelo clásico con los resultados experimentales mencionados puede abordarse examinando las críticas que se han vertido sobre el modelo clásico a lo largo de los años. Estas críticas hacen referencia a la imposibilidad de obtener atributos definitorios, al uso de atributos innecesarios en la clasificación, la existencia de miembros de categorías dudosos, de la tipicidad y los efectos de tipicidad y de la inconsistencia de muchos fenómenos relacionados con la organización taxonómica del modelo clásico.

1.2.1.3.1. Imposibilidad de Especificar Rasgos Definitorios.

El eje central del modelo clásico es su supuesto de que cada concepto posee una serie de rasgos necesarios y suficientes. Sin embargo, décadas de análisis han mostrado que tales rasgos definitorios son prácticamente imposibles de obtener para la mayor parte de las categorías (ver, por ejemplo Smoke 1933; Ryle 19451; Wittgenstein 1953).

1.2.1.3.2. Uso de Atributos Innecesarios en la Categorización.

Según el modelo clásico, la posesión de rasgos no definitorios por los miembros de una categoría es irrelevante para su pertenencia en la categoría y estos rasgos no deberían usarse a la hora de clasificar una determinada entidad como miembro de una categoría. Sin embargo, un creciente número de estudios muestran que los atributos “no necesarios” se usan en el proceso de la clasificación (ej. Hampton 1979; Malt 1993, 1994; Malt & Johnson 1992; Rips *et al.* 1973).

1.2.1.3.3. Miembros de Categorías Dudosos.

Al asumir que los juicios acerca de la pertenencia o no de una entidad a una categoría se basan en propiedades o atributos definitorios, este tipo de juicios no debería llevar a errores o a variaciones en la clasificación de una entidad como miembro de una determinada categoría. Los categorizadores simplemente tienen que comparar los rasgos definitorios de la entidad que ha a ser clasificada con los de la descripción abstracta de la categoría. Si dicha entidad reúne todos los rasgos necesarios y suficiente del concepto, será un miembro de la categoría; si no lo hace, no será miembro de la misma. No hay razones para esperar inconsistencia en la clasificación.

Sin embargo, esta predicción frecuentemente falla. Las categorías tienen miembros relativamente ambiguos en el sentido de que no son clasificados de la misma manera el 100% de los casos como muestra un creciente número de estudios (ej. Barsalou 1983; Bellezza 1984; Brownell & Caramazza 1978; Hampton 1979; Kempton 1978, 1981; Labov 1973; McCloskey & Glucksberg 1978; Smith *et al.* 1974; etc.).

1.2.1.3.4. Tipicidad.

De todas las críticas vertidas en contra del modelo clásico destaca la de la tipicidad, que se refiere a lo representativo que es un miembro de una determinada categoría como miembro de esa categoría. Así, un ejemplo como “gorrión” parece ser mejor ejemplo del concepto de “pájaro” que “cigüeña” o que “pato”.

Los grados de tipicidad son algo en lo que los sujetos suelen estar muy de acuerdo como demuestran infinidad de estudios con toda clase de categorías en los que se ha observado este fenómeno (ej. Hampton & Gardiner 1983; Malt 1994; McCloskey & Glucksberg 1978; Rips *et al.* 1973; Smith *et al.* 1974; etc.).

1.2.1.3.5. Fenómenos Taxonómicos.

Las categorías no ocurren independientes unas de otras, sino que aparecen organizadas en sistemas en los que se relacionan de diversas formas. Debería remarcarse, relativo a esto, que la organización jerárquica es una manera particularmente eficiente de almacenar la información y que una gran parte de nuestra actividad cognitiva parece estar organizada jerárquicamente en taxonomías. Una taxonomía refleja una subdivisión del conocimiento en clases. En una taxonomía los conceptos se relacionan por la relación “es un tipo de”. En este sentido es importante mencionar que el estudio de las taxonomías populares en los hablantes de sociedades no occidentales y precientíficas ha mostrado que la visión del modelo clásico de la organización taxonómica no es del todo correcta como un reflejo de una teoría cognitivamente adecuada de aquel tipo de organización. En primer lugar, las taxonomías clásicas tienen un número de niveles de abstracción muy superior al de las taxonomías de la gente corriente (ej. Berlin 1978, 1992; Berlin *et al.* 1973; Brown *et al.* 1976; Rosch 1978; Rosch *et al.* 1976a). En segundo lugar, las taxonomías clásicas se caracterizan por su elegancia y exhaustividad, sin niveles incompletos. Sin embargo, la investigación antropológica muestra que la completa elaboración taxonómica rara vez ocurre (Berlin 1978, 1992; Berlin *et al.* 1973; véase también Brown *et al.* 1976). Aunque otras muchas críticas se han vertido sobre el modelo clásico de la organización taxonómica es sin duda la crítica relacionada con la existencia de un “nivel básico” la que más frecuentemente se ha aducido contra la visión clásica. Por nivel básico se entiende ese nivel de abstracción en una taxonomía que parece gozar de una prominencia cognitiva mayor que ningún otro nivel, tal y como muestran decenas de estudios que tienen que ver con el procesamiento de la información categorial de las categorías en cada nivel de abstracción, la percepción de los miembros de las categorías según el nivel de abstracción al que se adscriben, el procesamiento de las categorías pertenecientes a diversos niveles de abstracción, o incluso la adquisición de las categorías. En suma, parece que uno de los niveles de organización taxonómica (el correspondiente a *perro* en la taxonomía que va desde el concepto abstracto de “forma de vida” hasta una variedad muy específica de una raza de perro o incluso a un ejemplar concreto de esa raza) es más básico o fundamental a la hora de clasificar entidades, procesarlas o tratar con ellas de alguna forma. Según el modelo clásico, ningún nivel de abstracción debería, sin embargo, ser más básico que cualquier otro. Al mostrar diversos estudios experimentales que esto no es así, se cuestiona también pues el modelo clásico de la organización taxonómica.

1.2.1.4. La Utilidad del Modelo Clásico.

Las críticas al modelo clásico son fuertes, pero ninguna lo invalida totalmente. De hecho, son frecuentes las réplicas a tales críticas por defensores del modelo clásico. Debido a su larga

tradición, muchos investigadores han intentado defender aquel modelo tanto en la psicología cognitiva (ej. Armstrong *et al.* 1983; Gleitman *et al.* 1983; Martin & Caramazza 1980), la filosofía (ej. Margolis 1994; Rey 1983, 1985; Sutcliffe 1993), la lingüística (ej. Wierzbicka 1990), etc. Estos autores mantienen que el modelo clásico ha sido rechazado muy prematuramente. Sin embargo los intentos más prometedores de “salvar” al modelo clásico, englobados bajo el llamado modelo de “core-identification procedimiento” (ej. Armstrong *et al.* 1983; Gleitman *et al.* 1983; McNamara & Sternberg 1983; Medin & Smith 1984; Rey 1983, 1985; Smith & Medin 1981; Smith *et al.* 1984; Smith *et al.* 1974; Osherson & Smith 1981, 1982) han resultado infructuosos también. Este intento propone que en la representación mental existen dos tipos de atributos: los definitorios o centrales (“core”) que siguen las predicciones del modelo clásico y los no definitorios que son característicos de los miembros de la categoría pero no necesariamente aplicables a cada uno de los miembros y que se usan frecuentemente para realizar juicios de tipicidad, clasificar entidades, etc., y que explicarían los hallazgos aducidos en contra el modelo clásico. Sin embargo, esta moderna versión del modelo clásico no carece de problemas. El más llamativo es, sin duda la imposibilidad de distinguir experimentalmente entre atributos centrales necesarios y atributos meramente característicos (ej. Barsalou & Medin 1986; Barsalou *et al.* 1987; Malt 1994; Malt & Johnson 1992; Murphy 1995).

El modelo clásico parece poco válido como una explicación general de la habilidad de categorizar del ser humano. Sin embargo, parece que en algunos como en el de “diagnóstico diferencial” (Schank *et al.* 1986) sí parece tener sentido. En el diagnóstico diferencial, los atributos necesarios y suficientes son útiles para garantizar que los miembros de una categoría sean correctamente asignados a la categoría en cuestión. Esto sucede cuyo la exactitud en la clasificación es la máxima prioridad en la categorización. A esto se unen las frecuentes referencias hechas por los psicólogos cognitivos (ej. Estes 1986a; Kemler Nelson 1984; Medin & Smith 1984; Rosch 1977; Smith & Medin 1981) o lingüistas (ej. Lakoff 1987a; Ungerer & Schmid 1996) a que aunque se trata de un modelo insuficiente como una teoría general de la categorización humana, el modelo clásico cobra sentido cuyo hay una necesidad de definiciones y clasificaciones técnicas precisas y rígidas en contextos educacionales, científicos, o legales. Esto ocurre en las ciencias matemáticas, la geometría, la biología, la física o incluso en el sistema legal.

1.2.1.5. Modelos Alternativos de Categorización: Modelos Basados en la Similitud.

Debido a la limitada cantidad de evidencia experimental que apoya al modelo clásico y a que la gran parte de la evidencia experimental obtenida contradice sus presupuestos, un paulatino paso del modelo clásico a modelos de categorización alternativos como el

probabilístico, el de ejemplares o el mixto se produjo desde principios de los 70. Estas alternativas suponían modelos prometedores porque eran, entre otras cosas “experimentales” y “experienciales”. Son experimentales en el sentido de que derivan de la investigación empírica, al contrario que el modelo clásico, que era una posición filosófica no empírica. Además estos modelos era “experienciales” en el sentido propuesto por Lakoff (1982, 1987a) mientras que los modelos clásicos eran “objetivistas”.

El modelo clásico asumía que los atributos definitorios eran un reflejo de los atributos supuestamente necesarios y suficientes de los miembros de categorías supuestamente objetivas existentes en el mundo. Ya que el mundo “objetivo” contenía, según el modelo clásico, diversos tipos de entidades que compartían propiedades definitorias, la tarea de los sujetos consistía en encontrar esas propiedades necesarias y suficientes y representarlas en la memoria. En otras palabras, las categorías de la mente reflejaban las categorías objetivas del mundo. Sin embargo, los modelos alternativos de categorización eran experiencialistas porque asumían que las categorías conceptuales estaban determinadas por la estructura del mundo *además de* la experiencia e interacción individual siempre dentro del marco de nuestra experiencia corporal y organización cognitiva y cultural. Además, las categorías cognitivas no tenían que reflejar necesariamente ninguna organización supuestamente objetiva y predefinida en el mundo real sino que reflejaban una mezcla de estructura tal y como es percibida en el mundo y las estructuras cognitivas del sujeto. Los modelos alternativos de categorización son experiencialistas porque reflejan la experiencia de la gente y su conocimiento cambiante acerca del mundo, mientras que el modelo clásico asumía que, una vez conocidos los rasgos definitorios, estos no cambiarían mientras que los modelos alternativos asumen que las representaciones mentales de las categorías varían dinámicamente con la experiencia (ej. Barsalou & Medin 1986).

Además de su punto de vista “experiencialista”, la mayor asunción de los modelos de categorización propuestos en los 70 y 80 es que la categorización era motivada por la similitud y no por la satisfacción de todas y cada una de los atributos necesarios y suficientes por parte de las entidades que eran categorizadas y la representación mental de la categoría. Se consideraba a la similitud como un poderoso principio que podía explicar la mayor parte de los problemas que habían debilitado al modelo clásico. La similitud no tenía ningún tipo de papel en los modelos clásicos ya que según estos, una entidad pertenece a una categoría sólo si satisface estrictamente la regla de la categoría. Una entidad no puede ser un miembro de la categoría si es parecido a la regla. Por el contrario, la entidad debe de satisfacer la regla perfectamente, siendo una satisfacción parcial insuficiente.

1.2.2. Modelos Probabilísticos de Categorización.

El primer grupo de modelos que constituyen una alternativa al modelo clásico es el de los modelos probabilísticos de categorización, que aún comparten muchas presupuestos con el modelo clásico pero que difieren de él en aspectos fundamentales.

1.2.2.1. Aprendizaje, Representación y Almacenamiento Mentales.

Los modelos de categorización probabilísticos asumen que el categorizador calcula un resumen estadístico que describe a la categoría en su totalidad al tiempo que está expuesto a diversos miembros de dicha categoría. Como el modelo clásico, los modelos probabilísticos asumen que la representación de un concepto es una descripción de una clase entera que se aplica a todos los miembros de la categoría. Sin embargo, al contrario que los modelos clásicos, la representación no necesita incluir atributos necesarios y suficientes. Por el contrario, la representación incluye atributos que son representativos o característicos de los ejemplares de la categoría, atributos que son bastante probables en los miembros de la categoría.

A pesar de que los modelos probabilísticos comparten esta asunción, se diferencian normalmente en el tipo de información que se supone el aprendiz calcula y almacena. La mayor parte de estos modelos asumen que la información se basa en una tendencia central de las propiedades de los miembros de la categoría. Esta tendencia puede incluir medias (ej. Estes 1986b; Homa & Chambliss 1975; Posner 1969; Posner & Keele 1968, 1970; Posner *et al.* 1967; Reed 1972, 1978; Reed & Friedman 1973; Rips *et al.* 1973; Rosch *et al.* 1976), modas (ej. Collins & Loftus 1975; Estes 1986b; Goldman & Homa 1977; Hampton 1979; Hayes-Roth & Hayes-Roth 1977; Kellogg 1981; McCloskey & Glucksberg 1979; Neumann 1974; Rosch *et al.* 1976; Smith *et al.* 1974), o ambas (ej. el modelo “componente” de Smith & Medin 1981), etc.

1.2.2.2. Clasificación.

Los modelos probabilísticos acometen el acto de la clasificación rápidamente ya que, según ellos, no se necesita examinar un gran número de ejemplares sino tan solo la descripción calculada a la hora de aprender la categoría. Una vez que se ha accedido a esa descripción, el sistema cognitivo del categorizado compara la descripción estructural de la entidad que va a ser categorizada con la representación categorial a la que se ha accedido. Las entidades que superan un determinado nivel de similitud con la representación abstracta son entonces clasificadas como miembros de la categoría mientras que aquellas que no llegan a ese nivel no lo son.

1.2.2.3. Ventajas y Desventajas de los Modelos Probabilísticos de la Categorización.

Los modelos probabilísticos explican la mayor parte de los fenómenos que menoscababan a los modelos clásicos. Por ejemplo, los modelos probabilísticos no requieren atributos definitorios sino característicos o probables. En segundo lugar, ya que dichos modelos no requieren atributos definitorios, la clasificación de nuevas entidades como miembros de la categoría puede acometerse por medio de atributos que son meramente característicos y así el uso de atributos no necesarios en la categorización se explica satisfactoriamente. En tercer lugar, los modelos probabilísticos mantienen que los “casos dudosos” aparecen cuando la entidad que se ha de clasificar es lo suficientemente similar a las representaciones de más de un concepto o categoría. En cuarto lugar, el modelo probabilístico explica la tipicidad diciendo que ésta refleja lo similar que es un determinado miembro de una categoría a la representación probabilística de dicha categoría. A mayor similitud, mayor tipicidad. Los modelos probabilísticos también explican los fenómenos taxonómicos problemáticos para el modelo clásico pero limitaciones espaciales no permiten la más que larga explicación a la resolución de dicho problema por los modelos probabilísticos.

A pesar de las soluciones exitosas de los modelos probabilísticos a los problemas que pesaban sobre el modelo clásico, aquellos no están exentos de problemas. Entre ellos destacan el problema del uso de ejemplares específicos en la clasificación y/o su almacenamiento mental, hallazgos estos muy documentados y para los que los modelos probabilísticos carecen de explicación. En segundo lugar, se acusa a los modelos probabilísticos de no capturar todo el conocimiento que los categorizadores tienen acerca de las categorías, como es el caso de los atributos correlacionados (ej. Malt & Smith 1984; Medin & Schwanenflugel 1981; Medin & Shoben 1988; Medin *et al.* 1982; Medin *et al.* 1987; Murphy & Wisniewski 1989b; Wattenmaker 1991; Wattenmaker *et al.* 1986). En tercer lugar, los modelos probabilísticos no explican los efectos contextuales en el uso de los conceptos (Michalski 1993; Smith & Miden 1981) a pesar de la amplia evidencia existente al respecto.

1.2.3. Modelos de Ejemplares de Categorización.

En general la psicología ha tenido considerable fe en la existencia de mecanismos mentales de abstracción mecánica e inconsciente. Sin embargo, muchos estudios sugieren que en muchos casos la abstracción de regularidades puede estar muy lejos de ser automática y dudan de que incluso cuando las regularidades son conocidas, éstas sean suficientes para explicar el comportamiento categorizador. Esto es precisamente la asunción subyacente a la segunda alternativa a los modelos clásicos de categorización y conocida como el modelo o modelos de ejemplares. Los modelos de ejemplares no requieren la abstracción de información

de resumen nueva no dada previamente en alguno de los ejemplares sino que asumen que los miembros de la categoría que se encuentran son almacenados en la memoria. Un número importante de estudios parecen apoyar tal asunción (ej. Brooks 1978, 1987; Brooks *et al.* 1991; Estes 1986b; Heit 1992, 1994; Hintzman 1986; Hintzman & Ludlam 1980; Kossan 1981; Kruschke 1992; Lamberts 1994, 1995, 1998; Medin & Schaffer 1978; Reber & Allen 1978; Smith & Medin 1981; Walker 1975; Wattenmaker 1991; Whittlesea 1987, 1997).

1.2.3.1. Aprendizaje, Representación y Almacenamiento Mentales.

Según una versión genérica de los modelos de ejemplares de categorización, el categorizador no calcula un resumen abstracto cuando aprende la categoría sino que los sujetos, expuestos a los miembros de la categoría, memorizan descripciones individuales de esos ejemplos.

En general, los modelos de ejemplares mantienen que las categorías almacenadas en la memoria no incluyen generalizaciones de resumen abstraídas de los ejemplares pero asumen que, si lo hacen, tales generalizaciones no están tan robustamente codificadas en la memoria como las memorias de ejemplares y que no se usan en la clasificación. Los modelos de ejemplares mantienen que una categoría o concepto está principalmente representado/a por descripciones separadas de algunos de sus ejemplares específicos.

1.2.3.2. Clasificación.

Según los modelos de ejemplares de categorización, la gente clasifica nuevas entidades en base a la similitud de éstas a las memorias de ejemplares previamente experimentados, almacenados en la memoria a los que se accede a la hora de clasificar. Las entidades se asignan a la categoría que posea como media los ejemplares más parecidos a las entidades que se han de clasificar.

1.2.3.3. Ventajas y Desventajas de Modelos de Ejemplares de Categorización.

Los modelos de ejemplares de categorización proporcionan una explicación a los problemas a los que se enfrentaba el modelo clásico. En primer lugar, ya que no hay razón por la cual ejemplares diversos tengan por qué tener las mismas propiedades, no se espera la existencia de atributos definitorios. El único requisito es que las propiedades caractericen al menos a un ejemplar. En segundo lugar, ya que los atributos de los miembros de la categoría no tiene por qué ser verdaderos de otro miembro de la categoría, cualquier atributo o conjunto de atributos puede usarse para clasificar a una determinada entidad como miembro de una

categoría. El uso de propiedades no necesarias en la clasificación se explica así fácilmente. En tercer lugar, según los modelos de ejemplares, los casos dudosos aparecen cuando un objeto es similar a las memorias de ejemplares de más de una categoría o cuando un objeto no es muy similar a los ejemplares de varias categorías. En cuarto lugar, los modelos de ejemplares explican la tipicidad de la siguiente manera: cuanto más parecida es una determinada entidad a los miembros de su categoría y menos a los miembros de categorías opuestas, más alta será su tipicidad (Nosofsky 1988b). Finalmente, el modelo de ejemplares explica los fenómenos taxonómicos problemáticos para el modelo clásico de una forma satisfactoria que por razones de espacio no trataremos.

A pesar de proporcionar una explicación más o menos exitosa a los problemas del modelo clásico, los modelos de ejemplares también presentan ciertos problemas como la imposibilidad de explicar satisfactoriamente las numerosas abstracciones que, se ha demostrado, la gente induce de los ejemplares y usa (Barsalou 1992a; Michalski 1993). Los modelos de ejemplares tampoco explican la existencia de atributos correlacionados o estructuras de conocimiento más complejas (Medin 1989; Medin & Schaffer 1978), ni los efectos contextuales. Además, los modelos de ejemplares tienen problemas específicos que los modelos probabilísticos no tenían, como su inexplicable capacidad de aprender categorías nuevas sin haber experimentado anteriormente ejemplares de las mismas o su falta de restricciones a las propiedades que pueden entrar en la representación conceptual (o incluso en lo que puede llegar a constituir un concepto). Tampoco explican muy bien el hecho de que presuponen una capacidad memorística ilimitada aunque diversos autores parecen converger en la necesidad de proponer alguna forma de restringir el teóricamente infinito número de ejemplares que podrían entrar en la representación de categoría.

Los modelos de ejemplares tienen, sin embargo, ventajas con respecto a los probabilísticos como su capacidad de explicar el aprendizaje de categorías cuyos ejemplares varían enormemente y tienen poco en común (Barsalou 1992a: 27).

1.2.4. Los Modelos Mixtos de Categorización.

Los modelos probabilísticos y de ejemplares se han contrastado tradicionalmente como si fuesen mutuamente exclusivos. Recientemente, sin embargo, los investigadores han mantenido que el aprendizaje, almacenamiento mental y clasificación parecen explicarse mejor si se asume que tanto generalizaciones como ejemplares se encuentran almacenados en la memoria. De hecho, algunos modelos de categorización mantienen expresamente que la representación de una categoría incluye abstracciones en forma de resumen y memorias de ejemplares (ej. Barsalou 1990b; Bussemeyer *et al.* 1984; Homa 1984; Homa *et al.* 1981; Malt 1989; Medin

1986; Medin & Smith 1981; Medin *et al.* 1984; Medin *et al.* 1983; Smith & Medin 1981; Smith & Zárate 1990).

1.2.4.1. Aprendizaje, Representación Mental, Almacenamiento, y Clasificación.

Los modelos mixtos asumen que, al estar expuesto a nuevos ejemplares de una categoría, tanto memorias de esos ejemplares específicos como una abstracción que describe algún tipo de tendencia central son aprendidos y almacenados en la memoria. Lo que es más importante, estos modelos asumen que cada uno de los dos tipos de información se encuentra codificado con la misma intensidad. Los modelos mixtos mantienen que, a la hora de tomar decisiones con respecto a la pertenencia o no a una categoría, la representación abstracta es más accesible en algunos casos mientras que en otros lo son las memorias de ejemplares. En consecuencia, las entidades que se han de categorizar se comparan tanto con la representación categorial como con las diferentes memorias de ejemplares para llevar a cabo la clasificación.

1.2.4.2. Ventajas y Desventajas de los Modelos Mixtos de Categorización.

Los modelos mixtos explican los hallazgos problemáticos para el modelo clásico al integrar las explicaciones a tales fenómenos de los modelos probabilísticos y los modelos de ejemplares. Sin embargo, aunque los modelos mixtos constituyen un esfuerzo por reconciliar visiones de la categorización aparentemente contrapuestas, los mismos problemas en aquellos modelos de categorización reaparecen en los modelos mixtos.

1.2.5. Insuficiencia de los Modelos Basados en la Similitud.

Parte de la razón por la que los modelos de categorización alternativos al modelo clásico no son lo suficientemente satisfactorios es por su casi exclusiva dependencia del principio de “similitud”. En apariencia, la similitud es un poderoso mecanismo pero no suficiente para explicar todo lo que se conoce acerca de la categorización. El intento de explicar los fenómenos que los modelos de categorización basados en la similitud no solucionan satisfactoriamente motivó, al menos en parte, el segundo de los dos cambios mayores en la investigación de la categorización a lo largo de este siglo (Busemeyer *et al.* 1997; Medin 1989; Murphy 1993b). El primer cambio es de la asunción de que las representaciones conceptuales tienen propiedades definitorias (el modelo clásico) a la idea de que contienen propiedades que son sólo características de los miembros de la categoría (los modelos probabilístico, de ejemplares, y mixto). El segundo cambio consiste en considerar las representaciones conceptuales como

organizadas alrededor de estructuras de conocimiento más complejas llamadas “teorías” en vez de estar exclusivamente estructuradas por el principio de similitud.

Debería mencionarse aquí que, a pesar de diferir en la mayor parte de sus características, los modelos propuestos como alternativas al modelo clásico comparten diversos presupuestos. En primer lugar la coherencia categorial se basa en la similitud: la asociación de los diferentes ejemplares de una categoría parece coherente porque tales ejemplares son similares los unos a los otros y diferentes de los miembros de otras categorías. En segundo lugar, la clasificación se guía también por el principio de similitud. La clasificación implica un proceso de comparación de atributos y una computación de similitud de atributos que el sistema cognitivo acomete entre la entidad que ha de ser clasificada y la representación de la categoría (abstracción, memorias de ejemplares, o ambos) con la que la entidad se compara. La probabilidad de asignar a una determinada entidad a una categoría depende de que dicha entidad alcance o no un determinado umbral de similitud con la representación de la categoría. En tercer lugar, las representaciones en forma de resumen o las memorias de ejemplares están compuestas de colecciones de atributos relativamente independientes. En cuarto lugar la categorización basada en el principio de similitud asume que la memoria contiene representaciones invariables y estáticas y a las que se accede y que son usadas en todas las ocasiones en las que una categoría es procesada. Finalmente, los conceptos se consideran como independientes los unos de los otros y las pocas relaciones existentes entre ellos también se basan en la similitud.

Sin embargo, los investigadores de la categorización pronto comenzaron a encontrar muchos problemas con estos principios que finalmente llevaron a un cambio muy importante en la concepción de los mecanismos de la categorización.

1.2.5.1. La Emergencia de la Explicación Basada en “Teorías” de la Categorización.

La insatisfacción con los modelos basados en el principio de similitud y sus presupuestos llevó al surgimiento de la visión de la categorización y de la estructura conceptual basada en las “teorías”, lo cual constituye el segundo mayor cambio de orientación en la investigación acerca de la categorización de este siglo. Esta nueva visión consisten en contemplar a los conceptos como organizados alrededor de “teorías” y no exclusivamente basados en relaciones de similitud de atributos. Por “teorías” ha de entenderse el conjunto de creencias que la gente tiene acerca de las interrelaciones y conexiones causales entre los atributos de un concepto, entre los atributos y un concepto o entre los propios conceptos. Tales teorías pueden ser estereotipadas, breves, incorrectas (Rips 1995: 76) y se conciben como un conjunto de explicaciones mentales más que una explicación completa, organizada y científica (Murphy & Medin 1985: 290). El surgimiento de la visión basada en las “teorías” de la categorización se ha visto acompañada de un ingente conjunto de estudios experimentales que prueban que tales teorías son necesarias

para explicar los problemas que los modelos basados en las relaciones de similitud no conseguían resolver. El papel que juegan las “teorías” en los juicios de tipicidad (ej. Honeck *et al.* 1985; Rips 1991; Rips & Collins 1993), el aprendizaje de los conceptos o categorías (ej. Heit 1994; Kaplan & Murphy 1999; Medin *et al.* 1987; Murphy & Allopenna 1994; Nakamura 1985; Pazzani 1991; Spalding & Murphy 1996; Wattenmaker *et al.* 1986; Wisniewski 1995; Wisniewski & Medin 1994, 1991), la clasificación de entidades como miembros de determinadas categorías (ej. Keil 1987; Len & Murphy 1997; Len *et al.* 1997; Malt 1990; Pazzani 1991; Rips 1989; Smith & Sloman 1994; Wattenmaker 1995), y otros muchos fenómenos han sido suficientemente demostrados experimentalmente como llevar a los expertos a concluir que una explicación global de la categorización no puede pasar por alto las estructuras conceptuales del conocimiento que la gente utiliza diariamente para realizar todas las operaciones relacionadas con los conceptos y la clasificación.

1.2.6. Modelos de Categorización: Comentarios Finales.

Aunque existe un consenso general de que la categorización es una habilidad humana fundamental, no hay, sorprendentemente, un acuerdo global acerca de la naturaleza de tal habilidad (Medin 1989). Esto se debe en parte al hecho de que los experimentos rara vez han comparado modelos de categorización explícitamente formulados (excepcionalmente Elio & Anderson 1981; Estes 1986b; Hayes-Roth & Hayes-Roth 1977; Martin & Caramazza 1980; Reed 1972) o a que cuando tales comparaciones se han realizado, los resultados no han permitido dilucidar qué modelos explican mejor los resultados obtenidos. Además, hoy en día casi cualquier modelo de categorización puede construirse con la suficiente precisión como para explicar los fenómenos observados de una manera satisfactoria.

En general, sin embargo, parece razonable concluir que todas las presupuestos acerca del aprendizaje, modos de almacenamiento mental y clasificación defendidos en los diferentes modelos de categorización se encuentran, hasta cierto punto, disponibles en el sistema cognitivo y que todos son necesarios para proporcionar una explicación satisfactoria de la categorización.

1.3. La Investigación acerca de la Categorización con Otras Poblaciones.

La mayor parte de la investigación acerca de la categorización se ha centrado en los sujetos humanos adultos. No obstante, también se han llevado a cabo un gran número de estudios con humanos de menor edad así como con especies no humanas. Este conjunto de estudios es importante porque arroja una perspectiva diferente sobre la capacidad humana de categorizar (al menos en los adultos). Asimismo, un resumen panorámico de la categorización estaría incompleto si no se mencionasen algunas líneas referentes a estos estudios.

1.3.1 La Categorización en los Niños.

El periodo que va desde el comienzo de la adquisición del lenguaje hasta la pubertad representa uno de los más complejos e intrigantes por diferentes razones. Entre estas últimas se encuentra el aprendizaje de categorías que ha sido estudiado desde principios de los años 70 con gran intensidad. Hoy en día, y a través de numerosos estudios con niños en los cuales se utilizan diversas técnicas experimentales camufladas bajo la apariencia de juegos en los cuales los niños han de nombrar los miembros de diversas categorías u ordenar reproducciones de dichos miembros en diferentes grupos, por ejemplo, se han podido constatar el aprendizaje en el laboratorio de categorías naturales como plantas, flores, animales, etc. (ej. Blewitt 1994; Blewitt & Durkin 1982; Callanan 1983; Callanan & Markman 1982; Daehler *et al.* 1979; Déak & Bauer 1995, 1996; Gelman 1984, 1988; Gelman & Coley 1990; Gelman & Markman 1986, 1987; Gelman & O'Reilly 1988; Gelman *et al.* 1986; Tversky 1989), categorías de artefactos como muebles, vehículos, instrumentos, etc. (ej. Blewitt 1994; Blewitt & Durkin 1982; Callanan 1983; Callanan & Markman 1982; Daehler *et al.* 1979; Déak & Bauer 1995, 1996; Gelman 1984, 1988; Gelman & O'Reilly 1988; Hupp & Mervis 1982; Melkman *et al.* 1981; Mervis 1984; Mervis & Mervis 1982; Nelson 1973) o incluso categorías artificiales de diversos tipos (ej. Horton & Markman 1980; Kossan 1981; Markman *et al.* 1980; Merriman *et al.* 1991). Dichos estudios han arrojado considerable luz sobre aspectos ya estudiados con los adultos como la existencia de casos dudosos, de distintos niveles de abstracción en las taxonomías y la existencia de grados de tipicidad entre los miembros de una misma categoría, etc.

1.3.2. La Categorización en los Bebés.

Los psicólogos han vuelto su atención también hacia el estudio de la capacidad de categorizar de los bebés prelingüísticos. Tal interés es importante para un mejor entendimiento de las primeras etapas del desarrollo cognitivo (Eimas 1994) o del más temprano estadio de la adquisición del vocabulario (ej. Gopnik & Meltzoff 1987, 1992; Roberts & Horowitz 1988; etc.).

Aunque el estudio de la categorización en los bebés puede parecer metodológicamente difícil debido a la corta edad de los sujetos, recientes avances metodológicos han facilitado tal estudio a través, por ejemplo, de la técnica del habituamiento visual y posterior observación de la preferencia visual por la novedad en el estímulo (ej. Cohen & Strauss 1979; Cohen & Younger 1983; Colombo *et al.* 1987; Quinn 1987; Roberts & Horowitz 1986; Ross 1980; Strauss 1979). Con esta técnica, los bebés son familiarizados en un primer término con diferentes miembros de una determinada categoría. Posteriormente se les enseña un ejemplar

nuevo de la categoría o un ejemplar nuevo de una categoría diferente o ambos al mismo tiempo. Si los bebés aprendieron la categoría al principio, normalmente pasan más tiempo mirando un estímulo nuevo de la segunda (y nueva) categoría que el estímulo de la categoría ya familiar ya que aquél es más atrayente por su absoluta novedad.

Usando este u otros procedimientos, numerosos estudios han proporcionado abundante evidencia de que los bebés hasta de tres meses pueden aprender categorías naturales, de artefactos, e incluso artificiales al ser expuestos a objetos físicos de esas categorías o a diapositivas que los representan. Entre las categorías naturales de las que se tiene evidencia experimental se encuentran las de “gatos”, “perros”, “pájaros”, etc. (ej. Behl Chadha 1994, 1996; Eimas & Quinn 1994; Eimas *et al.* 1994; Mandler & Bauer 1988; Quinn & Eimas 1996a; Quinn *et al.* 1993; Younger 1990a). Entre las categorías de artefactos tenemos las de “vehículos” (ej. Mandler & Bauer 1988; Mandler *et al.* 1991; Nelson 1973; van de Walle 1997), “muebles” (ej. Behl Chadha 1994, 1996; Mandler *et al.* 1991), etc. Ejemplos de categorías artificiales son las de formas geométricas (ej. Bomba & Siqueland 1983; Quinn 1987; Younger & Gotlieb 1988), dibujos de caras humanas esquemáticas (ej. Sherman 1985; Strauss 1979; Younger 1992), etc.

1.3.3. La Categorización en las Especies no Humanas.

Parece hoy en día, por la evidencia que se tiene, que hasta las criaturas más simples pueden discriminar entre sustancias nocivas y no nocivas, entornos peligrosos o no peligrosos, etc. Un número de estudios en aumento apoya, sin embargo, la afirmación de que la habilidad de categorizar no se encuentra limitada a la especie humana. La pregunta de si los animales pueden adquirir conceptos o no ha sido reemplazada hoy en día por qué especies pueden adquirir qué tipo de conceptos y bajo qué condiciones. En este sentido, cabe resaltar que la mayor parte de los estudios de categorización con especies no humanas se han centrado en las habilidades categorizadoras de las aves y de los primates.

Diversos estudios demuestran que varios tipos de especies de aves tales como los periquitos (Trillmich 1976), los pollos (ej. Ryan & Lea 1994) o el loro gris Africano (Peppenberg 1987) pueden categorizar muchos estímulos visuales. Sin embargo, se trata de las habilidades categorizadoras de los palomos de las que más se ha investigado. Un ingente número de estudios demuestran que estas aves pueden categorizar una diversa variedad de estímulos como las categorías “humanos” (Astley & Wasserman 1992; Edwards & Honig 1987; Herrnstein *et al.* 1976; Malott & Siddall 1972; Siegel & Honig 1970), otros palomos (Poole & Lander 1971), peces (Herrnstein & de Villiers 1980), gatos (Bhatt 1988; Bhatt *et al.* 1988; Wasserman *et al.* 1988), árboles (Herrnstein 1979, 1985), etc. Se ha demostrado también que los palomos pueden aprender categorías de artefactos e incluso categorías construidas con estímulos

artificiales (ej. Huber & Lenz 1993, 1996; Lea & Ryan 1990; Lea *et al.* 1993; etc.) o incluso categorías tan complejas como los conceptos “similar-diferente” (Cook *et al.* 1995; Cook *et al.* 1997; Santiago & Wright 1984), “familiar-nuevo” (MacPhail & Reilly 1989), etc. Todos estos estudios muestran que la categorización en las aves es un fenómeno central y no marginal y que es esencial en ellas como lo es en los humanos.

Otro gran número de estudios muestra que los primates también son buenos categorizadores. Varias especies de monos han sido estudiadas y se ha constatado la facilidad que dichos primates tienen para aprender tanto categorías naturales (ej. Roberts & Mazmanian 1988; Schrier *et al.* 1984; Yoshikubo 1985) como categorías de artefactos (ej. Morgan *et al.* 1976; Schrier *et al.* 1984) o incluso artificiales (ej. Jitsumori 1994).

1.3.3.1. Comparación entre Especies.

Al igual que el estudio con niños y bebés ha arrojado considerable luz sobre las diferencias y similitudes entre la habilidad de categorizar en los adultos y en aquellos grupos de sujetos de menor edad, la investigación sobre las habilidades categorizadoras de las especies no humanas ha mostrado también su interés (y ha arrojado también luz) sobre las similitudes y/o diferencias entre la habilidad de categorizar en los humanos y en diversas especies no humanas. Es claro que numerosas preguntas no han obtenido una respuesta satisfactoria y que mucho más trabajo sistemático es necesario para llegar a comprender, por un lado, la habilidad categorizadora en los humanos y la de las especies no humanas por otro lado, antes de llegar a una comprensión mayor de los puntos de unión o desunión entre los dos tipos de población. Sin embargo, las comparaciones ya establecidas constituyen un primer paso y allanan el camino en el siempre difícil terreno de la psicología comparativa.

CAPÍTULO II: SOBRE LA CATEGORIZACIÓN DE LOS SONIDOS.

2.1. La categorización de los sonidos: Evidencia experimental.

Los seres humanos parecen tener categorías para prácticamente cualquier cosa y esto incluye las categorías de sonidos. Aunque la mayor parte de los estudios de la categorización han trabajado con estímulos visuales, existen otro tipo de estímulos que se perciben a través de otros modos de percepción y, si una categoría implica una clase o agrupación coherente de entidades o eventos que ocurren en el mundo, parece obvio que hay un gran número de sonidos que los organismos encuentran y categorizan que podrían considerarse como categorías ejemplificadas por una variedad de eventos acústicamente diferentes. De hecho, según Nathan (1996: 112), “los sonidos.... son categorizados de la misma manera que el resto de cosas en el mundo lo son”.

El propósito del resto de esta sección es repasar toda la literatura que ha tratado sobre la categorización de sonidos producidos naturalmente o artificialmente (es decir, sintetizados) a imitación de los sonidos naturales del inglés. Además, se hará especial hincapié en las técnicas de categorización usadas en los diversos estudios experimentales.

2.1.1. La categorización de los sonidos por los Adultos.

El estudio de la categorización o clasificación que los adultos hacen de los sonidos tiene una larga historia en los campos de la percepción del habla y la fonología experimental. Diversas técnicas experimentales han sido usadas para estudiar la categorización de los sonidos por adultos. Entre ellas destacan la técnica del “phoneme monitoring”, las tareas de identificación absoluta, los test de discriminación diferencial y el paradigma de formación de conceptos.

La técnica de la “monitorización de fonemas” consiste en pedirle a los sujetos que presionen un botón tan rápido como puedan tras oír un estímulo que contenga un determinado tipo de sonido previamente especificado. El paradigma, desarrollado por Foss (Foss 1969; Foss & Lynch 1969), ha sido empleado para estudiar diversos fenómenos como la unidad de percepción o identificación (ej. Foss & Swinney 1973; Healy & Cutting 1976; Mills 1980; Savin & Bever 1980; Segui *et al.* 1981; Swinney & Prather 1980), los efectos del acento sobre el reconocimiento de fonemas (Cutler 1970), o, sobre todo, las influencias de factores fonéticos preléxicos o no fonéticos y postléxicos. La gran cantidad de estudios realizados y la diversidad

de resultados obtenidos parecen apoyar la hipótesis de que solo los modelos que prevén la existencia de efectos léxicos y preléxicos pueden explicar satisfactoriamente el gran número de contradictorios resultados (ej. Cutler *et al.* 1987; Dell & Newman 1980; Eimas *et al.* 1990; Dess & Blank 1980; Dess & Lynch 1969; Dess *et al.* 1980; Frauenfelder & Segui 1989; Segui & Frauenfelder 1986; Segui *et al.* 1981; Stemberger *et al.* 1985).

Las técnicas que más han sido utilizadas para estudiar la categorización de los sonidos son la de la “identificación absoluta” y “discriminación diferencial” (Weitzman 1993: 141). La primera consiste en hacer que los sujetos escuchen una serie de sonidos y decidan a cual categoría de un número previamente definido cada uno de los estímulos de la serie pertenece (véanse, por ejemplo, los estudios de Barclay 1972; Davis & Kuhl 1992; Ganong 1980; Healy & Cutting 1976; Lotto 2000; Miller & Micely 1955; Miller & Volaitis 1989, 1992; Oden & Massaro 1978; Samuel 1982). En cuanto a la técnica de la discriminación diferencial, cabe resaltar que existen dos versiones. La discriminación “igual-diferente” es aquella en la que los sujetos deben indicar si un par de sonidos presentados uno tras otro son, para ellos, el mismo o diferentes. Una segunda versión, la llamada tarea “ABX”, consiste en discriminar de un trío de sonidos aquel que es diferente de los dos restantes (ver, por ejemplo, el estudio clásico de Liberman *et al.* 1967). Estas dos técnicas, (identificación y discriminación) han dominado los estudios de la percepción del habla desde sus comienzos y normalmente son empleadas ambas en dichos estudios (ej. Barclay 1972; Carney & Widin 1976; Carney *et al.* 1977; Frazier 1976; Liberman *et al.* 1957; Mandler 1976; Miller & Eimas 1989; Pisoni & Lazarus 1974; Pisoni & Tash 1974; Samuel 1977). Entre los temas más destacados están los de la percepción categorial/percepción continua (véase Repp 1984 para un resumen) y el del emplazamiento de los límites entre categorías (véase Repp & Liberman 1987).

Finalmente, otra técnica útil en el estudio de la categorización de los sonidos es la de la “formación de conceptos”, técnica que consiste de dos partes: una fase de aprendizaje y otra de generalización o test. En la fase de aprendizaje, los sujetos aprenden a discriminar entre estímulos que ejemplifican una categoría que ha de aprenderse y estímulos que no la ejemplifican. La tarea de los sujetos es pues adivinar si cada uno de los estímulos que se encuentran es un miembro de aquella categoría o no e indicar si piensan que es un miembro o no inmediatamente después de oír el estímulo. Los sujetos saben si han acertado o por el contrario se han equivocado porque se les proporciona feedback en relación a cada estímulo presentado acerca de si era o no un miembro de la categoría. En la sesión de test, los sujetos intentan generalizar sus respuestas a nuevos estímulos pero no se les proporciona feedback. La técnica de formación de conceptos ha sido utilizada en su mayor parte en estudios de fonología experimental (ej. Jaeger 1980a, 1980b, 1984, 1986b; Jaeger & Ohala 1984; Ohala 1983, 1986; Wang 1985) o incluso en el campo de la percepción del habla (ej. Weitzman 1990a, 1990b, 1992, 1993).

2.1.2. La categorización de los sonidos por los Niños, Los Bebés, y las Especies No Humanas.

Como ocurren con los estudios de categorización que utilizan estímulos visuales, los estudios de la categorización de los sonidos también se han extendido a lo largo de los años a poblaciones no adultas de humanos e incluso a especies no animales. En muchos casos, las mismas técnicas utilizadas con los adultos pero adaptadas para facilitar a los sujetos humanos menores o incluso no humanos han sido empleadas. En ocasiones, nuevas técnicas se han tenido que utilizar debido a las limitaciones motoras de los propios sujetos. Estudios clásicos con niños son, por ejemplo, los de Rebecca Treiman (Bruck & Treiman 1990; Treiman & Baron 1981; Treiman & Breaux 1982; Treiman 1985a, 1985b, 1988, 1992; Treiman *et al.* 1994), que ejemplifican prácticamente todas las técnicas experimentales utilizadas y temas tratados y que son un buen comienzo para familiarizarse con la literatura sobre la categorización de los sonidos por los niños.

La categorización de los sonidos por bebés ha experimentado en los últimos años un gran auge. Su estudio presenta, no obstante, problemas similares a los de la categorización de estímulos visuales debido a la corta edad de los sujetos pero diferentes versiones de un genérico paradigma de aprendizaje discriminativo han logrado profundizar en el entendimiento de los primeros estadios de la percepción de los sonidos en los humanos. Estudios clásicos son los de Patricia Kuhl (ej. Grieser & Kuhl 1983, 1989; Kuhl 1977, 1979, 1980, 1987, 1991; Kuhl & Miller 1975) que, como en el caso de Treiman, resumen muy bien lo que ha sido esta línea de investigación.

Finalmente deberían destacarse algunos estudios realizados sobre la categorización de los sonidos por grupos de especies no humanas, como los insectos (ej. Thorson *et al.* 1982; Weber *et al.* 1981), las ranas y sapos (ej. Capranica 1966; Gerhardt 1978), las aves (ej. Beletsky *et al.* 1980; Gottlieb 1980; Kluender *et al.* 1987; Kluender *et al.* 1998), los primates (ej. Kuhl 1991; Mose & Snowdon 1975; Shyan *et al.* 1987; Wright *et al.* 1990), las chinchillas (ej. Burdick & Miller 1975; Kuhl 1981; Kuhl & Miller 1975a, 1978), los ratones (ej. Ehret & Haack 1982), etc. Al igual que los estudios con estímulos visuales, los experimentos sobre las habilidades auditivas categorizadoras de las especies no humanas están arrojando luz no sólo sobre sus propias habilidades sino también sobre las humanas desde una perspectiva comparada.

2.2. La Categorización de los Sonidos del Inglés y su Relación con la Fonología.

La categorización de los sonidos y la fonología parecen tener mucho que ver la una con la otra. En primer lugar, y debido a la importancia de la categorización para la percepción,

cognición, memoria, conducta, y lenguaje (véanse las secciones de la 1.1.1. a la 1.1.5.), el estudio experimental de las habilidades categorizadoras del hablante permitiría a fonólogos desarrollar una teoría explicativa del lenguaje y llegar así a un mejor entendimiento de las fuentes de sus regularidades estructurales. Como la estructura de los lenguajes naturales parece estar determinada en gran medida por las características tanto del aparato vocal humano como de su aparato auditivo/perceptual y de factores cognitivos generales (Nathan 1986, 1994; Pierrehumbert *et al.* 2000), la categorización, que es una habilidad cognitiva básica, podría ser la causante de muchos fenómenos estructurales de los que los fonólogos se han preocupado con frecuencia. En segundo lugar, podría decirse que los datos obtenidos del estudio de la categorización de los sonidos por los hablantes podría arrojar luz sobre cuestiones tradicionalmente controvertidas en la teoría fonológica. Sin embargo, considerar la categorización de los sonidos por hablantes reales como evidencia valiosa para solventar controversias teóricas en fonología es una importante asunción que no es compartida por la mayor parte de los fonólogos y que es explícitamente rechazada por algunos (ej. Bloch & Trager 1942: 40). El uso de evidencia experimental obtenida de los hablantes de una lengua no es muy común en muchas teorías fonológicas. Tal evidencia se descarta o considera irrelevante para el objeto de estudio, es decir, el patrón de los sonidos de la lengua. Además, recoger tal tipo de evidencia es aún menos usual. Finalmente, recurrir a hablantes reales para decidir qué análisis de entre los diversos posibles existentes o propuestos es más apropiado parece implicar que las descripciones fonológicas o análisis que son válidos teóricamente son aquellos que tienen algún tipo de apoyo psicolingüístico experimental. Sin embargo, la asunción de que los objetivos de la fonología deberían ser proporcionar descripciones psicológicamente válidas no ha sido la visión dominante en la historia de la fonología ni la más exitosa. A continuación se exponen las diferencias entre las dos grandes visiones acerca de lo que constituye la labor de la fonología y posteriormente las razones por las que la visión “psicológica” de la fonología no ha tenido gran éxito. Asimismo se exponen las críticas al uso de experimento las ventajas del uso de la experimentación y finalmente se argumenta que las dos visiones son complementarias.

2.2.1. Fonología Descriptiva vs. Fonología Psicolingüística o Psicológica.

La cuestión de la relación entre los patrones descritos por los fonólogos y su relevancia psicológica ha sido un tema controvertido a lo largo de los años. Existen aparentemente dos visiones fundamentales: en primer lugar, tal relación se ve como irrelevante. Los fonólogos que defienden esta postura argumentan que la labor de la fonología es simplemente la de descubrir, capturar y sistematizar todas las estructuras fonológicas detectables en un determinado corpus de datos lingüísticos. En segundo lugar, otros fonólogos piensan que los análisis fonológicos no deberían ser meras descripciones de los datos lingüísticos, sino que deberían representar el

conocimiento (tácito o explícito) que los hablantes tienen acerca del lenguaje y ser por tanto capaces de ser válidos tanto como descripciones de los patrones lingüísticos en los que se basan y, sobre todo, como reflejos del conocimiento que los hablantes tienen de su lengua. Esta segunda visión sobre la labor de la fonología ha sido seguida por diversos autores, aunque en general, ha sido interpretada de una forma errónea por diversas razones (véanse los estudios Cutler 1979; Derwing 1979; Campbell 1986; Eddington 1996, 1999; Fromkin 1980; McCawley 1986; Wheeler 1980). Por poner un ejemplo, sirva decir que los fonólogos que se jactan de describir el conocimiento que los hablantes tienen sobre los patrones fonológicos, generalizaciones, estructuras, etc., de su lengua, normalmente llevan a cabo sus estudios sin consultar a hablantes reales sino que, por el contrario, basan sus afirmaciones en especulaciones basadas en datos no experimentales.

2.2.2. Críticas al Uso de los Experimentos en Fonología.

A pesar de que el uso de experimentos cuidadosamente diseñados y realizados parece ser un método apropiado para validar o invalidar las afirmaciones acerca de la pertinencia de los análisis fonológicos, es paradójico que la mayor parte de los fonólogos con orientación psicológica han sido negativos o indiferentes hacia los experimentos. Entre las razones de esta indiferencia pueden hallarse una insuficiente familiaridad con la naturaleza de los experimentos y de cómo procede la investigación experimental, con la psicología cognitiva, la falsa creencia de que la experimentación implica instrumental complejo, procedimientos y análisis estadísticos, etc. Asimismo, diversas críticas se han vertido sobre el uso y naturaleza de los experimentos a lo largo de los años. Estas críticas carecen, en la mayor parte de los casos, de una sólida argumentación que sobreviva a una réplica formal y argumentada a tales críticas. Por ejemplo, se critica de los experimentos que suelen producir resultados contradictorios pero normalmente se ignora que los resultados contradictorios ocurren en absolutamente todos los campos de la ciencia que incorporan experimentos y que aquellos son resultado de interpretaciones erróneas, práctica metodológicamente pobre, o de la presencia de variables no anticipadas. La respuesta a unos resultados contradictorios es realizar un nuevo experimento que intente neutralizar el fenómeno problemático que puede haber producido tales resultados contradictorios.

Se critica también de los experimentos el que la experimentación procede muy lentamente y que de esta forma se retarda en progreso en una disciplina como la fonología. Tras esta crítica subyace la idea de que si la teoría fonológica tuviese que esperar a una validación experimental de cada noción que propone, se lograría un progreso muy lento (Ohala 1986: 12). Sin embargo, es dudoso que si algún “progreso” se realice realmente en las teorías fonológicas que nunca someten sus hipótesis a una validación empírica. De hecho, sólo debemos mirar los

particular, un buen análisis (y evaluación de tal análisis) solo puede llevarse a cabo si hace explícito su dominio (descriptivo o psicológico) y los métodos empleados.

Sin embargo, los dos enfoques sobre la tarea de la fonología tratados no son mutuamente exclusivos o irreconciliables. De hecho, se trata de caminos complementarios que convergen ya que parece a todas luces posible proporcionar una explicación unificada de la estructura del lenguaje además del conocimiento declarativo o procedimental que los hablantes tienen de su lengua o de las que han aprendido con posterioridad. En otras palabras, el énfasis en distinguir entre una fonología puramente descriptiva y una fonología psicológica no significa que no deba existir una interacción entre ambas.

La presente tesis, sin embargo, ha de entenderse desde la perspectiva de la fonología psicológica y pretende encuadrarse dentro de los estudios que pretenden investigar la relevancia psicológica de las descripciones fonológicas, describir patrones fonológicos importantes conceptualmente para los hablantes y evaluar los análisis fonológicos desde su validación experimental.

2.3. Formación de Conceptos como Paradigma Experimental.

Para investigar la forma en la que los hablantes categorizan los sonidos del lenguaje y obtener así evidencia experimental válida para proponer análisis fonológicos que reflejen lo que es psicológicamente válido para los hablantes se deberían seleccionar técnicas útiles. Durante muchos años, diferentes técnicas se han empleado (véase la sección 2.1.). El paradigma de formación de conceptos parece especialmente apropiado para el fin citado. La técnica es un paradigma experimental muy bien conocido y utilizado originariamente (y de forma amplia) en psicología (ej. Bolton 1977; Bourne & Bunderson 1963; Bruner *et al* 1956; Dominowski 1970; Hull 1920; Smoke 1932). La técnica ha sido también empleada en la lingüística y en la fonología (ej. Baker *et al.* 1973; Jaeger 1980a: exp.4).

2.3.1. Partes de un Experimento de Formación de Conceptos.

Aunque la Formación de Conceptos (de ahora en adelante FC) se considera típicamente como un paradigma consistente en dos partes, sesión de aprendizaje y sesión de test, que sigue a aquella, los experimentos de FC normalmente se dividen en cuatro fases: instrucciones, sesión de aprendizaje, sesión de test, y entrevista post-experimental (Jaeger 1980a: 66-76; 1986: 214-221).

2.3.1.1. Las instrucciones.

Las instrucciones necesarias para llevar a cabo un experimento FC pueden presentarse oralmente o de forma escrita e informan al sujeto acerca de las características del experimento. En primer lugar, demandan de los sujetos la concentración de su atención en descubrir la presencia o ausencia de algún fenómeno especificado en los estímulos a los que se verán expuestos durante la tarea. En otras palabras, las instrucciones informan a los sujetos que el fenómeno específico aparece en algunos estímulos y no en otros. Los estímulos que contienen tal fenómeno pertenecen a una clase o categoría mientras que los que no lo contienen no son ejemplos o miembros de aquella categoría. En segundo lugar, las instrucciones piden a los sujetos que respondan a cada estímulo de la manera que ellas mismas especifican una vez que el sujeto tenga cierta idea acerca de qué estímulos contienen aquel fenómeno y cuáles no. En tercer lugar, las instrucciones informan a los sujetos que se les facilitará feedback en la fase de aprendizaje para que conozcan si los estímulos a los que serán expuestos ejemplifican el fenómeno que tienen que descubrir o no. Finalmente, los sujetos son informados de que después de una sesión inicial de entrenamiento o aprendizaje, vendrá una segunda sesión en la que continuarán encontrándose con estímulos que ejemplifican el fenómeno y otros que no. Sin embargo, los sujetos no dispondrán de feedback en esta sesión, en la que deben comportarse como en la primera.

2.3.1.2. La Sesión de Aprendizaje.

El objetivo de la sesión de aprendizaje es enseñar a los sujetos el fenómeno que se está investigando. Esto se hace entrenándolos a clasificar un conjunto de estímulos en diferentes grupos o categorías que han sido predefinidos por el experimentador. En la sesión de aprendizaje, los sujetos son por tanto entrenados en responder de un modo particular a un determinado grupo de estímulos que contiene el concepto investigado y que por tanto lo ejemplifican (estímulos positivos) y en responder de otra forma a los estímulos que no incluyen el fenómeno y, por tanto, no ejemplifican dicho concepto.

En la sesión de aprendizaje de un experimento FC existen tres eventos críticos, la presentación del estímulo, la respuesta, y el feedback informativo. Estos tres eventos, que ocurren en el citado orden, constituyen cada uno de los intentos de aprender el concepto. Asimismo, en la sesión de aprendizaje, dos fuentes de información están disponibles para el sujeto, los propios estímulos y el feedback informativo proporcionado por el experimentador. Finalmente, después de examinar cada estímulo y antes de que se le proporcione el feedback informativo, el sujeto debe realizar una respuesta como reacción al estímulo para indicar su decisión relativa a la pertenencia o no de dicho estímulo en la categoría que está formando.

2.3.1.3. La Sesión de Test.

La sesión de test de un experimento FC es básicamente idéntica a la de la sesión de aprendizaje excepto una diferencia muy importante: la ausencia de feedback informativo. Esta ausencia de feedback se justifica ya que la intención de la fase de test es comprobar si los sujetos han aprendido realmente la categoría enseñada en la fase de entrenamiento y, en muchos casos, observar los patrones de clasificación de los sujetos relativos a los llamados estímulos “test”, que no son más que nuevos casos que podrían ser considerados como potenciales miembros de la categoría pero cuya pertenencia a la misma no está del todo clara. Los patrones de clasificación de los sujetos relativos a los estímulos test permiten al experimentador trazar los límites de la categoría que se está investigando.

2.3.1.4. La Entrevista Post-Experimental.

La parte final de una sesión de FC es una entrevista post-experimental considerada como una fuente de información verbal que complementa a la evidencia motora obtenida en las sesiones de aprendizaje y test.

2.3.2. Interpretación de los Datos.

Una vez que un experimento FC ha sido llevado a cabo, los experimentadores tienen a su disposición diferentes fuentes de información. Éstas incluyen el número de sujetos que formaron la categoría correctamente, el número total de respuestas en la sesión de aprendizaje, el tiempo requerido para lograr la solución de la tarea, la proporción de respuestas correctas en comparación con las incorrectas, los tiempos de reacción en responder a los estímulos que son presentados, etc. Además, si una entrevista post-experimental es llevada a cabo, se dispone del conocimiento verbalizado de los sujetos relativo a la categoría o fenómeno investigado. Un análisis post-experimental de todos los datos mencionados y muchos otros permite arrojar luz sobre las cuestiones investigadas.

CAPÍTULO III. EVIDENCIA EXPERIMENTAL PROCEDENTE DE LA CATEGORIZACIÓN DE LOS SONIDOS: UNA INVESTIGACIÓN FONOLÓGICA PSICOLINGÜÍSTICA.

3.1. Introducción.

En el capítulo I se revisó la literatura acerca de la categorización con estímulos visuales mayormente en diversa poblaciones mientras que el capítulo II hizo lo propio con los estímulos auditivos. Se argumentó entonces que la categorización de los sonidos por sujetos experimentales proporcionaría evidencia útil para resolver controversias en la teoría fonológica de compromiso explícitamente psicológico. Una técnica experimental se seleccionó entonces para proporcionar evidencia sobre algunos fenómenos específicos. El propósito del capítulo III es el de describir la realización de diversos experimentos realizados para obtener evidencia sobre ciertos fenómenos fonológicos y/o fonéticos controvertidos y proporcionar así evidencia experimental que pueda arrojar algo de luz sobre dichos fenómenos y poder así proponer análisis fonológicos psicológicamente plausibles.

3.1.1. Aspectos Generales Relacionados con los Experimentos Descritos.

El propósito de la siguiente sección es describir aspectos relacionados con los experimentos que se describen en esta tesis de tal manera que la presentación de aquellos y su entendimiento y claridad se verán beneficiados de antemano.

3.1.1.1. Propósitos Generales y Específicos de los Experimentos y Selección de los Fenómenos Fonológicos Investigados.

El propósito general y explícito de los experimentos que se describen en esta tesis es el de obtener evidencia experimental para proporcionar una explicación a ciertos temas fonológicos que corresponda muy de cerca al conocimiento que los hablantes tienen acerca de los sonidos del lenguaje. En otras palabras, el objetivo de los experimentos es proponer descripciones fonológicas que representen el conocimiento que los hablantes tienen de los fenómenos investigados tal y como se revela de su actuación categorizadora.

Los fenómenos investigados en la presente tesis han sido escogidos cuidadosamente. Estos fenómenos son los relacionados con la categorización de las llamadas semivocales o

semiconsonantes, la clasificación de las africadas inglesas, la clasificación de las oclusivas bilabiales tras [s] tautosilábica, y la discriminación de la unidad lingüística llamada “alófono”, ejemplificada en esta tesis por la alófono bilabial sordo oclusivo y aspirado del fonema / p /.

3.1.1.2. Selección de los Sujetos.

Lo sujetos seleccionados para participar en los experimentos descritos en sucesivas secciones fueron 80 hablantes nativos del inglés de entre 18 y 45 años. La homogeneidad de este grupo de sujetos fue perseguida imperiosamente. Cabe destacar que todos ellos eran estudiantes universitarios o lo habían sido en el pasado y que prácticamente ninguno había estudiado fonética y/o fonología en el pasado y, si lo había hecho (normalmente no más de 4/6 meses), declararon no recordar prácticamente nada. Se pagó a los sujetos 6 euros por su participación en los experimentos.

3.1.1.3. Asignación de las Tareas Experimentales a los Sujetos.

Se decidió que cada uno de los 4 experimentos fuese realizado por un grupo de 20 sujetos que no habrían de intervenir en ningún otro experimento para evitar el cansancio producido por la acumulación de tareas o las diferencias en el grado de satisfacción de cada prueba debidas a la falta de familiaridad con la técnica experimental en la primera tarea en relación con las sucesivas.

3.1.1.4. Estímulos, Materiales, y Aparatos Experimentales.

Los estímulos utilizados fueron 400 monosílabos cuidadosamente escogidos que representaban diferentes palabras inglesas. Estas 400 palabras fueron divididas en cuatro listas de 100 palabras cada una.

Los estímulos fueron producidos por una hablante nativa del inglés de acento estándar que desconocía la finalidad de la grabación. Los estímulos fueron grabados digitalmente y tratados posteriormente una sesión de laboratorio para su posterior uso como estímulos en cada uno de los experimentos. A tal fin, se hizo uso de un ordenador en el que se hallaba instalado un programa específicamente diseñado para monitorizar automáticamente los eventos experimentales de cada estudio. El ordenador se encargaba pues de presentar los estímulos, que eran escuchados por los sujetos a través de auriculares, y de proporcionar el feedback visual en forma de rectángulos rojos o verdes (dependiendo de si un determinado estímulo era negativo o positivo) así como de grabar las respuestas motoras de los sujetos, que estos realizaban presionando un botón verde o rojo en el teclado del ordenador para indicar, respectivamente,

que un determinado estímulo era un ejemplo de la categoría que estaban aprendiendo o no lo era.

3.1.1.5. Aspectos Procedimentales Específicos de los Experimentos CF Descritos.

Los experimentos FC descritos en esta tesis tenían además las siguientes particularidades: 1) fueron realizados individualmente y de forma aislada del experimentador, que permanecía detrás de un panel durante la realización del experimento 2) la fase de aprendizaje constaba de 32 estímulos positivos y 28 negativos mientras que en la de test de 19 positivos, 12 negativos y 9 estímulos *test*; 3) el intervalo entre la presentación del estímulo y la presentación del feedback era de 5 segundos; 4) las instrucciones fueron extremadamente similares entre los cuatro experimentos para permitir una posterior comparación entre ellos.

3.2. Experimento 1. La categoría “Sonidos Consonánticos” en Inglés.

3.2.1. Diseño.

Los términos “vocal” y “consonante” han sido usados durante siglos (Gimson 1980: 27) y proceden en última instancia de los gramáticos griegos (Abercrombie 1967: 39; Trubetzkoy 1969: 223). En fonética y fonología estos términos han sido usados durante más de 120 años aunque una gran controversia ha rodeado su definición, sus características, y sus respectivos inventarios.

Aunque se han usado diversos criterios para definir las vocales y las consonantes tanto en fonología como en fonética, es interesante observar que existen un número de segmentos que resisten a una nítida clasificación como vocales o consonantes en términos fonéticos o fonológicos. Entre estos segmentos destacan, en inglés, la famosa *h* aspirada o las tradicionalmente llamadas semivocales [j] y [w].

La ya larga controversia relativa a la definición de las vocales y consonantes, a la línea divisoria entre ambas categorías y al estatus categorial de segmentos como los arriba mencionados se presenta como un atractivo caso para el experimentador interesado en la forma en la que los hablantes conciben las categorías “vocales” y “consonantes” y en reflejar tal concepción en los análisis fonológicos.

El propósito del siguiente estudio es, por tanto, proporcionar evidencia acerca de la habilidad de los sujetos de clasificar diferentes miembros de una categoría como la de “sonidos consonánticos” y valorar así la afirmación que hacen van Ooijen y otros (1992: 102) de que los hablantes normalmente tienen una conciencia de la distinción entre vocales y consonantes. El propósito del estudio es también cómo clasifican los hablantes los [h], [w], y [j]. Las preguntas de investigación examinadas son:

- 1) ¿Pueden los hablantes del inglés formar la categoría “sonidos consonánticos”?
- 2) Si es así, ¿cómo clasifican [h], [j], y [w] en posición inicial de palabra y seguidas por vocal?

Las dos hipótesis que formulamos son que los hablantes podrán formar la categoría y que clasificarán [h], [w], y [j] como miembros de la categoría “sonidos consonánticos”.

3.2.2. Método.

3.2.2.1. Sujetos.

Veinte sujetos de edades comprendidas entre los 18 y 37 años fueron seleccionados al azar del grupo inicial de 80 sujetos reclutados para los experimentos CF descritos en esta tesis. Los sujetos tenían las características mencionadas en la sección 3.1.1.2.

3.2.2.2. Estímulos.

Los estímulos del presente experimento consistían de 100 palabras monosilábicas. En la sesión de aprendizaje habían 32 estímulos positivos y 28 negativos. Asimismo, la sesión de test presentaba 19 estímulos positivos y 12 negativos junto con 9 estímulos test (3 ejemplos de [w], 3 de [j], y 3 de [h]). La tabla 1 ofrece un resumen de los tipos de estímulos usados y la tabla 2 la lista completa de tales estímulos.

Tabla 1. Resumen de los Estímulos Usados en el Experimento 1.				
Tipo de estímulo	SESIÓN DE APRENDIZAJE		SESIÓN DE TEST	
	Subtipo de estímulo	No de ejemplos	Subtipo de estímulo	No de ejemplos
EJEMPLOS POSITIVOS	Oclusivas / p b t d k g /	12	Oclusivas / p b t d k g /	6
	Fricativas / s z f v θ ð /	12	Fricativas / s z f v θ ð ʃ /	9
	Africadas / tʃ dʒ /	4	Africadas / tʃ dʒ /	2
	Nasales / m n /	4	Nasales / m n /	2
	TOTAL	(32)	TOTAL	(19)
EJEMPLOS NEGATIVOS	Monoptongos / ʌ æ ɑ: u: i: ɒ ɪ e ɜ: /	17	Monoptongos / ɪ ʌ ɑ: ɒ ɜ: ɔ: /	8
	Diptongos y Triptongos / ɔɪ aɪ əʊ eɪ aʊə /	11	Diptongos / eɪ əʊ /	4
	TOTAL	(28)	TOTAL	(12)
EJEMPLOS TEST	----	----	[h]	3
			[j]	3
			[w]	3
			TOTAL	(9)

Tabla 2. Lista de Estímulos para el Experimento 1.					
Orden	Estímulo	Positivo (+) Negativo (-)	Orden	Estímulo	Positivo (+) Negativo (-)
SESIÓN DE APRENDIZAJE					
1.	path	+	53.	change	+
2.	ash	-	54.	aim	-
3.	boom	+	55.	jug	+
4.	ache	-	56.	ebb	-
5.	toad	+	57.	ace	-
6.	duck	+	58.	mug	+
7.	kid	+	59.	itch	-
8.	up	-	60.	name	+
9.	give	+	SESIÓN DE TEST		
10.	eat	-	1.	pet	+
11.	seethe	+	2.	bathe	+
12.	zone	+	3.	tooth	+
13.	edge	-	4.	ode	- control
14.	fish	+	5.	deep	+
15.	van	+	6.	hot	test
16.	at	-	7.	eight	-
17.	egg	-	8.	ape	-
18.	ill	-	9.	cab	+
19.	thing	+	10.	guess	+
20.	off	-	11.	oath	- control
21.	that	+	12.	fetch	+
22.	each	-	13.	heat	test
23.	hours	-int	14.	us	-
24.	cheese	+	15.	vague	+
25.	job	+	16.	all	- control
26.	miss	+	17.	seed	+
27.	out	-	18.	youth	test
28.	neck	+	19.	ship	+ control
29.	eve	-	20.	wit	test
30.	on	-	21.	zoom	+
31.	pub	+	22.	earn	-
32.	aid	-	23.	thin	+
33.	beach	+	24.	then	+
34.	teach	+	25.	urge	-
35.	oil	-	26.	shock	+ control
36.	odd	-	27.	orb	- control
37.	dove	+	28.	once	test
38.	arm	-	29.	chase	+
39.	call	+	30.	yet	test
40.	goose	+	31.	of	-
41.	ale	-	32.	judge	+
42.	safe	+	33.	wall	test
43.	earth	-	34.	if	-
44.	zip	+	35.	hard	test
45.	fang	+	36.	shell	+ control
46.	owl	-	37.	map	+
47.	ooze	-	38.	net	+
48.	ice	-	39.	use	test
49.	vet	+	40.	art	-
50.	thick	+			
51.	age	-			
52.	those	+			

3.2.2.3. Procedimiento.

3.2.2.3.1. Instrucciones.

Los sujetos leyeron una hoja de instrucciones donde se les informaba acerca de todos los eventos experimentales. Al final de dicha lectura se les permitió preguntar acerca de aquellos aspectos que no les hubieren quedado lo suficientemente claros. La sesión de aprendizaje comenzó una vez que el experimentador se había asegurado de que los sujetos habían entendido realmente las instrucciones.

3.2.2.3.2. Sesión de Aprendizaje, Sesión de test, y Entrevista Post-Experimental.

Las sesiones de aprendizaje y test procedieron según lo indicado en la 3.1.1.5. Los sujetos realizaron el experimento individualmente.

3.2.3. Resultados.

3.2.3.1. Sesiones de Aprendizaje y de Test.

De los 20 sujetos que participaron en el experimento se consideró que 12 aprendieron la categoría correctamente con una media de 48.67 respuestas correctas (rango 37-59, s.d.= 6.50) en la sesión de aprendizaje.

Los números de respuestas correctas (C), incorrectas (I), y nulas (NR) tanto para los estímulos positivos como negativos se muestra en la tabla 3. El mismo tipo de información en relación con los estímulos negativos se muestra en la tabla 4.

Tabla 3. Número de Respuestas Correctas, Incorrectas, y Nulas para los Estímulos Positivos y Negativos del Experimento 1 (Sesión de Aprendizaje).						
Tipo de Estímulo	Tipo de respuesta			% Correctas	Ejemplos	Número de Respuestas
	C	I	NR			
Ejemplos positivos	322	33	29	83.85%	32	384
Ejemplos negativos	286	30	20	85.12%	28	336
TOTAL	608	63	49	84.44%	60	720

Tabla 4. Número de Respuestas Correctas, Incorrectas, y Nulas para los Estímulos Positivos y Negativos del Experimento 1 (Sesión de Test).						
Tipo de Estímulo	Tipo de respuesta			% Correctas	Ejemplos	Número de Respuestas
	C	I	NR			
Ejemplos positivos	218	2	8	95.61%	19	228
Ejemplos negativos	128	15	1	88.89%	12	144
TOTAL	346	17	9	93.01%	31	372

Una comparación de los datos mostrados en ambas tablas muestra que los porcentajes de respuestas correctas a los estímulos tanto positivos como negativos incrementa substancialmente en la sesión de test en comparación con los porcentajes en la sesión de aprendizaje. En otras palabras, las respuestas correctas son significativamente más frecuentes en la sesión de test que en la de aprendizaje (84.44% vs. 93.01%; P-valor = 0.000 < 0.05 según un contraste de proporciones). Este incremento en la precisión de los sujetos muestra que su actuación fue mejor en la sesión de test que en la de aprendizaje, en la que ya lo estaban haciendo bastante bien.

En relación a los estímulos *test*, cabe decir que, como refleja la tabla 5, los sujetos respondieron que los estímulos eran ejemplos de la categoría “sonidos consonánticos” más frecuentemente que no.

Tabla 5. Respuestas a los Estímulos Test en el Experimento 1.						
Estímulo Tipo	Tipo de respuesta			% Y Respuestas	Items	Número de Respuestas
	Y	N	NR			
/ h /	30	2	4	83.33%	3	36
/ j /	24	11	1	66.67%	3	36
/ w /	24	11	1	66.67%	3	36
TOTAL	78	24	6	77.22%	9	108

Aunque el porcentaje de respuestas considerando a [j] y [w] como sonidos consonánticos (66.67%) es menor que el porcentaje para [h] (83.33%) las diferencias entre los tres tipos de estímulos test no son estadísticamente significativas según un contraste de proporciones (P-valor = 0.096 > 0.05).

Un análisis más cuidadoso de los estímulos *test* revela, como se ven en la tabla 6, que cuando [j] y [w] se deletrean con “u” u “o”, los sujetos tratan a estos sonidos como consonantes un 33.33% de las veces mientras que lo hacen un 83.33% cuyo [j] y [w] se escriben con “y” o “w”, lo cual parece indicar que las respuestas de los sujetos se ven, en relación a estos dos sonidos, influidas por la ortografía convencional de las palabras utilizadas.

Tabla 6. Respuestas a los Estímulos Test en el Experimento 1 Según la Ortografía del Segmento.							
Estímulo Tipo	Ortografía	Tipo de respuesta			% de Yes	Ejemplos	Número de Respuestas
		Y	N	NR			
/ h /	< h >	30	2	4	83.33%	3	36
/ j /	< y >	20	3	1	83.33%	2	24
	< u >	4	8	0	33.33%	1	12
/ w /	< w >	20	3	1	83.33%	2	24
	< o >	4	8	0	33.33%	1	12

2.4. Discusión.

El experimento 1 fue diseñado y realizado para obtener evidencia acerca de la formación de la categoría “sonidos consonánticos” y la clasificación por parte de los hablantes nativos del inglés de los segmentos [h], [j], y [w] en posición inicial de palabra.

Los resultados obtenidos permiten confirmar las hipótesis formuladas de que los hablantes son capaces de aprender la categoría investigada y de que los hablantes categorizan los sonidos [h], [j], y [w], en general, como miembros de la categoría “sonidos consonánticos”. La influencia de la representación ortográfica de [j] y [w] resultó ser un factor importante para los sujetos a la hora de clasificar dichos segmentos aunque, en general, los resultados obtenidos parecen apuntar al hecho de que el criterio seguido para el resto de los estímulos fue mayoritariamente fonético.

3.3. Experimento 2. La categoría “Oclusivas Orales” en Inglés.

3.3.1. Diseño.

En inglés existen 6 fonemas consonánticos diferentes que normalmente se agrupan en tres subgrupos / p /-/ b /, / t /-/ d /, y / k /-/ g /. El punto de articulación es bilabial para el primer par, alveolar para el segundo y velar para el tercero. El primer miembro de cada par es tradicionalmente descrito como sordo o *fortis* mientras que el segundo es descrito como sonoro o *lenis*.

Aunque / p /-/ b /, / t /-/ d /, y / k /-/ g / se realizan normalmente como consonantes oclusivas con una rápida separación de los articuladores que lleva a una explosión corta y abrupta, estos fonemas se realizan algunas veces de forma africada. Esto significa que la separación de los articuladores no se hace tan rápidamente y, entonces, un sonido fricativo, articulado en la misma área de articulación que la oclusiva, se oye. Esto ocurre, por ejemplo, en variedades del inglés como el habla popular de Londres, donde se escuchan frecuentemente oclusivas africadas de la forma descrita como [t^s] (ej. “time” [t^saim]) o [d^z] (ej. “door” [d^zɔ:]).

Las oclusivas africadas se diferencian de la realización de combinaciones de oclusiva+ africada mayormente en la brevedad de la fricción asociada con aquellas. Secuencias tales como [t] + [s] y [d] + [z] se conocen, normalmente, con el simple nombre de africadas. En inglés, existen varios tipos de africadas como [tθ] (ej. “eighth”), [dð] (ej. “width”), [tʃ] (ej. “tray”), [dʒ] (ej. “dread”), [tʃ] (ej. “chain”), [dʒ] (ej. “jet”), y, por supuesto, [ts], en “cats”, y [dz], como en “seeds”.

Sin embargo, de todas las posibles africadas, sólo el par palatoalveolar (/ tʃ / y / dʒ /) parece tener un cierto estatus especial en la fonología del inglés (véase Bloch & Trager 1942: 49; Bloomfield 1979: 20; Gimson 1980: 171-173; Gleason 1955: 22-23; Ladefoged 1993: 60; Roach 1983: 94-97; Trager & Bloch 1941: 229; Trager & Smith 1951; Wells 1982a: 48-49). Este estatus especial consiste en que esas dos africadas no se consideran como combinaciones de dos consonantes sino que tratan como unidades indivisibles por razones distribucionales, combinatorias e incluso psicolingüísticas (véase Gimson 1980: 171-172; Ladefoged 1993: 60; Lass 1984: 26-27; Roach 1983: 37-38).

Aunque la controversia sobre el análisis bisegmental o unisegmental de [tʃ] y [dʒ] ha sido un problema tradicional en la fonología, existe un segundo problema de categorización que ha sido tratado en menor medida. Aunque / p t k / y / b d g / son claros ejemplos de la categoría “oclusivas orales”, / tʃ / y / dʒ / podrían, en principio, considerarse como “oclusivas” ya que en su producción existe una etapa de cierre al paso del aire y de compresión del aire atrapado en la boca tal y como ocurre con / p t k / y / b d g /. / tʃ / y / dʒ / serían ejemplos poco típicos de la categoría oclusivas quizás. / tʃ / y / dʒ / podrían, por otro lado, ser consideradas como un tipo especial de fricativas, ya que su fase final es parecida a la producción de las fricativas. Sin embargo, la posición más común es que / tʃ / y / dʒ / son africadas fonéticas que constituyen los únicos miembros de una tercera categoría fonológica llamada “africadas” distinta de las categorías fonológicas “oclusivas” y “fricativas”.

Teniendo en cuenta lo dicho, parecería interesante investigar cómo perciben los hablantes las africadas dejando a parte consideraciones estrictamente fonéticas o fonológicas. El propósito del experimento que se describe a continuación fue el de obtener evidencia acerca de la habilidad de los sujetos para formar la categoría “oclusivas orales”, y arrojar luz sobre la forma en la que los sujetos clasifican / tʃ / y / dʒ /. Las preguntas de investigación que este estudio hizo son las siguientes:

- 1) ¿Pueden los hablantes agrupar diferentes oclusivas orales juntas?, es decir, ¿pueden reconocer las oclusivas como un grupo?
- 2) Consideran los hablantes / tʃ / y / dʒ / como miembros de la categoría “oclusiva oral”?

Las hipótesis formuladas son que los sujetos formarían la categoría “oclusivas orales” y que no considerarían a las africadas como miembros de la categoría investigada. Esta segunda hipótesis se basa en el hecho de que la parte fricativa de las africadas puede ser lo suficientemente relevante como para hacer que los sujetos excluyan mayormente las africadas de la categoría “oclusivas orales”.

3.3.2. Método.

3.3.2.1. Sujetos.

Veinte sujetos de edades comprendidas entre los 20 y 39 años fueron seleccionados al azar del grupo inicial de 80 sujetos reclutados para los experimentos CF descritos en esta tesis. Los sujetos tenían las características mencionadas en 3.1.1.2

3.3.2.2. Estímulos.

Los estímulos del presente experimento consistían, al igual que en el experimento 1, de 100 palabras monosilábicas. En la sesión de aprendizaje habían 32 estímulos positivos y 28 negativos. Asimismo, la sesión de test presentaba 19 estímulos positivos, 12 negativos, y 9 estímulos test (5 ejemplos de [tʃ] y 5 de [dʒ]). La tabla 7 ofrece un resumen de los tipos de estímulos usados y la tabla 8 la lista completa de tales estímulos.

Tabla 7. Tabla 1. Resumen de los Estímulos Usados en el Experimento 2.				
Tipo de estímulo	SESIÓN DE APRENDIZAJE		SESIÓN DE TEST	
	Subtipo de estímulo	No de ejemplos	Subtipo de estímulo	No de ejemplos
EJEMPLOS POSITIVOS	/ p b t k g /	32	/ p b t d k g /	19
	TOTAL	(32)	TOTAL	(19)
EJEMPLOS NEGATIVOS	/ s z f v ʃ θ, ð /	22	/ s z f v ʃ θ, ð /	9
	/ m n /	6	/ m n /	3
	TOTAL	(28)	TOTAL	(12)
EJEMPLOS TEST	----	----	/ tʃ /	5
			/ dʒ /	4
			TOTAL	(9)

Tabla 8. Lista de Estímulos para el Experimento 2.

Orden	Estímulo	Positivo (+) Negativo (-)	Orden	Estímulo	Positivo (+) Negativo (-)
SESIÓN DE APRENDIZAJE					
1.	push	+	53.	cave	+
2.	fall	-	54.	zone	-
3.	bus	+	55.	gull	+
4.	verse	-	56.	shell	-
5.	tall	+	57.	thighs	- int
6.	cash	+	58.	path	+
7.	gas	+	59.	thus	- int
8.	safe	-	60.	ton	+
9.	pace	+	SESIÓN DE TEST		
10.	zeal	-	1.	pave	+
11.	beef	+	2.	bath	+
12.	tough	+	3.	tool	+
13.	shove	-	4.	file	-
14.	kill	+	5.	coal	+
15.	girl	+	6.	chill	test
16.	thief	- int	7.	seethe	-
17.	these	- int	8.	veil	-
18.	fish	-	9.	goose	+
19.	pill	+	10.	pause	+
20.	verve	-	11.	zoom	-
21.	bill	+	12.	ball	+
22.	sell	-	13.	jazz	test
23.	zoos	-	14.	shoal	-
24.	toes	+	15.	dish	+ control
25.	case	+	16.	thieve	- int
26.	gaze	+	17.	tale	+
27.	knife	- int	18.	chief	test
28.	pile	+	19.	deaf	+ control
29.	psalm	- int	20.	juice	test
30.	kneel	- int	21.	call	+
31.	beige	+	22.	those	- int
32.	shave	-	23.	give	+
33.	tail	+	24.	pull	+
34.	cough	+	25.	moth	- int
35.	thaws	- int	26.	dull	+ control
36.	miss	- int	27.	nose	- int
37.	goal	+	28.	choose	test
38.	nerve	- int	29.	buzz	+
39.	puff	+	30.	cheese	test
40.	booze	+	31.	fill	-
41.	this	- int	32.	time	+
42.	tease	+	33.	jail	test
43.	mill	- int	34.	gnash	- control
44.	cool	+	35.	choice	test
45.	gash	+	36.	deal	+ control
46.	five	-	37.	cause	+
47.	mass	- int	38.	guess	+
48.	veal	-	39.	jaws	test
49.	pass	+	40.	these	- int
50.	bush	+			
51.	sauce	-			
52.	toll	+			

3.3.2.3. Procedimiento.

3.3.2.3.1. Instrucciones.

Las instrucciones fueron exactamente las mismas que fueron empleadas por los sujetos del experimento 1.

3.3.2.3.2. Sesión de Aprendizaje, Sesión de Test, y Entrevista Post-Experimental.

Las sesiones de aprendizaje y test procedieron según lo indicado en la 3.1.1.5. Los sujetos realizaron el experimento individualmente.

3.3.3. Resultados.

3.3.3.1. Sesiones de Aprendizaje y Test.

De los 20 sujetos que participaron en el experimento se consideró que 14 aprendieron la categoría correctamente con una media de 47 respuestas correctas (rango 37-59, s.d.= 7.06) en la sesión de aprendizaje.

Los números de respuestas correctas (C), incorrectas (I), y nulas (NR) tanto para los estímulos positivos como negativos se muestra en la tabla 9. El mismo tipo de información en relación con los estímulos negativos se muestra en la tabla 10.

Tabla 9. Número de Respuestas Correctas, Incorrectas, y Nulas para los Estímulos Positivos y Negativos del Experimento 2 (Sesión de Aprendizaje).						
Tipo de Estímulo	Tipo de respuesta			% Correctas	Ejemplos	Número de Respuestas
	C	I	NR			
Ejemplos positivos	364	74	10	81.25%	32	448
Ejemplos negativos	290	98	4	73.98%	28	392
TOTAL	654	172	14	77.86%	60	840

Tabla 10. Número de Respuestas Correctas, Incorrectas, y Nulas para los Estímulos Positivos y Negativos del Experimento 2 (Sesión de Test).						
Tipo de Estímulo	Tipo de respuesta			% Correctas	Ejemplos	Número de Respuestas
	C	I	NR			
Ejemplos positivos	248	13	5	93.23%	19	266
Ejemplos negativos	140	25	3	83.33%	12	168
TOTAL	388	38	8	89.40%	31	434

Una comparación de los datos mostrados en ambas tablas muestra que los porcentajes de respuestas correctas a los estímulos tanto positivos como negativos incrementa substancialmente en la sesión de test en comparación con los porcentajes en la sesión de aprendizaje. En otras palabras, las respuestas correctas son significativamente más frecuentes en la sesión de test que en la de aprendizaje (77.86% vs. 89.40%; P-valor = 0.000 < 0.05 según un contraste de proporciones).

Analizando a continuación las respuestas de los sujetos a los estímulos *test* (véase tabla 11), queda claro que los sujetos trataron a / tʃ / y / dʒ / *no* como miembros de la categoría aprendida (es decir, “oclusivas orales”) más frecuentemente que como miembros, siendo la diferencia estadísticamente significativa tanto para / tʃ / (31.43% vs. 68.57%, P-valor = 0.00 < 0.05 según un contraste de proporciones) como para / dʒ / (39.28% vs. 60.71%, P-valor = 0.020 < 0.05 según un contraste de proporciones).

Tabla 11. Respuestas a los Estímulos Test en el Experimento 2.							
Tipo de Africada	Tipo de respuesta			% de Respuestas		Items	Número de Respuestas
	Y	N	NR	Como Plosiva	Como no Plosiva		
/ tʃ /	22	47	1	31.43%	68.57%	5	70
/ dʒ /	22	32	2	39.28%	60.71%	4	56
TOTAL	44	79	3	34.92%	65.08%	9	126

3.3.4. Discusión.

El experimento 2 fue diseñado y realizado para obtener evidencia acerca de la formación de la categoría “oclusivas orales” y la clasificación por parte de los hablantes nativos del inglés de los segmentos africados [tʃ] y [dʒ] en posición inicial de palabra.

Los resultados obtenidos permiten confirmar las hipótesis formuladas de que los hablantes son capaces de aprender la categoría investigada y de que los hablantes tienden a no categorizar las africadas como miembros de la categoría “oclusivas orales”. Estos resultados parecen indicar que la parte fricativa de las africadas es lo suficientemente poderosa como para hacer que los sujetos no clasifiquen a las africadas como ejemplos de oclusivas orales a pesar de compartir con éstas importantes rasgos fonéticos.

3.4. Experimento 3. La categoría “Fonema / p /” en Inglés.

3.4.1. Diseño.

El fonema ha sido una de las nociones básicas en la teoría lingüística (Jaeger 1980a: 26). Definido de diversas maneras, recientemente ha venido a operacionalizarse como una categoría mental realizada por diversos sonidos acústicamente diferentes que los hablantes categorizan como miembros de aquella misma categoría mental (ej. Nathan 1986, 1994, 1996, 1999; Taylor 1995). Esta visión mental del fonema debe mucho al trabajo experimental de Jeri Jaeger (ej. 1980a, 1980b) y John Ohala (1983, 1986) que mostraba que los hablantes clasifican varios alófonos de un determinado fonema como miembros de una misma categoría y que, por tanto, ha de considerarse al fonema como una unidad lingüística psicológicamente real y no como un mero artefacto teórico formal.

La concepción del fonema expuesta es muy interesante debido a su potencial utilidad a la hora de resolver problemas tradicionales de la fonología como son los de la asignación de ciertos alófonos controvertidos a ciertos fonemas. Por ejemplo, uno de los problemas más importantes en la teoría fonológica ha sido tradicionalmente el relacionado con la clasificación de las oclusivas orales sordas no aspiradas tras una [s] perteneciente a la misma sílaba. Aparentemente, existen cuatro posturas enfrentadas. En primer lugar, muchos investigadores mantienen que tales oclusivas deberían interpretarse como alófonos de / p t k /, mientras que otros fonólogos y fonetistas parecen decantarse por la interpretación / b d g /. Además, algunos autores prefieren ver en aquellas oclusivas una tercera categoría concebida como un archisegmento o archifonema mientras que otro grupo de fonólogos, fundamentalmente generativistas, aboga por una descripción de tales oclusivas en términos de rasgos distintivos binarios. Razones de espacio imposibilitan hacer una relación de todos y cada uno de los criterios aportados para apoyar cada una de las posiciones mencionadas. Sin embargo, el problema parece lo suficientemente interesante como para ser sometido a una investigación experimental en la que formulamos dos preguntas de investigación. Dichas preguntas son:

- 1) ¿Pueden los hablantes categorizar diferentes alófonos del fonema / p / como miembros de la misma categoría?
- 2) Si es así, ¿cómo clasifican las oclusivas bilabiales sordas no aspiradas tras / s /?

La primera hipótesis es que los sujetos experimentales clasificarían distintos alófonos del fonema / p / como miembros de la misma categoría y la segunda hipótesis es que considerarían a las oclusivas bilabiales que ocurren después de / s / como miembros de aquella categoría.

3.4.2. Método.

3.4.2.1. Sujetos.

Veinte sujetos de edades comprendidas entre los 20 y 34 años fueron seleccionados al azar del grupo inicial de 80 sujetos reclutados para los experimentos CF descritos en esta tesis. Los sujetos tenían las características mencionadas en 3.1.1.2

3.4.2.2. Estímulos.

Los estímulos del presente experimento consistían, al igual que en los experimento 1 y 2, de 100 palabras monosilábicas. En la sesión de aprendizaje habían 32 estímulos positivos y 28 negativos. Asimismo, la sesión de test presentaba 19 estímulos positivos y 12 negativos junto con 9 estímulos test (ejemplos de la bilabial oclusiva oral después de [s]). La tabla 12 ofrece un resumen de los tipos de estímulos usados y la tabla 13 la lista completa de tales estímulos.

Tabla 12. Resumen de los Estímulos Usados en el Experimento 3.				
Tipo de estímulo	SESIÓN DE APRENDIZAJE		SESIÓN DE TEST	
	Subtipo de estímulo	No de ejemplos	Subtipo de estímulo	No de ejemplos
EJEMPLOS POSITIVOS	<i>Prenuclear</i>		<i>Prenuclear</i>	
	[p ^h V-]	8	[p ^h V-]	4
	[p ^h _i V-]	4	[p ^h _i V-]	3
	[p ^h _j V-]	4	[p ^h _j V-]	2
			[p ^h _j V-]	1
	<i>Postnuclear</i>		<i>Postnuclear</i>	
	[-Vp]	8	[-Vp]	3
	[-Vmp]	3	[-Vmp]	3
	[-Vp ^o t]	2	[-Vp ^o t]	1
	[-Vsp]	2	[-Vpst]	1
	[-Vpθ]	1	[-mps]	1
	TOTAL	(32)	TOTAL	(19)
EJEMPLOS NEGATIVOS	Otros tipos de consonantes	28	Otros tipos de consonantes	12
	TOTAL	(28)	TOTAL	(12)
EJEMPLOS TEST	----	---	[spV-]	5
			[sp _i V-]	2
			[sp _j V-]	2
			TOTAL	(9)

Tabla 13. Lista de Estímulos para el Experimento 3.					
Orden	Estímulo	Positivo (+) Negativo (-)	Orden	Estímulo	Positivo (+) Negativo (-)
SESIÓN DE APRENDIZAJE					
1.	pet	+	53.	print	+
2.	sell	-	54.	fee	-
3.	up	+	55.	pond	+
4.	egg	-	56.	end	-
5.	pay	+	57.	grant	-
6.	plea	+	58.	top	+
7.	drip	+	59.	fist	-
8.	die	-	60.	trap	+
9.	apt	+	SESIÓN DE TEST		
10.	tray	-	1.	pit	+
11.	priest	+	2.	pear	+
12.	depth	+	3.	prow	+
13.	drill	-	4.	sheet	-
14.	path	+	5.	plane	+
15.	ape	+	6.	spend	test
16.	old	-	7.	near	-
17.	drift	-	8.	slow	-
18.	golf	-	9.	clamp	+
19.	pie	+	10.	pulse	+
20.	fish	-	11.	bear	- int
21.	pray	+	12.	cap	+
22.	ash	-	13.	spa	test
23.	bay	-int	14.	ground	-
24.	place	+	15.	prayer	+
25.	opt	+	16.	false	-
26.	stamp	+	17.	drop	+
27.	sphere	-int	18.	spy	test
28.	post	+	19.	glimpse	+ control
29.	graph	-int	20.	spoon	test
30.	blast	-int	21.	prince	+
31.	shop	+	22.	phone	- int
32.	east	-	23.	paste	+
33.	pea-p	+	24.	ship	+
34.	play	+	25.	sly	-
35.	self	-	26.	lapsed	+ control
36.	psalm	-int	27.	slob	- control (int)
37.	proud	+	28.	sponge	test
38.	sea-see	-	29.	plot	+
39.	asp	+	30.	spray	test
40.	clasp	+	31.	cross	-
41.	dry	-	32.	tramp	+
42.	damp	+	33.	sply	test
43.	clean	-	34.	nymph	- int
44.	keep	+	35.	spring	test
45.	paw	+	36.	pure	+ control
46.	act	-	37.	lamp	+
47.	bet	-int	38.	rapt	+
48.	trust	-	39.	split	test
49.	plough	+	40.	stealth	-
50.	group	+			
51.	fond	-			
52.	imp	+			

3.4.2.3. Procedimiento.

3.4.2.3.1. Instrucciones.

Las instrucciones fueron muy similares a las usadas en los experimentos 1 y 2.

3.4.2.3.2. Sesión de Aprendizaje, Sesión de test, y Entrevista Post-Experimental.

Las sesiones de aprendizaje y test procedieron según lo indicado en la 3.1.1.5. Los sujetos realizaron el experimento individualmente.

3.4.3. Resultados.

3.4.3.1. Sesión de Aprendizaje y Sesión de Test.

De los 20 sujetos que participaron en el experimento se consideró que todos aprendieron la categoría correctamente con una media de 56.55 respuestas correctas (rango 51-60, s.d.= 2.06) en la sesión de aprendizaje.

Los números de respuestas correctas (C), incorrectas (I), y nulas (NR) tanto para los estímulos positivos como negativos se muestra en la tabla 14. El mismo tipo de información en relación con los estímulos negativos se muestra en la tabla 15.

Tabla 14. Número de Respuestas Correctas, Incorrectas, y Nulas para los Estímulos Positivos y Negativos del Experimento 3 (Sesión de Aprendizaje).						
Tipo de Estímulo	Tipo de respuesta			% Correctas	Ejemplos	Número de Respuestas
	C	I	NR			
Ejemplos positivos	598	12	30	93.44%	32	640
Ejemplos negativos	532	12	16	95%	28	560
TOTAL	1130	24	46	94.17%	60	1200

Tabla 15. Número de Respuestas Correctas, Incorrectas, y Nulas para los Estímulos Positivos y Negativos del Experimento 3 (Sesión de Test).						
Tipo de Estímulo	Tipo de respuesta			% Correctas	Ejemplos	Número de Respuestas
	C	I	NR			
Ejemplos positivos	379	0	1	99.74%	19	380
Ejemplos negativos	233	2	5	97.08%	12	240
TOTAL	612	2	6	98.71%	31	620

Una comparación de los datos mostrados en ambas tablas muestra que los porcentajes de respuestas correctas a los estímulos tanto positivos como negativos incrementa substancialmente en la sesión de test en comparación con los porcentajes en la sesión de aprendizaje. En otras palabras, las respuestas correctas son significativamente más frecuentes en la sesión de test que en la de aprendizaje (94.17% vs 98.71%; P-valor = 0.000 < 0.05 según un contraste de proporciones).

Las respuestas de los sujetos a los estímulos *test* muestran que, como se puede ver, en la tabla 16, aquellos respondieron afirmativamente con diferencias no significativas entre los distintos subtipos de estímulos positivos ([spV-] 99% vs. [sp̥V-] 100%, P-valor = 0.315 > 0.05; [spV-] 99% vs. [sp̥V-] 92.5%, P-valor = 1.29 > 0.05; [sp̥V-] 92.5% vs. [sp̥V-] 100%, P-valor = 0.072 > 0.05, según tres contrastes de proporciones).

Tabla 16. Respuestas a los Estímulos Test en el Experimento 3.						
Contexto Fonético	Tipo de respuesta			% Respuestas Positivas	Ítems	Número de Respuestas
	Y	N	NR			
[spV-]	99	1	0	99%	5	100
[sp̥V-]	40	0	0	100%	2	40
[sp̥V-]	37	1	2	92.5%	2	40
TOTAL	176	2	2	97.77%	4	180

3.4.4. Discusión.

El experimento 3 fue diseñado y realizado para obtener evidencia acerca de la formación de la categoría “oclusivas orales” y la clasificación por parte de los hablantes nativos del inglés de las oclusivas bilabiales orales y no aspiradas tras [s] en la misma sílaba.

Los resultados obtenidos permiten confirmar las hipótesis formuladas de que los hablantes son capaces de aprender la categoría / p / y de que estos categorizan las oclusivas bilabiales orales y no aspiradas tras [s] mayoritariamente como ejemplos de / p /. Entre las razones por las cuales los sujetos categorizan los estímulos test de la forma indicada pueden esconderse tanto razones ortográficas, como distribucionales, o fonéticas.

3.5. Experimento 4. La categoría “Alófono [p^h]” en Inglés.

3.5.1. Diseño.

Las diferentes realizaciones de un fonema se denominan *alófonos*. El fonema / p /, por ejemplo, realiza de diferentes formas según el contexto: como una oclusiva bilabial aspirada en

posición acentuada y prevocálica ([p^h]), como una oclusiva bilabial no aspirada tras una [s] perteneciente a la misma sílaba ([p]), etc.

En las literaturas fonética y o fonológica se suele mencionar que, a pesar de las diferencias articulatorias y/o acústicas, los alófonos de un determinado fonema se aceptan, consideran, sienten, u oyen como lo “mismo” (ej. Bloch & Trager 1942: 40; Brown 1977: 13; Gimson 1980: 48; Hockett 1955: 144; Jones 1984: 171; Nathan 1986: 216; O'Connor 1973: 121; Pike 1947: 65; Stampe 1987: 295). Sin embargo, existen diferentes opiniones acerca de la idea de que los hablantes son conscientes o no de la variación alofónica. Algunos investigadores mantienen que los hablantes no son conscientes de dicha variación en absoluto y que tal grado de consciencia se alcanza con un entrenamiento fonético especial y normalmente largo. Otros autores mantienen que los hablantes pueden distinguir tal variación sin un entrenamiento especial al tiempo que afirman que el hecho de que los no son conscientes de la variación alofónica simplemente refleja el hecho de que tal consciencia no es normalmente necesaria, no su incapacidad para percibir tal variación. Otros autores simplemente dicen que los hablantes no son normalmente conscientes de la variación alofónica pero no se decantan por una u otra postura.

El propósito del experimento descrito a continuación es el de obtener evidencia sobre la capacidad o incapacidad de hablantes no entrenados fonéticamente para distinguir una de las variantes alofónicas del fonema / p /, la llamada “p aspirada” (es decir [p^h]), de otras realizaciones del mismo fonema. Las preguntas de investigación formuladas son:

- 1) ¿Pueden los hablantes del inglés clasificar ejemplos de un determinado alófono de un fonema ([p^h] como alófono de / p /) excluyendo otros alófonos del mismo fonema?
- 2) ¿Son capaces de percibir los hablantes tal variación alofónica?

Las hipótesis propuestas son que los hablantes serán capaces de formar dicha categoría alofónica y percibir la variación subyacente a tal distinción.

3.5.2. Método.

3.5.2.1. Sujetos.

Veinte sujetos de edades comprendidas entre los 18 y 45 años fueron seleccionados al azar del grupo inicial de 80 sujetos reclutados para los experimentos CF descritos en esta tesis. Los sujetos tenían las características mencionadas en 3.1.1.2.

3.5.2.2. Estímulos.

Los estímulos del presente experimento consistían de 100 palabras monosilábicas. En la sesión de aprendizaje habían 32 estímulos positivos y 28 negativos. Asimismo, la sesión de test presentaba 22 estímulos positivos y 18 negativos. La sesión experimental no incluyó ningún tipo de estímulo *test*. La tabla 17 ofrece un resumen de los tipos de estímulos usados y la tabla 18 la lista de las palabras usadas.

Tabla 17. Resumen de los Estímulos Usados en el Experimento 4.				
Tipo de Estímulo	SESIÓN DE APRENDIZAJE		SESIÓN DE TEST	
	Subtipo de Estímulo	Número	Subtipo de Estímulo	Número
EJEMPLOS POSITIVOS	Pre-nuclear		Prenuclear	
	[p ^h V-]	23		
	[p ^h _i V-]	5	[p ^h V-]	21
	[p ^h _j V-]	4	[p ^h jV-]	1
	TOTAL	(32)	TOTAL	(22)
EJEMPLOS NEGATIVOS	<i>Prenuclear</i>		<i>Prenuclear</i>	
	[spV-]	6	[spV-]	4
	[sp _i V-]	2		
	[sp _j V-]	2		
	<i>Postnuclear</i>		<i>Postnuclear</i>	
	[-Vp]	6	[-Vp]	5
	[-Vmp]	3	[-Vmp]	3
	[-Vp ^o t]	3	[-Vp ^o t]	2
	[-Vps]	2	[-Vps]	1
	[-Vtp]	2	[-Vsp]	2
	[-Vpθ]	1	[-Vmps]	1
	[-Vsp]	1		
	TOTAL	(28)	TOTAL	(18)

Tabla 18. Lista de Estímulos para el Experimento 4.					
Orden	Estímulo	Positivo (+) Negativo (-)	Orden	Estímulo	Positivo (+) Negativo (-)
SESIÓN DE APRENDIZAJE					
1.	paw	+	53.	poise	+
2.	spend	-	54.	drop	-
3.	push	+	55.	pan	+
4.	rapt	-	56.	spy	-
5.	post	+	57.	clasp	-
6.	pray	+	58.	pots	+
7.	power	+	59.	ship	-
8.	stamp	-	60.	paled	+
9.	pass	+			
10.	up	-			
			SESIÓN DE TEST		

11.	pulse	+	1.	payer	+
12.	plot	+	2.	pass	+
13.	spray	-	3.	pegs	+
14.	purr	+	4.	span	-
15.	pin	+	5.	pen	+
16.	spoon	-	6.	clamp	-
17.	spear	-	7.	wept	-
18.	damp	-	8.	top	-
19.	pond	+	9.	pin	+
20.	depth	-	10.	pea	+
21.	proud	+	11.	spa	-
22.	help	-	12.	pool	+
23.	shop	-	13.	palms	+
24.	pay	+	14.	camp	-
25.	pill	+	15.	pie	+
26.	pence	+	16.	glimpse	- control
27.	cap	-	17.	peace	+
28.	play	+	18.	keep	-
29.	split	-	19.	poles	+
30.	apt	-	20.	spill	-
31.	par	+	21.	pear	+
32.	spring	-	22.	tramp	-
33.	pain	+	23.	pun	+
34.	pile	+	24.	pubs	+
35.	spice	-	25.	lapsed	-
36.	caps	-	26.	pure	+ control
37.	prince	+	27.	group	-
38.	gulp	-	28.	pause	+
39.	pace	+	29.	paved	+
40.	pelt	+	30.	sponge	-
41.	trap	-	31.	gasp	-
42.	plea	+	32.	peel	+
43.	Splay	-	33.	opt	-
44.	paste	+	34.	ape	-
45.	print	+	35.	punch	+
46.	spare	-	36.	pull	+
47.	kept	-	37.	pave	+
48.	lamp	-	38.	pant	+
49.	plough	+	39.	drip	-
50.	priest	+	40.	grasp	-
51.	ropes	-			
52.	puff	+			

3.5.2.3. Procedimiento.

3.5.2.3.1. Instrucciones.

Las instrucciones proporcionadas a los sujetos fueron muy similares a las utilizadas en el experimento 4.

3.5.2.3.2. Sesión de Aprendizaje, Sesión de test, y Entrevista Post-Experimental.

Las sesiones de aprendizaje y test procedieron según lo indicado en la 3.1.1.5. Los sujetos realizaron el experimento individualmente.

3.5.3. Resultados.

3.5.3.1. Sesión de Aprendizaje y Sesión de Test.

De los 20 sujetos que participaron en el experimento se consideró que todos aprendieron la categoría correctamente con una media de 51.40 respuestas correctas (rango 37-59, s.d. 5.15) en la sesión de aprendizaje.

Los números de respuestas correctas (C), incorrectas (I), y nulas (NR) tanto para los estímulos positivos como negativos se muestra en la tabla 19. El mismo tipo de información en relación con los estímulos negativos se muestra en la tabla 20.

Tabla 19. Número de Respuestas Correctas, Incorrectas, y Nulas para los Estímulos Positivos y Negativos del Experimento 4 (Sesión de Aprendizaje).						
Tipo de Estímulo	Tipo de respuesta			% Correctas	Ejemplos	Número de Respuestas
	C	I	NR			
Ejemplos positivos	554	27	59	86.56%	32	640
Ejemplos negativos	473	39	48	84.46%	28	560
TOTAL	1027	66	107	85.58%	60	1200

Tabla 20. Número de Respuestas Correctas, Incorrectas, y Nulas para los Estímulos Positivos y Negativos del Experimento 4 (Sesión de Test).						
Tipo de Estímulo	Tipo de respuesta			% Correctas	Ejemplos	Número de Respuestas
	C	I	NR			
Ejemplos positivos	434	5	1	98.64%	22	440
Ejemplos negativos	346	10	4	96.11%	18	360
TOTAL	780	15	5	97.5%	40	800

Una comparación de los datos mostrados en ambas tablas muestra que los porcentajes de respuestas correctas a los estímulos tanto positivos como negativos incrementa substancialmente en la sesión de test en comparación con los porcentajes en la sesión de aprendizaje. En otras palabras, las respuestas correctas son significativamente más frecuentes en la sesión de test que en la de aprendizaje (85.58% vs. 97.5% respectivamente; $P\text{-valor} = 0.000 < 0.05$ según un contraste de proporciones).

3.5.4. Discusión.

El experimento 4 fue diseñado y realizado para obtener evidencia acerca de la formación de la categoría alofónica [p^h]” y la habilidad de los sujetos experimentales de percibir la variación alofónica entre la variante alofónica [p^h]” y del fonema / p / y algunas de sus restantes variantes. Los resultados obtenidos permiten confirmar las hipótesis formuladas: no sólo fueron los sujetos capaces de formar la categoría sino que también percibieron diferencias en la realización y/o producción de los miembros de la categoría investigada en comparación con las realizaciones alofónicas que sirvieron de estímulos negativos.

CAPÍTULO IV. CONCLUSIONES.

4.1. Introducción.

El propósito de este breve capítulo final es resumir la conexión entre los capítulos previos. El capítulo trata también de comparar los resultados de los cuatro experimentos y sacar algunas implicaciones teóricas. El capítulo también sugiere algunas direcciones para la aplicación futura de la técnica experimental FC en fonología y otras disciplinas relacionadas.

4.2. Categorización, La Categorización de los Sonidos, la Fonología, y La Experimentación.

El capítulo I de esta tesis doctoral comenzó con una descripción de la importancia de la habilidad cognitiva de la categorización para la percepción, la memoria, la cognición, la conducta, y el lenguaje. El capítulo prosiguió con un resumen de los hallazgos experimentales sobre la categorización con humanos adultos, no adultos, y otras especies animales. El capítulo trazó así la historia de la categorización en los últimos 80 años describiendo de una forma pormenorizada los diversos modelos de categorización propuestos.

Si bien el capítulo I se centró en el estudio de los estímulos visuales el capítulo II trató de los estudios llevados a cabo con estímulos auditivos. Dicho capítulo evidenció la importancia de la categorización para la clasificación de los sonidos e hizo especial hincapié en las técnicas utilizadas en el estudio de la categorización de los sonidos. Se argumentó entonces que la categorización de los sonidos por los hablantes podría ayudar a los fonólogos interesados en proponer descripciones fonológicas con pretensión de reflejar el conocimiento que los hablantes tienen de su propia lengua en lograr tales fines. Se delineó entonces la distinción entre una fonología descriptiva y otra psicológica y se argumentó que la experimentación es un método ideal de practicar una fonología psicológica que halla, en la categorización, una fuente inagotable de evidencia acerca de cómo los hablantes perciben, conciben, y tratan la estructura fónica de la lengua.

4.3. CF como un Paradigma Experimental.

Entre las técnicas experimentales más prometedoras para estudiar la categorización de los sonidos por los hablantes se encuentra la de “formación de conceptos”. Dicha técnica fue

descrita en detalle al final del capítulo II y una implementación específica de dicho paradigma fue descrito a su vez al principio del capítulo III, el cual también se encargó de describir cuatro estudios experimentales destinados a recoger evidencia empírica sobre la formación y concepción de cuatro categorías fonológicas y la clasificación de segmentos controvertidos en relación a su pertenencia o no en dichas categorías.

4.3.1. Evidencia Obtenida de las Categorías Individuales.

Los estudios llevados a cabo dejan claro que los sujetos fueron capaces de formar las categorías estudiadas y de que trataron a los segmentos [h], [w], y [j] mayoritariamente como consonantes (experimento 1), a las africadas *no* como oclusivas (experimento 2), a las bilabiales oclusivas orales después de [s] como ejemplos del fonema / p /, y mostraron asimismo que los hablantes pueden percibir la diferencia entre las realizaciones aspiradas del fonema / p / y otras realizaciones no aspiradas. Es necesario resaltar, sin embargo, que aunque muchos de los resultados obtenidos son muy sugerentes, es necesario obtener mucha más evidencia experimental para proporcionar una solución definitiva a los problemas estudiados en este trabajo. Habrá por tanto que diseñar nuevos y mejores experimentos y tratar de validar los resultados obtenidos con otras técnicas experimentales.

4.3.2. Evidencia Obtenida de las Categorías: Comparación entre Categorías.

Es claro, atendiendo a los resultados obtenidos, que algunas categorías resultaron más fáciles de aprender que otras en el sentido que más sujetos las aprendieron, menos errores fueron cometidos, el porcentaje de respuestas correctas en la fase de aprendizaje fue mayor, y las desviaciones típicas de los distintos números de respuestas individuales por sujeto son menores. Teniendo en cuenta estos datos, queda claro que las categorías / p / y [p^h] fueron las más fáciles (/ p / la que más), mientras que las categorías “oclusivas orales” y “sonidos consonánticos” las que menos, habiendo pocas diferencias entre ambas. Podría interpretarse, debido a la aparente relación taxonómica existente entre las cuatro categorías, que la categoría “fonema / p /” puede tener muchos de los rasgos del nivel básico descrito en la sección 1.2.1.3.5., aunque posteriores estudios experimentales deberían examinar esta posibilidad, que es meramente especulativa atendiendo a los resultados de la presente tesis.

4.3.3. La Utilidad del Paradigma de “Formación de Conceptos”: Direcciones Futuras para la Aplicación de la Técnica CF.

El presente trabajo ha hecho uso de la técnica experimental conocida como “formación de conceptos”, que ha resultado útil para lograr evidencia experimental acerca de las cuestiones fonológicas investigadas. Sin embargo, dicha técnica puede utilizarse no sólo para proseguir obteniendo evidencia experimental acerca de las categorías investigadas u otras relacionadas, sino también puede utilizarse para examinar prácticamente cualquier cuestión fonológica. Además la técnica puede emplearse en otros campos o líneas de investigación tales como la comparación de fenómenos fonológicos entre lenguas, el aprendizaje de la fonología de una segunda lengua por hablantes de un determinado idioma, etc. Parece que sólo la imaginación del investigador puede ser el límite de los temas que pueden ser estudiados con el paradigma CF. El presente trabajo pretende ser un ejemplo de la utilidad de aquella técnica en la fonología para obtener evidencia experimental que ayude a formular análisis fonológicos que sean acordes con la realidad psicológica de los patrones fónicos para el hablante.

REFERENCIAS.

Véase la sección “REFERENCES” de la versión inglesa del presente trabajo.