



UNIVERSIDAD DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO
TESIS DOCTORAL

Connected Speech Processes in Spontaneous
Discourse: Linking /r/, T-tapping, and Yod
Coalescence in Non-Rhotic English

Procesos de la cadena del habla en el discurso
espontáneo: linking /r/, t-tapping y yod coalescence
en inglés no rótico

Dña. Ana María Sánchez Leira

2023



UNIVERSIDAD DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO
TESIS DOCTORAL

Connected Speech Processes in Spontaneous
Discourse: Linking /r/, T-tapping, and Yod
Coalescence in Non-Rhotic English

Procesos de la cadena del habla en el discurso
espontáneo: linking /r/, t-tapping y yod coalescence
en inglés no rótico

Autora: Dña. Ana María Sánchez Leira

Director: D. José Antonio Mompeán González



DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD
DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

Aprobado por la Comisión General de Doctorado el 19-10-2022

D./Dña. Ana María Sánchez Leira

doctorando del Programa de Doctorado en

869 – Programa de Doctorado en Artes y Humanidades: Bellas Artes, Literatura, Teología, Traducción e Interpretación y Lingüística General e Inglesa (Plan 2013)

de la Escuela Internacional de Doctorado de la Universidad Murcia, como autor/a de la tesis presentada para la obtención del título de Doctor y titulada:

Connected Speech Processes in Spontaneous Discourse: Linking /r/, T-tapping, and Yod Coalescence in Non-Rhotic English

Procesos de la cadena del habla en el discurso espontáneo: linking /r/, t-tapping y yod coalescence en inglés no rótico

y dirigida por,

D. José Antonio Mompeán González

DECLARO QUE:

La tesis es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, de acuerdo con el ordenamiento jurídico vigente, en particular, la Ley de Propiedad Intelectual (R.D. legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, modificado por la Ley 2/2019, de 1 de marzo, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), en particular, las disposiciones referidas al derecho de cita, cuando se han utilizado sus resultados o publicaciones.

Si la tesis hubiera sido autorizada como tesis por compendio de publicaciones o incluyese 1 o 2 publicaciones (como prevé el artículo 29.8 del reglamento), declarar que cuenta con:

- La aceptación por escrito de los coautores de las publicaciones de que el doctorando las presente como parte de la tesis.*
- En su caso, la renuncia por escrito de los coautores no doctores de dichos trabajos a presentarlos como parte de otras tesis doctorales en la Universidad de Murcia o en cualquier otra universidad.*

Del mismo modo, asumo ante la Universidad cualquier responsabilidad que pudiera derivarse de la autoría o falta de originalidad del contenido de la tesis presentada, en caso de plagio, de conformidad con el ordenamiento jurídico vigente.

En Murcia, a 12 de julio de 2023

Fdo.:

Esta DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD debe ser insertada en la primera página de la tesis presentada para la obtención del título de Doctor.

Información básica sobre protección de sus datos personales aportados	
Responsable:	Universidad de Murcia. Avenida teniente Flomesta, 5. Edificio de la Convalecencia. 30003; Murcia. Delegado de Protección de Datos: dpd@um.es
Legitimación:	La Universidad de Murcia se encuentra legitimada para el tratamiento de sus datos por ser necesario para el cumplimiento de una obligación legal aplicable al responsable del tratamiento. art. 6.1.c) del Reglamento General de Protección de Datos
Finalidad:	Gestionar su declaración de autoría y originalidad
Destinatarios:	No se prevén comunicaciones de datos
Derechos:	Los interesados pueden ejercer sus derechos de acceso, rectificación, cancelación, oposición, limitación del tratamiento, olvido y portabilidad a través del procedimiento establecido a tal efecto en el Registro Electrónico o mediante la presentación de la correspondiente solicitud en las Oficinas de Asistencia en Materia de Registro de la Universidad de Murcia

Page intentionally left blank

To my family,

Acknowledgements

The journey I decided to undertake towards the completion of a PhD thesis has given me the opportunity to broaden my knowledge in an area of linguistics which sparked my interest in research several years ago. A journey that would have been impossible to fulfil without the support of a number of people who helped me walk through it in different ways, and to whom I would like to express my sincere gratitude.

First of all, I am immensely indebted to Jose A. Mompeán, my mentor, who has helped and advised me during all these years. I am extremely grateful for all your unconditional help and support, for your unlimited patience and for all the valuable lessons, explanations, and meetings that you shared with me.

This thesis has also been possible to conduct thanks to my family, to whom I dedicate it. To my parents, Leandro and Ángeles, who have always been there for me and whose encouragement guided me every single day. You always believed in me. You both taught me that no matter how challenging times are because if you carry on, you can make it. Thanks. To my siblings, Leandro, Esther, and Silvia, with whom I have shared so many moments. They always offered me their time and help when I needed it the most, even when they suffered my absence on many occasions. Thanks to you, too.

To my friends, Carolina, Cynthia, Carmen, María José, and Noelia. This journey would have been incredibly more difficult without them. They have always been there for me when I needed them, no matter what, even when we were far away, they felt really close. Words cannot describe how much in debt I am with you for all the moments we have shared and all your support while working on this project. You have encouraged me every step of the way, and that meant the world to me. Sharing this experience with me and understanding every single feeling I was going through was more helpful than you can imagine. Thanks, I will forever be in debt to you, my friends.

I would also like to thank Elena, my boss, and some special workmates, Mari Carmen, Arancha, Arturo, Miguel, and Sarah, who have made going to work more pleasant and helped me in many ways. Special thanks to Elena, Arancha and Mari Carmen, because I feel grateful for many reasons. Elena, thanks for being so understanding and kind, and making things easy for me at difficult times. Arancha, thanks for listening to me when you were busy with your own academic things, and for your never-ending positive attitude and interest in phonetic details that I could tell you. To Mari Carmen, to whom I

owe eternal gratitude for everything she has done for me over the past few years. It is impossible to put into words how much you helped me in every single way you could, work-related, thesis-related, every aspect-related. Thanks, Mari, your help has been invaluable, and I am sure it will continue to be.

I would also wish to thank Joanna Kopaczyk, who made it possible that I could do a research stay at the University of Glasgow. Joanna, thanks for your patience and kindness, and thanks for giving me the opportunity to experience Glaswegian life, meet fantastic people and discover the world of Scots to me in our meetings. It was an unforgettable experience.

Finally, I would like to thank to the Statistics Team at the University of Murcia as well for guiding and helping me with the different statistical tests and figures.

TABLE OF CONTENTS

List of figures	i
List of tables	iii
Resumen	v
Abstract	xiii
INTRODUCTION	1
CHAPTER I: Connected speech processes and the Speech Chain	5
1.1. Types of connected speech processes	8
1.1.1. Modification processes	14
<i>Assimilation</i>	14
<i>Regressive, progressive, and coalescence assimilation</i>	15
<i>Adjacent and distant assimilation</i>	19
<i>Partial and total assimilation</i>	20
<i>Dissimilation</i>	21
<i>Contact and distant dissimilation</i>	22
1.1.2. Sandhi processes	23
<i>Tone sandhi</i>	25
<i>Consonant sandhi</i>	27
<i>French liaison</i>	28
<i>/r/-sandhi</i>	30
1.1.3. Reduction processes	32
<i>Lenition</i>	32
<i>Tapping</i>	34
<i>Deletion</i>	36
1.1.4. Insertion processes	38
1.1.5. Phonological approaches	42

CHAPTER II: The variables under investigation	57
2.1. Linking /r/	57
2.1.1. Impressionistic accounts	59
2.1.2. Theoretical analyses.....	62
2.1.3. Empirical studies.....	64
2.2. T-tapping	68
2.2.1. Impressionistic accounts	70
2.2.2. Theoretical analyses.....	71
2.2.3. Empirical studies.....	74
2.3. Yod coalescence	82
2.3.1. Impressionistic and theoretical accounts	84
2.3.2. Empirical accounts.....	86
2.4. Factors affecting linking /r/, t-tapping and yod coalescence	90
<i>Phonetic factors</i>	90
<i>Linguistic factors</i>	93
<i>Sociolinguistic factors</i>	95
<i>Usage-based factors</i>	97
2.5. Concluding remarks.....	99
CHAPTER III: Objectives and methodology	101
3.1. Main objectives	101
3.2. Research questions and hypothesis	103
3.3. Methodology.....	105
3.3.1. Speakers	105
<i>A relatively homogenous sample</i>	106
<i>The individual speakers</i>	107
3.3.2. Data	111
3.3.3. Procedure.....	117

<i>Identification of potential cases</i>	117
<i>Auditory and acoustic analyses</i>	119
<i>Variables and data coding</i>	128
<i>Phonetic factors</i>	128
<i>Linguistic factors</i>	129
<i>Usage-based factors</i>	130
3.3.4. Statistical analysis	131
CHAPTER IV: Results and discussion	133
4.1. Results	133
4.1.1. Study 1: linking /r/	134
4.1.2. Study 2: t-tapping	143
4.1.3. Study 3: yod coalescence	152
4.1.4. General results	160
4.2. Discussion	169
4.2.1. Study 1: linking /r/	169
4.2.2. Study 2: t-tapping	177
4.2.3. Study 3: yod coalescence	185
4.2.4. General discussion	190
CHAPTER V: Conclusion, limitations of the study, directions for future research, and implications of the findings	195
5.1. Main findings	195
5.2. Limitations and directions for future research	198
5.3. Implications for applied work	201
References	205

List of figures

Figure 1. The Speech Chain model, adapted from Denes and Pinson (1993, p. 5).	6
Figure 2. Alameen and Levis' categories, adapted from Alameen and Levis (2015, p. 162).	10
Figure 3. Proposed taxonomy for the present study following Alameen and Levis (2015).	14
Figure 4. Tone sandhi representation with four tones applied to the same syllable. Adapted from Huang, Y., Athanasopoulou, A., & Vogel, I. (2018).	26
Figure 5. Third tone sandhi applied in connected speech in Mandarin Chinese.	26
Figure 6. Third tone sandhi applied in connected speech in Mandarin Chinese.	27
Figure 7. Outline of phonological approaches.	43
Figure 8. Hierarchical representation of the relevant domains in prosodic phonology.	52
Figure 9. The ranking of faithfulness and markedness constraints for <i>law</i> is from Uffmann (2007a, p. 466).	54
Figure 10. Andrew Marr with George Osborne at <i>The Andrew Marr Show</i> . Source: Creative Commons (https://www.bbc.co.uk/programmes/b01mh0ds).	113
Figure 11. Andrew Marr with Theresa May at <i>The Andrew Marr Show</i> . Source: Creative Commons (https://www.bbc.co.uk/programmes/b08plldm).	113
Figure 12. Example of linking /r/ in <i>here is</i> by Theresa May (31st October 2010)	122
Figure 13. Example of absence of linking /r/ in <i>there and</i> by Harriet Harman (31st July 2011).	122
Figure 14. Example of laryngeal strategy (i.e, creaky voice) in <i>poor ambition</i> by George Osborne (31 st January 2010).	123
Figure 15. Example of /r/ in <i>but also</i> by Ed Balls (20th June 2010).	124
Figure 16. Example of /t/ in <i>out of</i> by George Osborne (31st January 2010).	125
Figure 17. Example of [ʔ] in <i>what Ed</i> by Harriet Harman (7 th July 2013).	126
Figure 18. Example of /ʃ/ in <i>this year</i> by George Osborne (31st January 2010).	127

Figure 19. Example of /t/ followed by /j/ in <i>but you</i> by Theresa May (24 th April 2016)..	127
Figure 20. Bar plot by cases produced and not produced for linking /r/.	134
Figure 21. Linking strategies.	135
Figure 22. Linking /r/ production by gender.	136
Figure 23. Correlation matrix for linking /r/ tokens. Actual cases produced, laryngeal strategies, and tokens (left and right-hand tokens and word collocations) in the AMS corpus and BNC Spoken corpus.	138
Figure 24. Correlation matrix for linking /r/ production and word-collocations and stress patterns combinations for linking /r/ in the AMS corpus.	142
Figure 25. Bar plot by produced and not produced cases for t-tapping.	143
Figure 26. T-tapping production by gender.	146
Figure 27. Correlation matrix for t-tapping tokens. Actual cases produced, laryngeal strategies, and tokens (left and right-hand tokens and word collocations) in the AMS corpus and BNC Spoken corpus.	148
Figure 28. Correlation matrix for t-tapping production and word-collocations and stress patterns combinations for t-tapping in the AMS corpus.	151
Figure 29. Bar plot by cases produced and not produced for yod coalescence.	153
Figure 30. Yod coalescence production by gender.	154
Figure 31. Correlation matrix for yod coalescence tokens. Actual cases produced, laryngeal strategies, and tokens (left and right-hand tokens and word collocations) in the AMS corpus and BNC Spoken corpus.	156
Figure 32. Correlation matrix for yod coalescence production and word-collocations and stress patterns combinations for yod coalescence in the AMS corpus.	159
Figure 33. Overall cases produced and cases not produced.	161
Figure 34. Data per connected speech phenomena.	161
Figure 35. Males and Females overall production.	163

List of tables

Table 1. Information about speakers: gender, year of birth (YoB) and speaker's age at the time of production (SaP).....	109
Table 2. Information about speakers: gender, year of production (YoP), non-available interviews (NA), number of interviews (NoI) which were analysed for each speaker, and the total amount of words (W) per speaker.	115
Table 3. Information about speakers: total amount of tokens per phenomena and per speaker.	116
Table 4. Cases produced and not produced divided by speakers for linking /r/.	136
Table 5. Ten most frequent tokens for linking /r/.	140
Table 6. T-tapping strategies divided into categories.	144
Table 7. Number and rate of potential cases produced in the AMS corpus divided by speakers (t-tapping).	145
Table 8. Overall cases produced and not produced divided by gender for t-tapping. .	146
Table 9. Glottal stop production for t-tapping.	147
Table 10. Ten most frequent tokens for t-tapping.....	150
Table 11. Overall yod coalescence production divided by speakers.	154
Table 12. Cases produced and not produced divided by gender for yod coalescence.	155
Table 13. Ten most frequent tokens for yod coalescence.	158
Table 14. Cases produced divided by phenomena.....	162
Table 15. Cases produced by speakers for linking /r/, t-tapping, and yod coalescence.	163
Table 16. Word-class combinations per phenomena: potential/actual cases per phenomenon and rate of actual cases	164
Table 17. Lexical frequency of the left-hand and right-hand tokens in the AMS corpus.	165
Table 18. Collocational frequency in the AMS corpus.	166
Table 19. Lexical frequency in the reference corpus (BNC Spoken).	166

Table 20. Collocational frequency according to the BNC Spoken corpus.	167
Table 21. Stress pattern combinations per phenomena: potential/actual cases per phenomenon and rate of actual cases.	168

Resumen

Los fenómenos del habla en cadena han sido durante mucho tiempo el centro de diversos estudios en el campo la lingüística, en particular, en fonética y fonología; sin embargo, el estudio de los procesos del discurso en cadena no solo ha atraído la atención de fonetistas y fonólogos, sino que también se ha convertido en un campo de creciente interés entre los sociolingüistas (ej., Dressler & Wodak, 1982; Ingram, 1989; Kerswill, 1985). Los diferentes procesos del habla en cadena pueden ser entendidos como procesos que ocurren en el contexto de la interacción hablada y se pueden interpretar en términos articulatorios, acústicos, o de características perceptuales, que corresponden a las fases de la cadena del habla de la producción, transmisión y percepción del habla (Denes & Pinson, 1993).

La presente tesis doctoral pretende contribuir a la ya existente investigación empírica en fonética mediante el estudio de la variación de tres fenómenos del habla en cadena (linking /r/, t-tapping, y yod coalescence) en el discurso espontáneo, mediante el análisis de varios factores que, según la literatura, podrían afectar de manera potencial la producción de estos fenómenos. El principal objetivo de este estudio es aportar evidencia empírica sobre el uso de linking /r/, t-tapping, y yod coalescence en la variedad de inglés no róticos conocida como *Received Pronunciation* (RP).

Linking /r/ es un fenómeno que ocurre cuando un sonido /r/ a final de palabra está presente mediante la ortografía, pero ausente en la pronunciación de la palabra de manera aislada, siendo únicamente pronunciado en el habla en cadena en los acentos no róticos (ej., Collins & Mees, 2003; Wells, 1982; Wells & Colson, 1971). T-tapping es uno de los procesos de debilitamiento consonántico más comunes mediante el cual una consonante plosiva (/t, d/) se pronuncia como un tap alveolar (/r/) (Alameen & Levis, 2015). Yod coalescence, o asimilación por coalescencia, es un tipo de asimilación de tipo recíproco por el cual dos sonidos, principalmente, una plosiva en posición final de palabra (t, d/), se asimila con la palatal aproximante (/j/) en posición inicial de la palabra siguiente (ej., Crystal, 2003; Roach, 2009; Roach, Setter & Esling, 2011). A pesar de que este proceso de tipo coarticulatorio ha sido durante mucho tiempo asociado únicamente con consonantes plosivas, las alveolar fricativas (/s, z/) en posición final de palabra también pueden asimilarse con la palatal aproximante (/j/) en posición inicial de la palabra siguiente (Cruttenden (2014). De esta manera, los fenómenos del discurso en cadena normalmente implican unos cambios en la pronunciación de los lexemas de manera

aislada, lo que es a menudo entendido como parte de un proceso que ocurre en un dominio mayor que el de las palabras aisladas.

Durante los últimos 150 años, numerosas teorías lingüísticas, que han estudiado los procesos del habla en cadena, los han descrito de varias maneras dependiendo del enfoque teórico que adoptasen. Las teorías y paradigmas fonológicos, que han tratado los procesos del habla en cadena, ayudaron a comprender como los fenómenos del discurso en cadena. De entre todas las teorías y paradigmas, un enfoque funcional, basado en el uso parece el más apropiado para el análisis de los procesos de la cadena hablada ya que los datos lingüísticos que se obtendrán del análisis del presente estudio no representan material lingüístico descontextualizado si no que es material lingüístico dependiente del contexto en el que está siendo producido (ej., Bybee, 1999, 2001).

En cuanto a la investigación sobre la naturaleza específica de los diferentes procesos de habla en cadena objeto de estudio, las aproximaciones son de tres tipos: impresionista, teórica y empírica. Las aproximaciones impresionistas y teóricas pueden ser más limitadas en cuanto a ámbito de estudio, pues las primeras están basadas en descripciones del hablante del propio fonetista o una comunidad idealizada de hablantes, y las segundas se centran en interpretaciones de los fenómenos siguiendo un determinado marco teórico (ej., Haugen, 1938; Jones 1909/1956; McMahon, 2002; Wells, 1982). Sin embargo, los estudios empíricos emplean procedimientos metodológicos diversos y una operacionalización de los datos que permiten obtener conocimientos sobre los factores potenciales que motivan la producción de los procesos de la cadena hablada, que ayuda a revisar y actualizar los enfoques de tipo teórico (ej., Glain, 2012; Hannisdal, 2006; Mompeán & Gómez, 2011; Patterson and Connine, 2001; Pavlík, 2016).

En su contribución a la siempre creciente investigación fonética, los hallazgos principales muestran que existe variación en la cadena hablada y que está motivada por factores de diversa índole. Entre ellos, destacan factores de tipo lingüístico (ej., el estatus gramatical de la palabra y los tipos de combinaciones de palabras), sociolingüístico (ej., género), uso (ej., frecuencia léxica y de colocación), y fonético (ej., patrón prosódico) que se ha demostrado que tienen un impacto en la producción de los fenómenos. En cuanto a la producción de linking /r/ se refiere, los principales hallazgos revelan que el uso de linking /r/, aunque no es categórico, sí que exhibe altos porcentajes de producción, y que las estrategias de tipo laríngeo (ej., glotalización, creaky voice) son comunes cuando no se usa linking /r/.

Los resultados obtenidos en la literatura también evidencian que los factores fonéticos y de uso constituyen las variables que más influencia ejercen sobre la variación del fenómeno (ej., Hannisdal, 2006; Mompeán & Gómez, 2011; Pavlík, 2016). En cuanto al t-tapping, los hallazgos más importantes revelan que hay diferencia entre géneros (masculino y femenino) que la probabilidad de que t-tapping ocurra es cuasicategoría en inglés americano. Los datos aportados por los estudios empíricos también revelan un alto porcentaje de producción de tap en el límite entre palabras funcionales. Los enfoques de corte empírico sobre el t-tapping también revelan algunas incongruencias entre estudios sobre el grado de impacto de los factores de uso. Mientras ciertos estudios indican que ni la frecuencia ni la colocación parecen tener impacto alguno en la producción de este fenómeno, otros estudios, por el contrario, indican que lexemas de alta frecuencia sí que motivan la producción de t-tapping (ej., Connine 2004; Hannisdal, 2006; Patterson and Connine, 2001). En cuanto a los factores que motivan la producción de la yod coalescence, cabe mencionar que parece ser el fenómeno de la cadena hablada, de los tres mencionados, el que se ve más afectado por factores de corte sociolingüístico según los diferentes estudios. Como revelan los estudios, el fenómeno de la yod coalescence presenta una gran variación en torno a factores como el género y la edad, debido a que está normalmente asociado a hombres más que a mujeres, y a generaciones de hablantes jóvenes.

Respecto al impacto de factores fonético en la producción de yod coalescence, la investigación en fonología histórica revela que este fenómeno estaba estigmatizado en sílabas acentuadas, mientras que estudios recientes revelan que el fenómeno de la yod coalescence en sílabas acentuadas está cada vez más extendido (ej., Beal et al., 2020; Glain, 2012, 2014). Tal y como muestran los estudios previos, factores de tipo lingüístico, sociolingüístico, de uso y fonéticos motivan la producción de los procesos de la cadena hablada de manera diferente. A pesar de que los estudios llevados a cabo hasta ahora han revelado ciertos patrones que son relevantes al uso de linking /r/, t-tapping, y yod coalescence en inglés RP, la investigación sobre el discurso espontáneo u oral es escasa, ya que la mayoría de los estudios empíricos se han centrado mayormente, aunque no de manera exclusiva, en contextos informativos o de noticiario de la BBC o en un habla guionizada. La variedad de inglés RP fue escogida como el acento objeto de análisis por ser un acento que es abunda en cuanto a materiales de índole televisiva se refiere, y por ser el acento estándar de Inglaterra. Esto ha permitido un gran margen a la hora de

seleccionar discursos largos en los que los fenómenos del habla en cadena objeto de estudio pudieran ser investigados.

La contribución de la presente tesis doctoral es, por tanto, empírica, usando un procedimiento metodológico de análisis basado en corpus. El motivo de realizar investigación con una metodología basada en corpus se debe a que los estudios sobre el discurso espontáneo son más escasos en la literatura, como previamente se ha mencionado. Por lo tanto, poco se sabe sobre la producción real de los procesos de la cadena hablada en un contexto espontáneo. Esta decisión de analizar procesos de habla en cadena en un contexto conversacional asume que dicho contexto podría revelar datos fonéticos importantes. Además de usar un método basado en corpus, un enfoque basado en el uso hace posible modelar la probable correlación que exista entre los fenómenos lingüísticos y las variables (Bybee, 1999, 2001).

Siguiendo la literatura especializada, la presente tesis doctoral pretende describir la correlación entre los tres fenómenos del habla en cadena previamente mencionadas y varios factores: fonéticos (patrón acentual), lingüísticos (tipo de palabra y combinaciones de palabra), sociolingüísticos (género), y de uso (frecuencia léxica y de la colocación). Para ello, los contextos potenciales de linking /r/, t-tapping, y yod coalescence han sido identificados y analizados en busca de presencia o ausencia de la producción de los distintos fenómenos en un corpus a pequeña escala de entrevistas. Esta tesis intenta, por tanto, contribuir a la creciente investigación empírica en fonética y fonología mediante el estudio de linking /r/, t-tapping y yod coalescence en RP contemporáneo, y esclarecer la influencia de los estos factores en los procesos de la cadena hablada en un uso real (conversacional) de la lengua.

La investigación del presente estudio está guiada por las siguientes preguntas de investigación (PI) que permitan explorar el estado actual de linking /r/, t-tapping, y yod coalescence en RP contemporáneo y qué factores motivan su producción en el discurso en cadena:

PI1. ¿Cómo de frecuente es el uso de linking /r/, t-tapping, y yod coalescence en un uso conversacional de la lengua?

PI2. ¿Qué factores condicionan la variación en la producción de linking /r/, t-tapping, y yod coalescence?

Para responder a estas PI, la presente tesis examina cómo se comportan los diferentes procesos del habla en cadena anteriormente mencionados en un uso conversacional de la lengua. Los fenómenos del discurso (linking /r/, t-tapping, y yod coalescence) que constituyen el objeto de estudio del presente estudio son descritos en detalle teniendo en cuenta diversas perspectivas: investigación no empírica que use enfoques impresionistas y teóricos, así como investigación empírica basada en la evidencia de datos (ej., experimentos, estudios de corpus, etc.).

Basadas en la literatura especializada, las hipótesis principales implican que los procesos son no solamente variables si no que el porcentaje de producción entre ellos es diferente debido a los factores lingüísticos, sociolingüísticos, de uso y fonéticos que afectan su producción.

El estudio se centra en el habla de seis políticos británicos entrevistados en el programada *The Andrew Marr Show*, un programa de entrevistas que se emitía en el canal BBC One desde mediados de los años 2000 hasta principios de 2020. Los seis hablantes elegidos para el presente estudio fueron seleccionados siguiendo el criterio de ser políticos británicos y hablantes de inglés RP que fueron entrevistados en algún punto de la historia del programa. La presente tesis doctoral, por tanto, pretende alcanzar una mejora comprensión de los tres fenómenos de la cadena hablada objeto de estudio en el discurso espontáneo mediante la compilación de un corpus oral a pequeña escala de un programa de entrevistas de televisión, analizándolos en un entorno conversacionales natural y espontáneo, para contribuir de forma significativa a la investigación empírica ya existente. Los datos extraídos de las entrevistas fueron codificados y analizados auditiva y acústicamente.

Para analizar la transcripción de las entrevistas, se exploraron contextos ortográficos potenciales. Dado que el estudio se centraba en el análisis de los fenómenos del discurso en cadena en los límites de palabra, las búsquedas de los contextos potenciales para cada uno de los fenómenos no solo incluyó la búsqueda de combinación de letras concretas en los límites de palabra para encontrar los contextos potenciales para cada uno de los tres fenómenos, sino también una amplia y meticulosa búsqueda de potenciales letras mudas (ej., /h/ en *hour* para los contextos de linking /r/), pausas ortográficas para identificar si realmente eran pausas fonéticas o no, contracciones (ej., *there's*), y repeticiones involuntarias por parte de los hablantes, entre otras búsquedas.

Después de buscar los contextos potenciales, se llevó a cabo un análisis auditivo inicial para comprobar si los casos potenciales de los hablantes correspondían con los contextos fonéticos requeridos para la producción de los fenómenos. Cuando se identificaron los contextos ortográficos y potenciales, se analizaron auditivamente por segunda vez. Al inspeccionar los casos potenciales auditivamente, la producción de los fenómenos objeto de estudio fue categorizada siguiendo una decisión binaria de ocurre/no ocurre basada en el análisis perceptual, dejando el análisis de los casos ambiguos para ser comprobados por el software (*Praat*) para su desambiguación.

Los resultados del análisis muestran que los tres fenómenos exhiben una variabilidad de producción significativa. En término del porcentaje de producción, los hallazgos revelan que linking /r/ tiene un porcentaje más elevado de producción en comparación con los otros dos fenómenos. Aunque la producción de t-tapping y yod coalescence difiere de la elevada producción extraída del análisis de linking /r/, el porcentaje de producción de t-tapping y yod coalescence obtenido en el presente estudio revela que, a pesar de no tener porcentajes elevados de producción, son fenómenos del discurso en cadena que no son exclusivos de variedades no estándares de la lengua inglesa o de contextos situacionales coloquiales.

En cuanto a los factores que parecen motivar la producción de estos procesos de la cadena hablada, para la muestra de hablantes que se ha analizado (políticos británicos), los hallazgos revelan que una combinación de factores lingüísticos (ej., el tipo de palabra y su combinación), de uso (ej., frecuencia léxica y de la colocación), y fonéticos (ej., sílabas acentuadas/inacentuadas en los contextos potenciales) influye en cómo de variable es la producción de estos fenómenos, mientras que los factores de tipo sociolingüístico (ej., género) parecen no ser tan relevantes.

Los hallazgos principales del estudio indican que la producción de linking /r/ es notablemente alta, aunque no categórica, lo que guarda coherencia con su función lingüística y el uso de estrategias laríngeas para evitar el hiato cuando no se aplica dicho fenómeno, como apuntan varios estudios (ej., Cox, Palethorpe, Buckley & Bentink, 2014; McCarthy, 1993; Mompeán, 2022; Mompeán & Gómez, 2011). La ausencia de una producción categórica puede ser debido a que los hablantes evitan producir linking /r/ a modo de hipercorrección para evitar, de esta manera, producir intrusive /r/, cuya pronunciación está tradicionalmente estigmatizada (ej., Brown, 1988a; Cruttenden, 2014; Gimson, 1962/1980; Mompeán, 2022; Wells, 1982; Wells & Colson, 1971).

Aunque los hallazgos revelan que t-tapping y yod coalescence no son fenómenos de la cadena hablada muy frecuentes en el discurso espontáneo en inglés RP, este estudio sugiere que pueden ser considerados como innovaciones recientes en el acento, y, por lo tanto, no son procesos del discurso exclusivos de otras variedades del inglés ni de variedades o acentos no estándar (ej., Wells, 1982). En términos de factores sociolingüísticos, las hablantes femeninas destacan por tener los porcentajes más altos de linking /r/, un hallazgo que puede justificarse con el argumento de que las mujeres tienden a usar variantes de más prestigio o estándar que los varones (ej., Coates, 2004), aunque algunos estudios previos no encontraron ningún tipo de correlación entre la producción de linking /r/ y el género (ej., Bauer, 1984; Hannisdal, 2006; Mompeán & Mompeán-Guillamón, 2009; Pavlík, 2016).

En cuanto al t-tapping, su producción es más baja de lo esperada, lo que quizás sea debido a la edad de los hablantes (mediana edad). Aunque algunos estudios previos encontraron también porcentajes bajos de producción en hablantes jóvenes de RP, concretamente chicos que habían asistido a colegios privados (ej., Alderton, 2022), hay evidencia de que hablantes jóvenes de otras variedades del inglés que usan el tap frecuentemente (ej., Holmes, 1994). Como hipótesis, se puede plantear que se hubiera obtenido un porcentaje de producción mayor si se hubieran analizado hablantes más jóvenes. Aparte de la variable de la edad de los hablantes, el género, al igual que en estudios anteriores parece afectar la producción de taps (ej., Hannisdal, 2006; Holmes, 1994). Los hallazgos del estudio muestran que los hombres usan más taps que las mujeres, como indican estudios previos tanto en RP como en otras variedades del inglés, como es el inglés de Nueva Zelanda, donde se encontró una correlación entre el género y la producción de taps (ej., Badia Barrera, 2015; Holmes, 1994).

En cuanto a la producción de yod coalescence, hombres y mujeres exhiben porcentajes de producción similares, al igual que en el estudio de Hannisdal (2006), aunque otros estudios encontraron que el género motivaba la producción de taps (ej., Altendorf, 2003; Amos, 2007). Con respecto a los factores lingüísticos, el estudio revela que tanto linking /r/ como t-tapping se producen con frecuencia en combinaciones de palabras funcionales, un hallazgo que se encuentra en la literatura previa (ej., Bauer, 1984; Foulkes, 1997; Patterson & Connine, 2001; Viollain, 2014). Una posible explicación podría venir dada por factores fonéticos, ya que las palabras funcionales van normalmente inacentuadas (Hannisdal, 2006), lo que estaría relacionado con los hallazgos del presente

estudio de la estructura prosódica de las combinaciones de palabras analizadas, ya que, al igual que indican estudios previos, las combinaciones de palabras inacentuadas representan el patrón acentual donde linking /r/ y t-tapping se producen con un porcentaje más alto (ej., Eddington & Elzinga, 2008; Mompeán, 2022; Warner & Tucker, 2011). Sin embargo, yod coalescence se produce con más frecuencia en combinaciones de palabras formadas por dos palabras de contenido y en combinaciones de palabras acentuadas, hallazgos también encontrados en estudios anteriores (ej., Amos, 2007; Glain, 2012; Hannisdal, 2006).

Los hallazgos del presente estudio relacionados con los factores de uso revelan que tanto la frecuencia léxica como la frecuencia de la colocación tienen un impacto en la producción de linking /r/, t-tapping y yod coalescence. Esto subraya el hecho de que, en el caso de linking /r/, un porcentaje alto de linking /r/ se deba a la lexicalización de la linking /r/ como algo característico del discurso en cadena en ciertas colocaciones léxicas (ej., Bybee, 2001). En cuanto a la correlación entre t-tapping y yod coalescence y la frecuencia de uso, una posible explicación apuntaría a que las palabras funcionales de alta frecuencia estarían motivadas por un proceso de reducción, el cual afecta a las palabras más comunes o frecuentes primero mediante un proceso de difusión léxica (Shockey, 2003).

De igual modo, los hallazgos también indican que existe cierto grado de fijación en ciertos lexemas y combinaciones de palabras en cuanto a la producción de yod coalescence. Una conclusión interesante extraída de la presente tesis doctoral es que múltiples factores condicionan la producción de los fenómenos del habla en cadena objeto de estudio (linking /r/, t-tapping, y yod coalescence), aunque el impacto de cada factor es diferente. Un hallazgo notable que hace diferente al presente estudio de investigaciones anteriores basadas en corpus es el hecho de que, los resultados obtenidos proporcionan un avance empírico, contribuyendo así a la incipiente necesidad de actualizar la descripción sobre el inglés RP (ej., Cruttenden, 2014; Wells, 1982), y sobre cómo operan en esta variedad acentual de la lengua inglesa los procesos de la cadena del habla.

Abstract

Connected speech phenomena have long been the centre of many studies in the field of linguistics. The current doctoral thesis aims to contribute to the existing empirical body of research in phonetics by studying the variation in three connected speech phenomena (linking /r/, t-tapping, and yod coalescence) in spontaneous speech through the analysis of several factors that, according to the specialised literature, may potentially affect their production.

Linking /r/ is a type of sandhi phenomenon that refers to the use of a word-final r-sound, which is orthographically present but absent in the citation form of a given lexical item (e.g., Collins & Mees, 2003; Wells, 1982; Wells & Colson, 1971). T-tapping is one of the most common consonantal weakening processes whereby a stop consonant (/t, d/) turns into an alveolar tap ([ɾ]) (Alameen & Levis, 2015). Yod coalescence, or coalescence assimilation, is a type of reciprocal assimilation whereby two sounds, namely a word-final plosive (/t, d/) (but also the fricatives (/s, z/), fuse a following word-initial palatal approximant (/j/) (e.g., Cruttenden, 2014; Crystal, 2003; Roach, 2009; Roach et al., 2011).

In terms of the research accounts, they are mainly of three types: impressionistic, theoretical, and empirical approaches (e.g., Glain, 2012; Hannisdal, 2006; Haugen, 1938; Jones 1909/1956; McMahon, 2002; Mompeán & Gómez, 2011; Patterson & Connine, 2001; Pavlík, 2016; Wells, 1982). In their contribution to the ever-growing phonetic body of data, the key findings obtained from these studies suggest that variation occurs at the connected speech level and that it is motivated by diverse factors. Among them, phonetic, linguistic, sociolinguistic, and usage-based factors have proven to have an impact on their production.

Following the specialised literature, the present dissertation aims to describe the correlation between the three connected speech phenomena mentioned earlier and several factors: phonetic (e.g., prosodic environment), linguistic (e.g., grammatical status and word-class collocations), sociolinguistic (e.g., gender), and usage-based factors (e.g., lexical frequency and collocational frequency). In doing so, the potential contexts of linking /r/, t-tapping, and yod coalescence have been identified and analysed for the presence or absence of rhoticity, tap production, and coalescence assimilation in a small-scale corpus of interview speech.

The research conducted in the current study addresses the following general research questions (RQ) to explore the status of linking /r/, t-tapping, and yod coalescence in contemporary RP and what factors motivate their production in connected speech:

RQ1. How frequent is the use of linking /r/, t-tapping, and yod coalescence in conversational language use?

RQ2. What factors condition variation in the production of linking /r/, t-tapping, and yod coalescence?

The study focuses on the speech of six UK political figures, all of them RP speakers, interviewed at *The Andrew Marr Show*, a talk show broadcast on BBC One from the mid-2000s to the early 2020s. Therefore, the intended contribution of this study is, empirical, involving a central methodological corpus-based procedure. In addition to using a corpus-based method, a usage-based approach makes it possible to model a correlation between the linguistic phenomena under study as well as the variation across the distinct factors (Bybee, 1999, 2001).

The results show that the three phenomena exhibit significant variability. The key findings indicate that linking /r/ production is notably frequent, although not categorical, which is consistent with its hiatus-breaking function alongside the use of laryngeal strategies to avoid hiatus when linking /r/ is not applied, as pointed out by previous studies (e.g., Cox et al., 2014; McCarthy, 1993; Mompeán, 2022; Mompeán & Gómez, 2011). Even though the findings suggest that t-tapping and yod coalescence are not very frequent connected speech phenomena in spontaneous speech in RP, the present study reveals that the processes can be recent developments in RP and that they are not exclusive of other varieties of English or non-standard variants (e.g., Wells, 1982).

One interesting conclusion from the present thesis is that from the analysis of the factors that condition linking /r/, t-tapping, and yod coalescence usage is that multiple variables contribute to their variation, yet the weight of each of the factors is different across the phenomena. A remarkable finding that makes the current study different from previous ones (mostly corpus-based) is the fact that the results represent relevant and novel insights concerning empirical research and contribute to the necessity for an updated description of RP and the operationalisation of connected speech phenomena in this accentual variety (e.g., Cruttenden, 2014; Wells, 1982).

Keywords: connected speech, linking /r/, t-tapping, yod coalescence

INTRODUCTION

Speaking is an extraordinarily complex linguistic skill, it is unconscious as well as automatic, and connected speech processes are among the most complex and phonological phenomena (Alameen & Levis, 2015; Beňuš, 2021). Studies concerning connected speech phenomena, where both word-internal and word-boundary processes are the focus, have not only drawn the attention of phoneticians and phonologists but also among sociolinguists (e.g., Dressler & Wodak, 1982; Ingram, 1989; Kerswill, 1985). To put in context the study of connected speech processes involved in speech production, they can be interpreted in terms of the Speech Chain, which is a simple model of spoken communication that highlights the transformation of the speakers' communicative purpose into a decoded message in the listeners' mind (Denes & Pinson, 1993). The Speech Chain model enables to contextualise the occurrence of connected speech processes in human spoken interaction, and connected speech phenomena can be interpreted in terms of their articulatory, acoustic, and/or perceptual characteristics as well as their structural properties. In addition to this, this simple model helps us to contextualise the study of connected speech phenomena, which can be characterised as phonetic variations occurring in discourse when compared to citation forms (Gimson 1962/1980; Jones, 1909/1956).

Consequently, connected speech processes are said to have great relevance in speech production from the viewpoint of speech perception. Between the complexities of spoken language as the result of coarticulation and its segmentation in terms of phonemes, connected speech can be described as the changes citation forms undergo given articulatory and temporal constraints that operate in spontaneous speech (Hieke, 1987). However, modifications of the structural properties of segments may operate not only in spontaneous conversation but also in lexical items regardless of the phonetic surrounding context. While the former refers to connected speech processes, the latter implies allophonic variation. Therefore, connected speech processes usually imply some changes to the citation forms of lexical items; and they are frequently understood as occurring in a larger domain than the lexical word in isolation, as is the case of allophonic variation (Hannisdal, 2006). For this reason, this current research aims to delve into the connected speech phenomena of linking /r/, t-tapping, and yod coalescence in spontaneous,

conversational discourse and to shed some light on the factors that motivate their production.

The current dissertation is divided into five chapters. The first two chapters of the thesis, Chapters I and II, provide the background to the study. Chapter I explores some significant concepts related to connected speech and includes detailed sections on several types of connected speech processes, with an adapted taxonomy using one of the most recent and comprehensive taxonomies up to the present time (Alameen & Levis, 2015). The adapted taxonomy will be used in the present study to discuss some of the connected speech processes identified and studied in the literature (e.g., Cruttenden, 2014; Crystal, 2003; Wells, 1982). In addition to the description of connected speech processes, Chapter I concludes with a section that examines the numerous linguistic theories that have proliferated during the past 120 years and that have long been concerned with the description of human language.

Chapter II comprises three different sections, each for one of the phenomena under investigation. The three main sections in this chapter describe in depth each of the phenomena, which complement the succinct definitions given in the introductory chapter, alongside impressionistic, theoretical, and empirical approaches. A detailed revision of the different types of approaches that can be found in the specialised literature for each connected speech phenomenon is useful to gain a more comprehensive picture of the phenomena under investigation by contextualising existing work on the three phenomena (e.g., Amos, 2007; Bauer, 1984; Glain, 2012; Hannisdal, 2006; Kahn, 1976; Mompeán & Mompeán-Guillamón, 2009).

The main findings of the approaches are summarised in a section exclusively dedicated to the factors affecting the production of the connected speech phenomena, which is useful for the interpretation of the results. This section is subdivided into the four main factors which, according to the relevant literature, may potentially affect linking /r/, t-tapping, and yod coalescence production (i.e., phonetic, linguistic, sociolinguistic, and usage-based factors). Chapter II concludes with some notable remarks on such factors.

Chapter III is devoted to detailing the object of study, the research questions, and the hypotheses that guided the research. This chapter includes a detailed description of the methodology, the speaker sample that was used, the data collection procedure, the

identification and coding of the variables adopted for the study as well as the auditory and acoustic analyses that were carried out.

The results and discussion of the analysis conducted for the present dissertation concerning how linking /r/, t-tapping, and yod coalescence operate in spontaneous discourse are described in Chapter IV. The chapter consists of two main sections. The first part corresponds to the description of the quantitative results for each of the three connected speech processes under investigation. This first half concludes with a last section on general results for the three connected speech phenomena. The second part deals with the discussion of the results. The discussion section is again divided into three sections, one for each of the connected speech phenomena. Chapter IV concludes with a last section containing a general discussion of the key findings. Finally, Chapter V summarises the main findings obtained, presents the limitations of the study, and provides directions for future research.

CHAPTER I: Connected speech processes and the Speech Chain

Speaking is undoubtedly one of the most complex cognitive and linguistic skills, and connected speech processes are among the most diverse, complex, and fascinating phonological phenomena (Alameen & Levis, 2015). These include, among others, alterations such as deletions, additions, liaison, or even the modification of one sound resulting in a new one. The study of connected speech processes not only has drawn the attention of phoneticians but has also been a field of growing interest among sociolinguists (e.g., Dressler & Wodak, 1982; Ingram, 1989; Kerswill, 1985). For this reason, this dissertation presents a sociophonetic study of connected speech processes that investigates phonetic variation in speech production.

To contextualise the study of connected speech processes in the production of speech in communication, the latter can be understood in simple terms as occurring in human spoken language interactions. To model such interactions, the Speech Chain represents a simple model of spoken communication and highlights the transformation of the speakers' communicative purpose into a decoded message in the listeners' minds. This model consists of five stages on three distinct levels: linguistic, physiological, and acoustic (Cruttenden, 2014; Denes & Pinson, 1993). The physiological level can be further divided into speech production and speech perception, while the acoustic level corresponds to speech transmission.

According to Cruttenden (2014), the formulation of the message is created in the speakers' minds to reflect a communicative intention (linguistic or psycholinguistic level). At this stage, listeners arrange their thoughts to encode messages into linguistic forms by choosing words and phrases to convey meaning and ordering them according to grammatical rules (Denes & Pinson, 1993).

Second, the nervous system transmits a message to the organs of speech that will create a specific pattern of sound (physiological or articulatory level). At this level, neural and muscular activity occurs, resulting in speech production. Third, the airflow modulated by the articulators becomes a speech sound wave -a simple example of a complicated phenomenon that involves wave motion- that is transmitted through an elastic medium (typically air) between the speaker and the listener (physical or acoustic level), resulting in speech transmission. The listeners' hearing mechanism, which is activated by the sound waves, leads to the fourth stage where the sound wave is perceived and transmitted to the

listeners' brain through the sensory nerves (physiological level). This is generally known as speech perception. At this stage, which represents the listeners' physiological level, neural activity occurs (Denes & Pinson, 1993).

Finally, the last stage is where linguistic interpretation occurs as the listeners' mind decodes the speakers' message (linguistic level). At this phase, each phonetic element used by the speakers to create their message has a number of distinctive features that the listeners have to decode in order to understand the message (Denes & Pinson, 1993; Manchaiah, 2007).

Figure 1 represents a visual illustration of the Speech Chain model as described by Denes and Pinson (1993). As mentioned above, the model helps us contextualise the study of connected speech processes, which can be characterised as the phonetic variations taking place in discourse when compared to citation forms (Gimson 1962/1980; Jones, 1909/1956). Therefore, connected speech processes are said to occur in speech production but also have great relevance from the speech perception point of view.

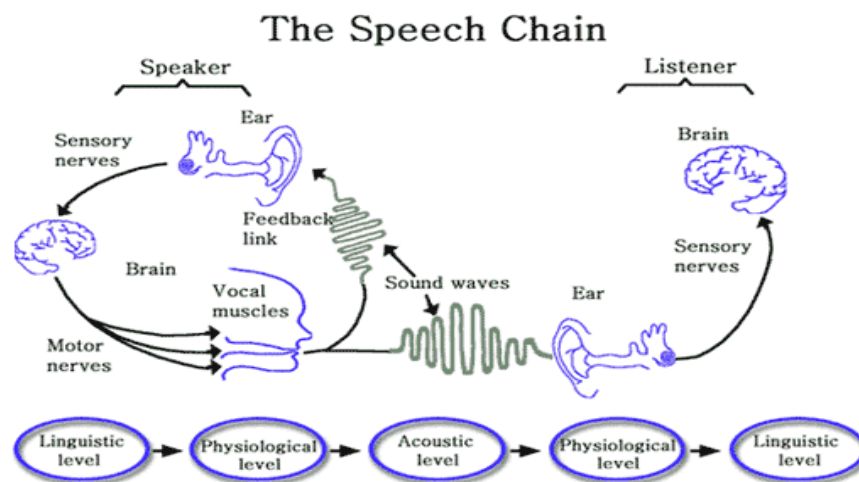


Figure 1. The Speech Chain model, adapted from Denes and Pinson (1993, p. 5).

The Speech Chain model allows us to contextualise the occurrence of connected speech processes in human spoken interaction, and connected speech processes can be interpreted in terms of their articulatory, acoustic, or perceptual characteristics. That is, phonetically, or in terms of their structural properties, that is, phonologically. Phonetics

deals with the articulatory, acoustic, and auditory description and analysis of speech sounds with reference to their production, transmission, and perception. On the other hand, the subject matter of phonology is sound patterns by describing their nature and use (e.g., Cruttenden, 2014; Giegerich, 1992; McMahon, 2002; Roca & Johnson, 1999).

Between the complexities of spoken language as the result of coarticulatory movements and the clean segmentation typically carried out by phonologists in terms of phonemes, connected speech can be described as “the changes which conventional forms undergo given the temporal and articulatory constraints upon spontaneous, casual speech” (Hieke, 1987, p. 41). In other words, connected speech evidences the difference between the ideal phonological form and the modified phonetic form.

Thus, understanding connected speech processes may become a complete challenge to intelligibility because the modifications might be unpredictable (Alameen & Levis, 2015). This mismatch produced between the citation forms and the modified segments may create lexical ambiguity. As a result, the success of this mapping process depends on retrieving relevant information effectively as well as the degree to which the representation of the input is well-matched with a stored lexical representation.

These phonological modifications are derived from different factors, such as the linguistic context, the rate of speech, or even the register being used. In rapid and casual speech, for example, successive articulations may either overlap or blend their characteristics completely, given that the speech organs tend to link sounds together using either the same or an intermediate articulatory gesture to save time as well as energy (Alameen & Levis, 2015; Clarey & Dixon, 1963).

However, connected speech processes not only occur when speakers unconsciously pay little attention to how they articulate their utterances in spontaneous discourse. Some literate speakers avoid using many connected forms in their discourse in formal situations because they consider them colloquial or inappropriate to the communicative context. For this reason, the process of production of speech does not have to do exclusively with style, since social, linguistic, and extralinguistic factors also affect how speech is produced (Alameen & Levis, 2015; Anderson-Hsieh, Rinev & Koehler, 1994; Lass, 1984).

Modifications of the structural properties of segments may operate not only in discourse but also in lexical items regardless of the phonetic context. While the former refers to connected speech processes, the latter implies allophonic variation. Allophonic

variation occurs word-internally in individual words and involves variation in the phonetic makeup of segments, whereas connected speech processes are normally referred to as phonetic alternations that are dependent on several phonetic, linguistic, sociolinguistic, and usage-based factors (Farnetani & Recasens, 2010; Ingram, 1989).

Hence, connected speech processes typically imply some changes to the citation forms of lexical items; these are often understood as taking place in a larger domain than the lexical word in isolation, as in the case of allophonic variation, thus needing a string of words or connected speech.

The assimilation of the alveolar /n/ to a velar nasal /ŋ/ before a velar plosive /k/ in *pancake* (/ˈpænkɛk/ > [ˈpæŋkɛk]), for example, or the elision of /ə/ leading to a syllabic consonant as in *London* (/ˈlʌndən/ > [ˈlʌndŋ]) can be considered to be cases of allophonic variation. The assimilation of the alveolar /n/ to a following bilabial (e.g., *ten boys* /ten ˈbɔɪz/ > [tʰem ˈbɔɪz]) or velar (e.g., *ten girls* /ten ˈgɜːlz/ > [tʰeŋ ˈgɜːɹz]) at word boundaries, or the /h/-dropping in object pronouns at word boundaries (e.g., *tell him* /tel hɪm/ > [tel ɪm]), however, can be considered cases of connected speech processes (Cruttenden, 2014).

1.1. Types of connected speech processes

When examining connected speech processes, two remarkable issues need to be addressed: differences in terminology and the scarcity of relevant research.

On the one hand, there seems to be no agreement on how to name and even classify connected speech processes, according to Alameen and Levis (2015). Regarding the issue of differences in terminology and classification schemes, the names given to connected speech processes and phenomena vary widely. Moreover, throughout the phonetic and phonological literature, there has been little agreement on how to classify connected speech processes (e.g., Blevins & Garret, 2004; Ingram, 1989).

As Alameen and Levis (2015) stated, “[n]ot only do different researches and materials designers use different terms for CSPs (e.g., sandhi variations, reduced forms, absorption), they also do not always agree on how to classify them” (p. 162). As a way

of illustration, Ingram (1989) at the end of the 20th century grouped connected speech processes into five broad different categories: lenitions, reductions and deletions, assimilations, linking processes, strengthening processes, and phonetically unmotivated or obscure processes.

Nevertheless, more recently, at the beginning of the 21st century, connected speech processes appear to be catalogued under different descriptions or labels. Blevins and Garret (2004), for example, pointed out that assimilation, dissimilation, deletion, insertion, or coalescence, among other phenomena, have traditionally been categorised as types of alternations between sounds, in which further subdivisions of phenomena are included within each category.

Alameen and Levis (2015), however, also at the beginning of the 21st century, proposed six main categories: linking, deletion, insertion, modification, reduction, and multiple processes (see Figure 2). Since this is one of the most recent and comprehensive taxonomies to date, it will be used here to initially discuss some of the connected speech processes identified and studied in the literature.

On the other hand, experimental studies on connected speech processes seem to be scarce. This may be due to the fact the variables in those studies are normally not controllable so it is difficult to predict the number of tokens a particular process may yield, making the application of statistical measures difficult (Shockey, 2003, p. 9).

In addition, the different modifications that citation forms may undergo in connected speech cannot even be predicted by applying phonological rules (Alameen & Levis, 2015; Anderson-Hsieh et al., 1994; Lass, 1984; Temperley, 1987).

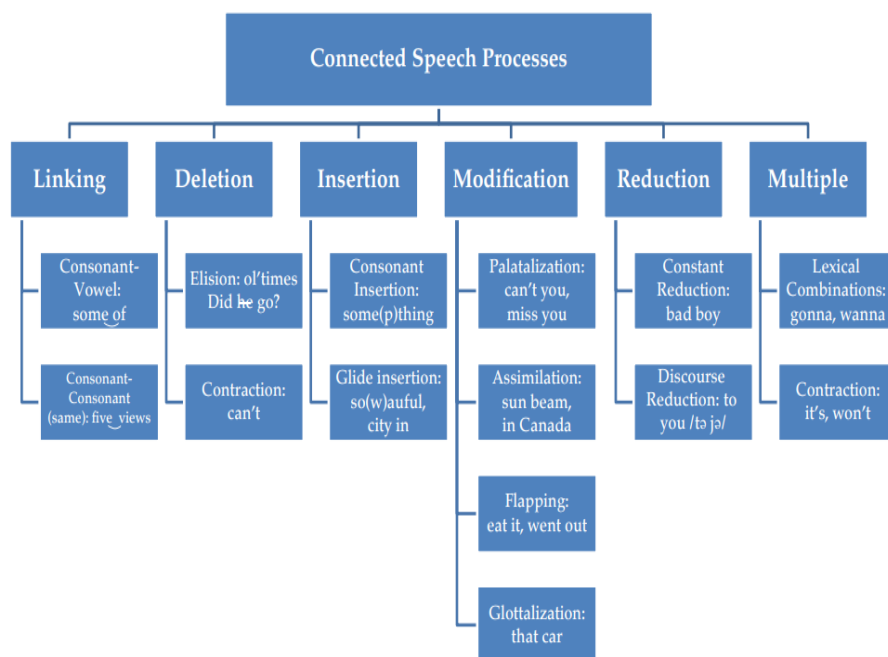


Figure 2. Alameen and Levis' categories, adapted from Alameen and Levis (2015, p. 162).

As the figure above illustrates, according to this taxonomy, linking refers not only to processes that join a consonant with the next vowel (1) but also to two consonants (2), while deletion processes include deletion (3) as well as contractions (4) (Alameen & Levis, 2015, p. 162):

(1) *some of* /'sʌm əv/ > ['sʌm əv]

(2) *five views* /faɪv vju:z/ > [faɪv:jʊ:z]

(3) *did he go?* /dɪd hi 'gəʊ/ > [dɪd i 'gəʊ]

(4) *cannot* /'kænət/ > *can't* [kɑ:nt]

In example (1), the linking process of a consonant followed by a vowel does not imply any phonological change but links the two words. Similarly, example (2) shows a

gemination process due to identical sounds occurring at word boundaries. In example (3), the glottal fricative [h] is elided, a very frequent process that unstress pronominal forms starting with /h/ (e.g., *he, him, her*, etc.) undergo in English. Finally, example (4) is a clear illustration of a contraction process which includes not only elisions (i.e., the vowel [ə] from the negative particle has been lost) but also vowel changes (i.e., the vowel [æ] changes to [ɑ:]).

Insertion processes, on the other hand, refer to the addition of sounds such as consonants or glides, as in examples (5) and (6) below:

(5) *the idea of* > *the idea(r) of* /ðɪ: aɪ'dɪə əv/ > [ðɪ: aɪ'dɪər əv]

(6) *I am what I am* [aɪ ˈjæm wɒt aɪ ˈjæm]

In example (5), an epenthetic consonant /r/ is inserted at word junctures to avoid hiatus, while example (6) shows the insertion of glides to link two vowels across word boundaries.

Modification processes, therefore, involve changes in pronunciation that substitute one phoneme for another in connected speech. Together with the processes mentioned previously, other types of processes that fall within the scope of modification are assimilation of place, manner or voicing, taps, and t-glottaling. Therefore, modification processes, according to Alameen & Levis (2015), group several connected speech processes, such as coalescence and assimilation, as illustrated in examples (7) and (8), respectively:

(7) *did you* /'dɪd ju:/ > ['dɪdʒu]

(8) *on point* /ɒn 'pɔɪnt/ > [ɒm 'pɔɪnt]

In example (7), the two sounds /t, j/ fuse or coalesce into a new segment [tʃ], whereas in example (8), the assimilation process whereby the alveolar nasal /n/ changes to a bilabial nasal [m] due to the influence of the following bilabial plosive /p/.

Other modification processes include tapping and glottalisation, as shown in (9) and (10) below:

(9) *part of* /'pɑ:t əv/ > ['pɑ:r əv]

(10) *can't make it* /kɑ:nt 'meɪk ɪt/ > [kɑ:nʔ 'meɪk ɪt]

In example (9), the voiceless plosive /t/ is produced as an alveolar tap intervocally while in example (10), the voiceless plosive /t/ is produced as a glottal fricative.

Reduction processes refer to both consonant reduction as well as vowel reduction, which normally occur in weak forms and are motivated by rhythmic factors, as in (11). In phonological terms, this type of process implies a weakening in articulation.

(11) *to you* /tu: ju:/ > [tə jə]

Example (11) shows vowel reduction of the strong vowel /u:/ of the citation forms of the function words *to* and *you* for rhythmic reasons –the destressing of those function words.

Multiple processes, as defined by Alameen and Levis (2015), can be considered to be those that exhibit “multiple CSPs simultaneously in each lexical combination” (p. 163), as shown in (12) below:

(12) *going to* > *gonna* /'gəʊɪŋ tu:/ > ['gɒnə]

In this example, simultaneous processes are applied. The velar nasal [ŋ] changes to [n]; there is also a vowel reduction in *to* from [u:] to [ə] as well as in the diphthong [əʊ] to the monophthong [ɒ], and the deletion of the sounds [ɪ] and [t].

As previously stated, the classification of connected speech processes varies widely, as does connected speech terminology, which results in a lack of a unified classification for the different connected speech phenomena. In the present dissertation, the taxonomy proposed by Alameen and Levis (2015) has been followed, although it has been adapted. Alameen and Levis' (2015) taxonomy is, therefore, the starting point for the following classification that is proposed. Firstly, the proposed new scheme reduces the classification to four broad categories (i.e., modification, linking, reduction and insertion processes); secondly, it considers connected speech phenomena not included in theirs; and thirdly, it classifies some of the previously discussed phenomena (e.g., t-tapping) in a different category (see also Figure 3).

The first category (i.e., modification processes), includes assimilation and dissimilation processes (see 1.1.1.), which refer to the changes sounds undergo given the influence of neighbouring sounds. The second category (i.e., sandhi processes) comprises phenomena such as French liaison, tone sandhi, or the well-known /r/-sandhi (see 1.1.2.), in which phonemes are articulated with the function of joining word boundaries. The third category (i.e., reduction processes) includes processes that imply a simplification in one or some of the parameters of the articulation of sounds due to the phonetic context. Some reduction processes included under this label are lenition, tapping, or deletion (see 1.1.3.). Finally, the last category (i.e., insertion processes) includes that type of articulation in which the speaker inserts a sound not supported by the orthography. Prothesis (insertion at the beginning of a word), epenthesis (insertion in the middle of a word) and paragoge (insertion at the end of a word) are some examples of connected speech processes where an epenthetic sound is inserted (see 1.1.4.).

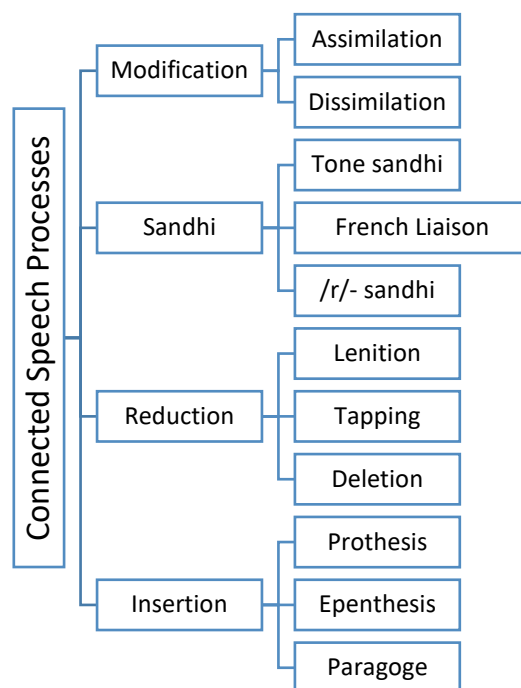


Figure 3. Proposed taxonomy for the present study following Alameen and Levis (2015).

1.1.1. Modification processes

Modification processes are types of processes that often create such dramatic changes. They include processes such as assimilation (progressive, regressive, coalescent) and dissimilation.

Assimilation

The first category refers to connected speech processes involving phonetic or allophonic segment alternations. An assimilation process, as Crystal (2003, p. 38) describes it, refers to “the influence exercised by one segment upon the articulation of another so that the sounds become more alike, or identical”. That is, two concomitant segments become more alike, regardless of whether the two segments belong to the same syllable/word, as assimilation may operate both within word and at word junctures (Brown, 2014).

This type of connected speech process is a frequent phenomenon across morpheme as well as word boundaries in natural languages (McMahon, 2002). It is also a type of connected speech phenomenon that varies widely according to the speaker’s

speech tempo and style, which is likely to be found in casual, spontaneous speech rather than in a slow and careful discourse (Roach, 2009).

Assimilation involves a type of coarticulation that mostly affects the place of articulation of a segment, manner of articulation or voicing, resulting in contextual variability (e.g., Brown, 2014; Hannisdal, 2006; Pavlík, 2009). Researchers have distinguished different types of assimilation based on parameters such as the direction in which the assimilation takes place (regressive, progressive, or coalescent assimilation), the distance between the sounds involved in the assimilation (adjacent vs. distance assimilation), or the degree or extent to which assimilation takes place (partial vs. total assimilation).

Regressive, progressive, and coalescence assimilation

In terms of the direction in which the assimilation process works, three types of assimilation can be distinguished: a) regressive; b) progressive; and c) coalescent. Even though assimilation processes are widely found in natural languages, as previously stated, Shiler (2000) pointed out that not all assimilation processes are equally frequent. In this respect, regressive assimilation is by far the most frequent type (Crowley & Bower, 2010), whereas progressive assimilation is a rather infrequent phenomenon (e.g., Crowley & Bower, 2010; Cruttenden, 2014; Shapiro, 1995; Skandera & Burleigh, 2016).

Regressive assimilation can modify the place of articulation, manner of articulation or voicing of the sound to the left of the assimilation. Regressive or anticipatory assimilation occurs when one sound changes because of the influence of the following sound (Sihler, 2000). In other words, the force or direction of the change operates backwards (Crowley & Bower, 2010). In this respect, as Roach (2009) describes it, “[a]ssimilation of place is most clearly observable in some cases where a final consonant (C^f) with alveolar place of articulation is followed by an initial consonant (C^i) with a place of articulation that is not alveolar” (p. 111). Some examples of assimilation of place are illustrated in (13) and (14) below:

(13) *that person* /ðæt 'pɜ:sən/ > [ðæp 'pɜ:sən]

(14) *ten bikes* /ten 'baɪks/ > [tem 'baɪks]

In example (13), the alveolar plosive [t] in *that* changes its place of articulation and becomes a bilabial plosive [p] before the bilabial plosive [p] in *person* (Roach, 2009, p. 111). In (14), the alveolar nasal /n/ has been influenced by the following bilabial plosive /b/ and has consequently adopted the bilabial characteristic of the stop consonant becoming a bilabial nasal /m/ (Crystal, 2003, p. 40).

Another example of regressive assimilation is the phrase *Church Street* that can be assimilated as [ˈtʃɜːtʃ ˈtriːt]. In this phrase, the alveolar fricative [s] acquires properties of a post-alveolar fricative [ʃ] given the influence of the preceding post-alveolar affricate [tʃ] (Skandera & Burleigh, 2016, p. 93).

Assimilation of manner is nonetheless found if the segments are produced in very rapid and casual speech. This type of assimilation is achieved when a word-final consonant adopts the manner of articulation of the following word-initial consonant, as in (15) below:

(15) *that side* /ðæt ˈsaɪd/ > [ðæs ˈsaɪd]

In the example above, the word-final alveolar plosive [t] becomes an alveolar fricative [s] given the influence of the following alveolar fricative /s/ (Roach, 2009, p. 112).

Assimilation of voice occurs when the voicing of the segments changes from voiced to voiceless, or vice versa, from voiceless to voiced, due to the influence of the neighbouring sound (Brown, 2014). Assimilation of voice is found across word boundaries when a lenis consonant (i.e., voiced) word-finally is devoiced given the influence of the following fortis consonant (i.e., voiceless) word-initially. For instance, the word-final segments of verbs such as *have* lose their voiced quality in connected speech, as in (16) below (Brown, 2014, p. 100).

(16) a. *have* /hæv/ cf. *have to* [hæf tə]

In example (16), the voiced labiodental fricative [v] becomes a labiodental fricative [f] when followed by the preposition *to*, which begins with voiceless /t/, given the influence of the following voiceless alveolar plosive [t] (Roach, 2009).

Progressive or lag assimilation occurs when one sound changes because of the articulation of the preceding sound. In other words, the force or direction of the change operates forwards (Crowley & Bowern, 2010). An example of progressive assimilation is (17), provided by Crystal (2003, p. 40):

(17) *lunch score* /'lʌntʃ 'skɔ:/ > ['lʌntʃ 'ʃkɔ:]

In example (17), the place of articulation of the alveolar fricative [s] is modified and realised as a postalveolar given the influence of the preceding /tʃ/, also postalveolar.

Another example is when a plosive or nasal follows a word-initial dental fricative /ð/ across word boundaries, as the fricative becomes similar in manner of articulation to the preceding consonant, as exemplified by Roach (2009, p. 112) in (18) below:

(18) *read these* /'ri:d ði:z/ > ['ri:d di:z]

In (18), the dental fricative [ð] is realised with a dental place of articulation given the influence of the preceding alveolar plosive /d/.

Coalescent (or reciprocal) assimilation refers to a type of coarticulatory process whereby two sounds - a word-final alveolar plosive /t, d/ assimilates or coalesces with the following word-initial palatal approximant /j/ - assimilate due to one segment exercising articulatory influence on the articulation of another leading to coalescence, resulting in the post-alveolar affricates /tʃ, dʒ/ (e.g., Crystal, 2003; Roach, 2009).

Yod coalescence is, therefore, a process that modifies not only the place of articulation but also the manner of articulation; however, alveolar plosives /t, d/ do not undergo voicing assimilation. While both /d/ and /j/ are voiced and voicing is preserved,

they coalesce into [ɖ]; however, the voiceless alveolar plosive [t] does not acquire the voicing of /j/ when it changes into [tʃ].

The process can also involve the alveolar fricatives /s, z/, which may coalesce with a word-initial semivowel /j/ resulting in [ʃ] and [ʒ] (Wells 1982, p. 331), respectively, as exemplified by Cruttenden (2014, p. 313) in (19), and (20):

(19) *miss you* /mɪs ju:/ > [mɪʃu:]

(20) *as yet* /əz jet/ > [əʒet]

In the examples above, the alveolar fricatives [s, z] and the palatal approximant [j] have merged into voiced postalveolar affricates [ʃ] and [ʒ]. Another example of coalescent assimilation is found in (21), from Crystal (2003, p. 82).

(21) *could you?* /'kʊd ju/ > ['kʊɖʒu]

In this example, the voiced alveolar plosive [d] and the palatal approximant [j] have merged into a voiced postalveolar affricate [ɖ]. This is an example of what is known as yod coalescence: the merging of phonemes /t, d/ when followed by /j/ into the affricate phonemes /tʃ, ɖʒ/ (e.g., Roach, 1992; Shockey, 2003; Wells, 1982, 1994a, 2000).

This assimilation process is not only understood as a type of progressive assimilation but also as a subcategory of assimilation of place, which has also been referred to as palatalization since the articulation becomes palatalised given the influence exercised by the palatal approximant on the preceding alveolar segment (e.g., Roach, 2009; Shockey, 2003). However, this categorisation is unclear as yod coalescence implies mutual assimilation of sounds, as Hannisdal (2006) highlighted, “the term is somewhat misleading because it is only the first element, the alveolar, which is palatalised” (p. 120), as exemplified in (22) and (23) below:

(22) *this year* /ðɪs jɪə/ > [ðɪʃ ɪə] or [ðɪʃ jɪə]

(23) *last year* /lɑːst jɪə/ > [lɑːʃ ɪə] or [lɑːʃ jɪə]

In the examples above, the alveolar fricative [s] and the palatal approximant [j] have merged into a voiced postalveolar affricate [ʃ].

Adjacent and distant assimilation

Apart from the direction of the assimilation, a common distinction often made in the literature considers the distance between the sounds involved in the assimilation. Based on this, we can distinguish adjacent from distance (or long-distance) assimilation.

Adjacent assimilation happens when a sound is influenced by an adjoining sound. This type of assimilation seems to be the most usual form of assimilation according to Burridge and Bergs (2016). For instance, the articulation of a nasal before stops changes to bilabial by the influence of the next segment, as in example (24):

(24) *ten beans* /ten 'bi:nz/ > [tem 'bi:nz]

In the example above, the alveolar nasal [n] changes the place of articulation to a bilabial nasal [m] given the influence of the following bilabial plosive /b/.

Distance assimilation refers to the case whereby a sound is influenced by a distant one. Even though it is rare in the case of consonants, long-distance consonant assimilation, or ‘consonant harmony’ can be found in Sidaama, a language of Ethiopia where the causative suffix /-is/ is realised as [iʃ] when postalveolar fricatives or affricates appear in the preceding stem (Rose, 2011, p. 1814), as in (25):

(25) /hank'-is/ ‘cause to get angry’ vs. /miʃ-iʃ/ ‘cause to despise’

Another type of long-distance assimilation is often referred to as vowel harmony and it is long-distance insofar as the vowel triggering the change is often separated from the affected vowel by several consonants (Reis, 2006). A clear example could be found in Murcian Spanish *mañanas* ‘tomorrow’ pronounced as [mæɲænæ] in which the plural marker is represented acoustically by the openness of the vowel likely to carry the plural suffix. Then, in order not to confuse singular and plural numbers, the rest of the syllables are pronounced with open vowels (Campoy & Jiménez-Cano, 2003; Hernández-Campoy & Trudgill, 2002). This kind of long-distance vowel assimilation exemplified by Murcian Spanish is an example of a synchronic type of alternation.

Related to this, the term metaphony is a type of sound change by which vowels are assimilated (Kaze, 1989). As a case in point, the case of progressive assimilation of a vowel given the influence of the next vowel sound is called *umlaut*, from German meaning ‘sound alternation’, also called mutation, a very frequent phenomenon in Germanic languages (Malmkjaer, 2002).

Partial and total assimilation

In terms of the extent to which assimilation takes place, another common distinction made is between partial and total assimilation.

Assimilation is considered to be partial if the assimilated sound retains at least one of its original phonetic features and adopts only some of the phonetic features of another sound, so that the unchanged features may be easily distinguished (Crowley & Bower, 2010). For example, in *that day* /t/ becomes /d/ when preceding /d/, pronounced as /ðæd 'deɪ/ > [ðæd 'deɪ]. In this example, both /t/ and /d/ are alveolar since they are produced in the same place of articulation, however, they are not completely identical, since /t/ is voiceless while /d/ is voiced.

Assimilation is understood as complete or total assimilation if the assimilated sound adopts all the phonetic features of another sound and becomes identical to it, as exemplified in (26) below:

(26) *this shoe* /ðɪs 'ʃu:/ > [ðɪʃ 'ʃu:]

In example (26), the alveolar fricative [s] is realised with a postalveolar place of articulation given the influence of the following postalveolar fricative /ʃ/ (Roach, 2009, p. 111). Since the alveolar was already a voiceless fricative, as is the case of /ʃ/, the assimilation is total.

A process that also operates in this context is called gemination. A geminated consonant implies only one articulatory process where the outcome of the process is a double or geminate segment due to consonant strengthening (Blevins & Garret, 2004; Hock, 1991).

Dissimilation

Another process involving the modification of segments is dissimilation, whereby two similar sounds become different or less alike given the influence exercised by one sound segment upon the articulation of the other (Crystal, 2003; Hannahs & Bosch, 2018). Given this, dissimilation can be considered to be the opposite phenomenon of assimilation (Bye, 2011; Crowley & Bower, 2010; Crystal, 2003).

According to Grammont (1936), dissimilation refers to two or more sounds that lose their phonetic characteristics because of being phonetically similar (as cited in Posner, 1961, p. 26). Dissimilation could be regarded as a natural or general linguistic process that operates to preserve some distinctiveness in the phonetic features of the segments (Johnson, 1973, p. 1).

Dissimilation is often regarded as an irregular and unpredictable connected speech process that reduces the number of common phonetic features between two sounds (Hall, 2007). Dissimilation has been generally grouped with other similar speech processes, such as assimilation, metathesis, and epenthesis, under the category of sporadic changes (Posner, 1961); and, together with other connected speech phenomena, such as metathesis, they are often treated in the specialised literature as speech errors or spoonerisms, which represent a faulty phonetic production (Hock & Joseph, 2009).

As to why dissimilation occurs, some researchers claim that one of the main reasons may be to reduce the lack of difficulty when articulating (Fasold & Connor-

Linton, 2014). This is an assumption based on the fact that when two adjacent sounds are similar but not the same, articulation is difficult for speakers. In other words, making two sounds more different helps hearers to discriminate segments since the speaker is responsible for making their speech as intelligible as possible (e.g., Ohala, 1981), although it is also argued that dissimilation cases might also arise from perceptual errors on the part of the listener (McMahon, 1994, p. 16).

Two types of dissimilation processes can be found: contact (or immediate) dissimilation and distance dissimilation. The former occurs when the segment that has been dissimilated is contiguous to the segment responsible for dissimilation; in contrast, the latter occurs when the segment likely to produce the dissimilation is non-adjacent (Campbell, 2013; Ohala, 1981, Posner, 1961).

Contact and distant dissimilation

One example of contact dissimilation is sometimes found in English when two consecutive fricatives occur. Given the difficulty in pronouncing two similar fricative sounds contiguously, some speakers avoid this by articulating one of them with a different place of articulation, which makes them less similar and easier to pronounce (Denham & Lobeck, 2012). A clear example of dissimilation concerning fricative sounds is exemplified in (27) below:

(27) *months* /mʌnθs/ > [mʌnts]

In the example illustrated above, the dental fricative [θ] changes to a different place and manner of articulation becoming an alveolar plosive [t].

Dissimilation at a distance can be found in tongue twisters since speakers tend to have articulatory difficulties when many similar sounds occur together frequently, as in (28):

(28) *Peter Piper picked a peck of pickled peppers*

When articulating this tongue twister, there can be a high chance of pronouncing ‘*pecker*’ instead of ‘*pepper*’ to avoid two bilabial plosives /p/ in the same lexical item (Crowley & Bowern, 2010).

Romance languages are particularly important for research on consonant dissimilation (Posner, 1961). For instance, liquid dissimilation has long been observed in the historical development of Latin and Romance languages such as Spanish. By way of example, Latin *arbor* developed into Spanish *árbol* ‘tree’, or Latin *peregrīnus* whose first [r] sound dissimilated to [l] when incorporated into the English language as *pilgrim*. Another example is the English loan from the Latin *purpura* that was modified as *purple* in English (e.g., Bye, 2011; Crystal, 2003; McMahon, 1994). The dissimilation processes in the English words *pilgrim* and *purple* are illustrated by McMahon (1994, p. 15) as shown in (29) and (30) below:

Latin	>	Old French	> English loan
(29) <i>peregrīnus</i>	>	<i>pelerin</i>	> <i>pilgrim</i>
(30) <i>purpura</i>	>	<i>purpre</i>	> <i>purpre</i> > <i>purple</i> > <i>purple</i>

Dissimilation has also been defined as derived from synchronic processes that apply in limited morphophonological contexts, and hence they are likely to be fixed vestiges of non-dissimilated patterns (Hall, 2007). For instance, a common dissimilated pronunciation is in *library* in which the first [r] sound dissimilates so radically that the result is its disappearance in non-standard varieties, resulting in a pronunciation similar to lib(r)ary as ['laɪbəri] (Hock & Joseph, 2009). A similar example is that of *February* where its first [r] sound may disappear as well, which derives in pronouncing *Feb(r)uary* ['feɪjuəri] (Hall, 2007, p. 15).

1.1.2. Sandhi processes

After dealing with the first category of speech processes in the revised taxonomy (i.e., modification processes), the second category of speech processes described here can be referred to as sandhi processes. The function of the sandhi phenomenon is to make two

words sound like only one, as well as avoid hiatus and assimilation (Alameen & Levis, 2015; Kolachina, Sharma, Gadde, Vijay, Sangal & Bharati, 2011).

The term sandhi comes from Sanskrit meaning ‘joining’ (Andersen, 1986, p. 11), and Crystal (2003), for example, describes sandhi as the term used in linguistics to refer to the phonological modifications of grammatical forms that are juxtaposed either at word boundaries or within words. Moreover, sandhi is also defined as a morphophonological speech phenomenon that produces resyllabification since sandhi phenomena may operate at either morpheme or word boundaries (Kolachina et al., 2011; Masutti, 2014).

Two types of sandhi can be distinguished according to the position in the lexical item(s): internal and external sandhi. The former refers to morphological changes within words that operate at morpheme boundaries, whereas the latter refers to processes that operate across word boundaries.

Internal sandhi, as shown in (31), is a similar process to assimilation, as it involves morphophonological changes at morpheme boundaries (Kolachina et al., 2011). However, a process of internal sandhi normally involves phonological variability across morpheme boundaries in derivational words after applying certain morphemes, such as when adding prefixes to a lexeme.

(31) *in* (prefix) /ɪn/ + *possible* (base word) /'pɒsɪbl/ > *impossible* [ɪm'pɒsɪbl̩]

An instance of external sandhi in English would appear in the colloquial contraction forms of certain verbs as in (32) and (33) that are followed by *to* (Kolachina et al., 2011; Ternes, 1986):

(32) *want to* /wɒnt tu/ > *wanna* [wɒnə]

(33) *have to* /hæv tu/ > *hafta* [hæftə]

Examples of external sandhi can also be found in tonal languages and the historical development of word-final consonant sandhi in some Western languages. Researchers

refer to these morphotonemic alternations as tone sandhi, in the case of tonal languages (Chen, 2000, p. xi). This process refers to a change in the tone of a phonetic segment given the influence of the pronunciation of the next tone. Regarding word-final consonants in some Western languages, the process of consonant sandhi can be illustrated by the case of French liaison and /r/-sandhi in English. French liaison refers to the resyllabification of an empty syllable onset, while r-sandhi involves the pronunciation of a post-vocalic /r/ and its link with the vowel at the beginning of the following lexical item. A discussion of tonal and consonant sandhi is provided below.

Tone sandhi

In tone languages, syllables have a lexically specified tone that differentiates lexical meaning. Many Asian and African languages illustrate several types of tonal reorganization that involve the phonetic or phonological influence of adjacent tones. When two or more tones occur in a sequence, one or more tones frequently change in connected speech (Wang & Smith, 2013; Yuan & Chen, 2014).

In this respect, tone sandhi alters drastically the phonetic shape of adjacent tones in connected speech, a phenomenon that is found particularly in the majority of languages spoken in China and the surrounding areas of East and Southeast Asia (Chen, 2000, p. xi).

Standard Mandarin Chinese, for example, has four distinctive tones. The first tone (T1) is a high-level tone; the second tone (T2) is a rising tone; the third tone (T3) is a low (dipping) tone; the fourth tone (T4) is a falling tone (Duanmu, 2007). When the four tones are applied to the same syllable, four different lexical meanings may result (Wang & Smith, 2013), as shown in Figure 4 below -tones are indicated with Pinyin diacritics (ā T1 'high'; á T2 'rising'; ǎ T3 low; à T4 falling), accompanied by simplified Chinese characters:

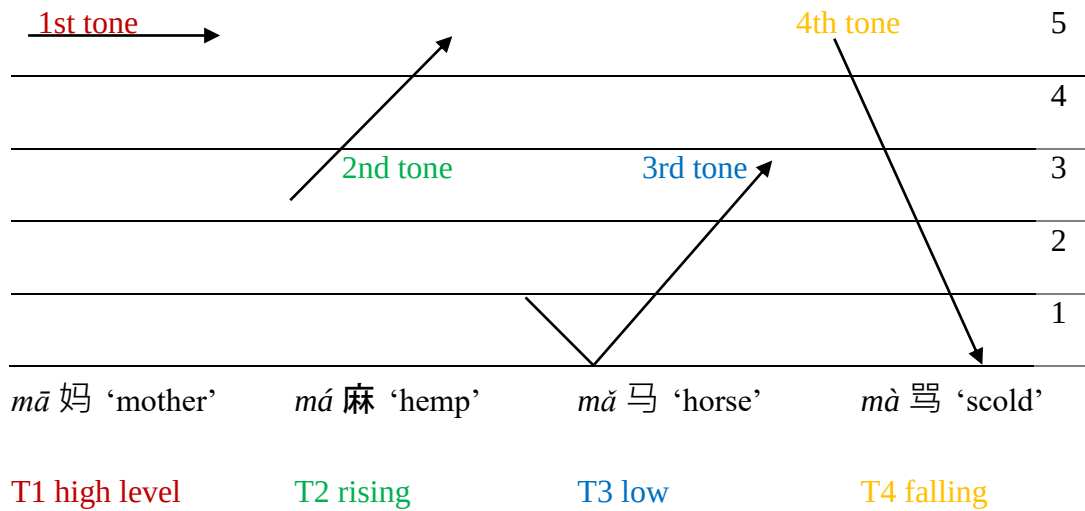


Figure 4. Tone sandhi representation with four tones applied to the same syllable. Adapted from Huang, Y., Athanasopoulou, A., & Vogel, I. (2018).

Tone sandhi also turns a third low-dipping or ‘dip-rise’ tone (i.e., T3) into a T2 tone when followed by another T3. This process, as Chen (2000) claimed, is traditionally known as third tone sandhi (Figure 5 below):

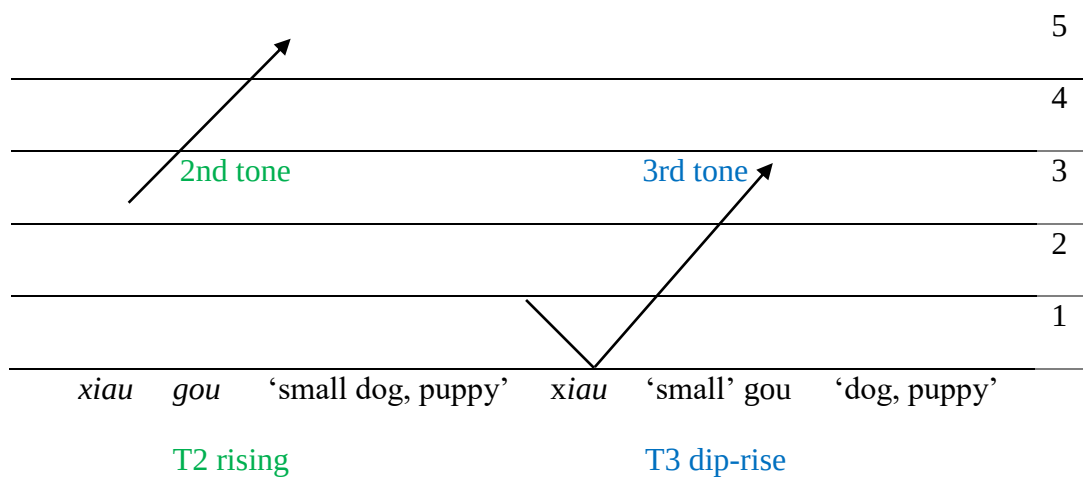
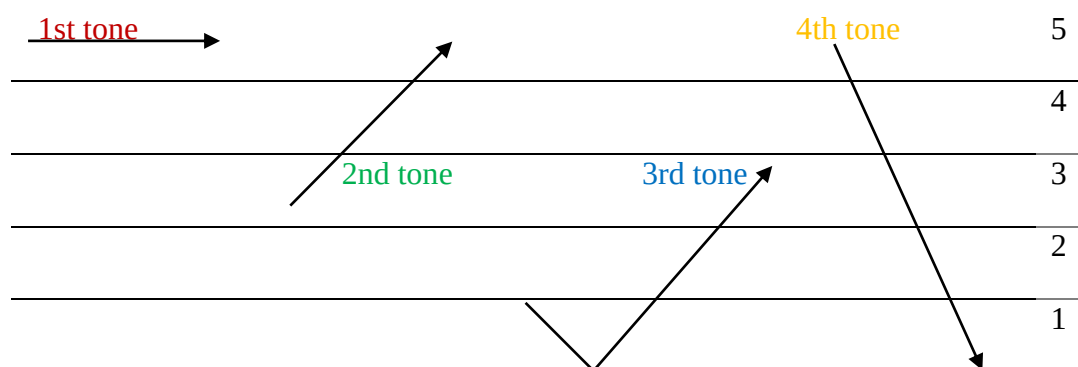


Figure 5. Third tone sandhi applied in connected speech in Mandarin Chinese.

In the examples above, the third tone sandhi rule turns the falling-rising tone in *xiao* ‘small’ into a T2 or high-rising tone as it is followed by another T3 or falling-rising tone in *gou* ‘dog, puppy’.

Another third-tone sandhi rule, often referred to as the neutral tone rule, is the traditional term for toneless syllables, and concerns neutral-toned syllables in which pitch level is determined by the preceding full tone (Yip, 2002). Toneless syllables are not meaning carrier words, they only appear in a non-initial position preceded by, at least, one syllable that carries a tone. A neutral tone is predictably high after T3, mid after T2 and T1, and low after T4. An example is illustrated by Chen (2000) using the toneless enclitic *–de* that functions as the nominalizer (see Figure 6):



xin de ‘the new one’ *hong de* ‘the red one’ *zi de* ‘the purple one’ *da de* ‘the big one’

T1 high level

T2 rising

T3 dip-rise

T4 falling

Figure 6. Third tone sandhi applied in connected speech in Mandarin Chinese.

Consonant sandhi

Word-final consonant sandhi is a common phenomenon in Western languages such as English and French, where the processes have long been described. This refers to a linking consonant between two words in an appropriate phonetic and syntactic context, the phenomenon is often referred to in these languages, particularly French, as *liaison*. Sandhi, or *liaison*, is an umbrella term that normally refers to the pronunciation used to break hiatus across word boundaries in different contexts (e.g., Côté, 2011).

French liaison

In French, for example, consonant sandhi, or *la liaison*, involves the pronunciation of commonly mute consonantal letters in word-final position, which will be uttered if the following word begins with a vowel (Côté, 2011, 2013; Masutti, 2014).

French liaison is thought to date back between the 12th and 16th century when fixed final consonants progressively ceased to be pronounced before other consonants in the first place, and subsequently ceased to be pronounced before a pause. However, some liaison consonants are thought to have evolved through an analogy process (Côté, 2011). An example of French liaison is illustrated in (34) below:

(34) *petit* /pti/ ‘small’ cf. *petit ami* [ptit_̣ami] ‘boyfriend’ (lit. ‘little friend’)

The example above shows the liaison consonant /t/ at the end of the word *petit* ‘small’, which can be represented in transcription simply by adding the consonant with or without linking the two words (e.g., [ptit ami]/[ptitami]), or using a bottom tie bar to show the link (e.g., [ptit_̣ami]). A bottom tie bar, as in the latter transcription, is used for all the examples involving cases of consonant sandhi processes in the current dissertation.

French liaison may apply categorically or variably across word boundaries as in example (35) (Côté, 2011; Masutti, 2014):

(35) *Ils ont attendu* /ilz ɔ̃(t) atãdy/ ‘they have waited’

In (35), the alveolar fricative [z] in the third plural personal pronoun is usually pronounced as a categorical liaison by linking the personal pronoun to the following vowel-initial verb. However, the word-final /t/ remains a variable liaison phenomenon, as variation in French liaison often depends on stylistic and sociolinguistic factors.

In terms of the French liaison variation, Selkirk (1972) argued that there are three different styles of speech in French where liaison applies. The first one is defined as an

informal style, in which liaison mainly involves function words, as shown in (36) below (Selkirk, 1972, p. 221):

(36) *on le considère très incommode*

/õ lə kōsiderə tʁɛ ɛ̃kɔmɔd/

>

[õ lə kōsiderə tʁɛz ɛ̃kɔmɔd]

The second type of style is related to formal conversations that are characterised by a larger number of potential contexts for liaison, including lexical items that retain their final consonant, as shown in (37) below (Selkirk, 1972, p. 229):

(37) *leurs rapports avec l'administration sont assez intimes*

/lœʁ rapɔʁ avek ladministrasjõ sõt ase ɛ̃timə/

>

[lœʁ rapɔʁz avek ladministrasjõ sõt aseɛ̃timə]

The third style is concerned with reading and discourse and displays the largest number of potential contexts since inflected nouns, verbs, and adjectives are likely to make the liaison more easily with their counterpart complements, as shown in (38) below (Selkirk, 1972, p. 234):

(38) *des langues étrangères* /de lãgə ɛtrãʒɛʁə/ > [de lãgəz ɛtrãʒɛʁə]

However, liaison is not categorical since the use of consonant liaison may be completely blocked in certain syntactic contexts, as it may also depend on the syntactic environment, as shown in (39a) in contrast to (39b) (Masutti, 2014; Meinschaefer et al., 2015) exemplified below:

(39) a. *bon* /bɔ̃/ ‘good’ cf. *un bon enfant* [œ̃ bɔ̃n ɑ̃fɑ̃] ‘a little child’

b. *mauvais* /mɔvɛ/ ‘bad’ cf. *mauvais en maths* [mɔvɛ ɑ̃ mat] ‘bad at maths’

Liaison is obligatory in the syntactic environment shown in (39a) but is blocked between the adjective and its complement in (39b) despite phonological conditions being optimal for French liaison to occur. Another example of a syntactic structure having an impact on the application of liaison in French is found in (40), where French liaison is mandatory when occurring between a numeral such as *deux* /dø/ ‘two’ and a noun (40a) but is completely ruled out between the numeral and a preposition (40b):

(40) a. *il a eu deux enfants* [il a y døz ɑ̃fɑ̃] ‘he had two children’

b. *il a eu deux en maths* [il a y dø ɑ̃mat] ‘he’s had a two in Maths’

/r/-sandhi

Apart from French liaison, another well-known process of word-final consonant sandhi in Western languages is that of */r/-sandhi* –or */r/-liaison*– in English. It involves the pronunciation of a post-vocalic */r/* in word-final position when the initial sound in the following word is a vowel. */r/-sandhi* is found in non-rhotic accents of English, where the rhotic consonant */r/* is found in syllable-onset positions but not in word-final, coda position.

In the English language, where this phenomenon occurs, accents have traditionally been described as rhotic and non-rhotic according to the distribution of the phoneme */r/*. Rhotic accents are characterised by the pronunciation of */r/* -typically spelt <r(e)>- in all phonological positions (e.g., most of the US, Canada, Scotland or Ireland) whereas in non-rhotic accents (e.g., most of England, Wales, Australia or New Zealand), */r/* is pronounced in syllable onset positions (e.g., *red* [red]), word-medially in intervocalic positions, with (e.g., *staring* [ˈsteərɪŋ]) or without (e.g., *very* /ˈveri/) an intervening

morpheme boundary but not word-finally (e.g., *car* [kɑ:]) or followed by another consonant (e.g., *fork* [fɔ:k]).

In spite of /r/ not being produced word-finally in non-rhotic accents of English, post-nuclear word-final /r/ can be pronounced if the grapheme <r(e)> is immediately followed by a vowel sound across morpheme boundaries, if there is a complete absence of any rhythmical hiatus or pause, and the preceding vowel is not a high (or close) vowel, central monophthong or centring diphthong in heterosyllabic contexts (i.e., V₁[-high] followed by V₂), as shown in (41) below (e.g., Collins & Mees, 2003; Lewis, 1975/2014; Mompeán, 2022; Mompeán & Mompeán-Guillamón, 2009; Wells, 1982; Wells & Colson, 1971):

(41) *stare at* ['steə̃r æt] cf. *stare* /'steə/

The phenomenon of /r/-sandhi comprises two cases: linking /r/, an etymologically and orthographically justified /r/; and intrusive /r/, which implies the use of an epenthetic r-sound not justified by the spelling (Minkova, 2014). Typical examples are (42) and (43) below, for linking /r/ and intrusive /r/, respectively:

(42) *for ever* [fə̃r 'evə]

here and there ['hɪər ən 'ðeə]

(43) *the idea[r] of* [ði: əɪ'diə̃r əv]

media[r] event ['mi:diə̃r 'ɪvənt]

In non-rhotic English, linking /r/ is the most familiar example of the /r/-sandhi process given its high rate of occurrence (Crystal, 2003). Intrusive /r/ is, on the other hand, a phenomenon that has arisen by analogy and that might date back to the time when <r> ceased to be pronounced in final positions (Jones, 1909/1956). Linking /r/ is, therefore, more common than intrusive /r/ since the former may have originated because of the loss of a post-vocalic /r/, while the latter did not exist and the insertion of an epenthetic /r/ has generally been attributed to an analogy process (e.g., Cruttenden, 2014; Gimson, 1962/1980; Wells, 1982).

Apart from being a well-known sandhi phenomenon, it should be born in mind that, in terms of its distribution, linking /r/ is categorical across word boundaries in word-internal position (e.g., *storing* ['stɔːrɪŋ]) but variable across word boundaries and that intrusive /r/ is also used in word-internal, intervocalic positions (e.g., *draw[r]ing* ['drɔː[r]ɪŋ]) by analogy with word-internal linking /r/ (see Chapter II).

The earliest evidence of intrusive /r/ in British English dates to the 18th century a time when the rhoticity of final constrictive /r/s was already disappearing in lower-class London speech (e.g., Heselwood, 2006; McMahon, 1994). As present-day orthography does not contain any <r>, intrusive /r/ is somehow stigmatised and perceived by some as a vulgarism (e.g., Crystal, 1984; Jones, 1909/1956; Knowles, 1987; Wells, 1982). However, even though intrusive /r/ appears to be generally censured by speech-conscious speakers (Roach, 2009), it is widespread even in broadcast speech (e.g., Brown, 1988b; Fox, 1978; Lewis, 1975/2014 1977; Pring, 1976).

1.1.3. Reduction processes

Crystal (2003) describes reduction as “a process of simplification which affects certain types of sound sequence” (p. 406). Reduction processes, such as lenition, tapping or elision, therefore, imply a change in the articulation of a sound that leads to the simplification of one or even several articulatory parameters in connected speech. For example, some of the parameters that may be affected refer to the duration of the segment, which can be shortened in articulation, the quality of the vowel, which can be reduced to a more neutral or central type of vowel, and even in some extreme cases, reduction may lead to the loss of a segment.

Lenition

Lenition (from Latin *lenis* meaning ‘weak’) and deletion could be included within this group of processes as both imply a weakening or reduction in the articulation of segments. The former refers to a type of process that alters sounds, especially consonants. It involves a weakening in their articulation, most commonly in weakened prosodic conditions, given a reduced use of articulators (Burridge & Bergs, 2016). Lenition refers to a gestural

reduction that occurs in weak prosodic positions in casual speech (Blevins & Garret, 2004), and lenition may also involve a reduction in consonant articulation stricture (Ashby & Maidment, 2005; Ashby and Przedlacka, 2010; Lass, 1984). An example of consonantal lenition is shown in (44) below:

(44) *not now* /'nɒt nau/ > ['nɒʔ nau]

In the example above, the stop /t/ has undergone a reduction or lenition process whereby the alveolar plosive has been replaced by a glottal stop allophone involving a loss of the articulatory alveolar contact gesture (Ashby & Przedlacka, 2010). As glottal consonants do not have oral articulation, they are weaker than any oral sound (Crowley & Bower, 2010).

Consonant lenition or weakening may also involve shortening processes such as degemination, voicing, approximantisation, debuccalisation and consonant loss (Blevins & Garret, 2004).

Lenition represents a shift towards deletion, and the latter phenomenon may be considered to be an extreme form of lenition, as it commonly refers to the elision of a sound or group of sounds (Lass, 1984). Deletion, or elision, is mostly motivated by articulatory economy and affects vowels, consonants, and consonant clusters (Crystal, 2003). In connected speech, deletion is optional but widespread. Function words in English, for example, often undergo deletion (as well as lenition) in connected speech (Crystal, 2003), as in (45):

(45) *boys and girls* /'bɔɪz ænd 'gɜ:lz/ > ['bɔɪz (ə)n(d) 'gɜ:lz]

In the example above, the conjunction *and* has undergone the weakening of a full vowel to a weak, centralized one (/ænd/ > [ənd]) in an unstressed position (Bates, 1995). Following this, the alveolar /d/ is elided, which is typical of /t, d/ when medial in a cluster of three consonants in present-day English in colloquial speech (Cruttenden, 2014, p.

257). The elision of /d/ may also be accompanied by the elision of schwa, which is normally elided before liquids and nasals in unstressed syllables resulting in a syllabic consonant (i.e., /ŋ/), which functions as the nucleus of the syllable.

Tapping

Tapping is one of the most common types of consonantal weakening (Lavoie, 2001). It is a process whereby a stop consonant turns into an alveolar tap (Alameen & Levis, 2015), which implies a decrease or reduction in the duration of the closure, as in (46) below:

- (46) a. *lady* /'leɪdi/ > ['leɪrɪ] *had a* /həd ə/ > [hər ə]
b. *better* /'betə/ > ['berə] *that is* /ðæt ɪz/ > [ðær ɪz]

The reduction process involved in the production of taps, from an articulatory point of view, is associated with the movement of the articulators since the duration of the articulatory contact between them is longer in the production of plosive consonants than in taps. The difference, therefore, concerns the rate of articulation (e.g., Hannisdal, 2006); together with the fact that taps tend to be considerably shorter in duration in contrast to plosives (Trusk, 1992). In the example above, a voiced alveolar plosive /d/ has turned into a voiced alveolar tap [ɾ] (46a), and a voiceless alveolar plosive /t/ has also turned into a voiced alveolar tap [ɾ] (46b), which implies not only a reduction in the duration of the stop closure but also a change in vocal fold vibration.

The examples in (46) above also show that t-tapping can occur both word-internally –with or without an intervening morpheme boundary– or externally across morpheme boundaries. Word-internally, t-tapping can be regarded as an allophonic process, while at word boundaries it can be considered to be a connected speech process.

This phenomenon is usually known as t-tapping or t/d-tapping or simply tapping among British, Australian or New Zealand linguists (e.g., Bell, 1991; Holmes, 1994; Tollfree, 1999), although American linguists normally refer to this phenomenon as t-flapping, t/d-flapping, or simply flapping (e.g., Eddington & Elzinga, 2008; Patterson &

Connine, 2001; Zue & Laferriere, 1979). Despite the different terminology associated with the phenomenon, both tap and flap are two terms generally considered to be synonyms that are used interchangeably to refer to a sound produced by a brief contact between the articulators (e.g., Fukaya & Byrd, 2005; Kahn, 1976). However, some authors differentiate them systematically with flaps implying the active articulator touching the passive one in a ballistic fashion (i.e., tangentially to the place of articulation in passing), while taps would imply a contact more similar to that of a stop, with the active articulator moving directly towards the roof of the mouth (e.g., Ladefoged, 1971; Ladefoged & Maddieson, 1996; Shockey, 2003; Trask, 1996). The term t-tapping, or simply tapping, is the one adopted in this study, regardless of any articulatory or terminological distinction between taps and flaps.

As Hannisdal (2006) argued, t-tapping is one of many reduction or lenition processes due to a tendency to aim for ease of articulation. The articulatory production of an intervocalic voiceless plosive /t/ “requires a switching off and on again the vocal cords, and it is easier to maintain the voicing throughout the alveolar articulation” (p. 114). Thus, tapping seems to be consistent, as the author suggested, with the principle of least effort whereby speakers tend to pronounce utterances with a minimum articulatory effort (see also Wells, 1982, p. 94). Furthermore, t-tapping has not only been related to economy when articulating sound, it has also been pointed out that t-tapping might be associated with the principle of lexical diffusion, given the tendency for t-tapping to affect high-frequency words first (e.g., Shockey, 2003).

In terms of its phonetic distribution, t-tapping is generally, although not categorically, favoured when followed by unstressed syllables (Eddington & Elzinga, 2008; Shockey, 2003), with the preceding syllable typically being stressed (e.g., *article* /'ɑ:tɪkəl/ > ['ɑ:rɪkəl]) or not (e.g., *ability* /ə'bɪləti/ > [ə'bɪləri]), that is, in a 'VCV context, while a stop pronunciation is the norm in V'CV contexts (see also 2.2.). The tap may also be followed by a syllabic consonant (e.g., *button* /'bʌtɪ/ > ['bʌrɪ]; *brutal* /'bru:təl/ > ['bru:rɪ]) (Eddington & Elzinga, 2008; Wells, 1982), and two consecutive taps can be found word-internally if the corresponding alveolar stops are found in unstressed syllables as in *competitive* /kəm'petətɪv/ > [kəm'perətɪv] or [kəm'perərɪv] (Harris & Kaye, 1990), that is, in a 'VCVCV context.

T-tapping is found in many varieties of English, including North American, Australian, British, or Irish accents (e.g., Shockey, 2003; Tollfree, 1999; Wells, 1982). Even though t-tapping is practically categorical in all environments in American English, in England, for example, is by no means uncommon and it appears as a realisation in certain casual styles at word boundaries and, to a certain extent, at morpheme boundaries in high-frequency words (e.g., Ashby & Przedlacka, 2010; Hannisdal, 2006; Wells, 1982). Indeed, the phenomenon of t-tapping across British accents is believed to represent the diffusion of American dialectal borrowing (Wells, 1982, p. 250).

T-tapping is a process that has been studied in the specialised literature from different perspectives (see 2.2.). For instance, some authors looked at its possible phonetic origins (Silverman, 2004; Turk, 1992), while others were interested in how the phenomenon varied according to linguistic and sociolinguistic variables, such as social class, age, gender or style (Byrd, 1994; Strassel, 1998; Zue & Laferriere, 1979), and the linguistic factors that condition its variability (e.g., Connine, 2004; Egido & Cooper, 1980; Gregory, Raymond, Bell, Fosler-Lussier & Jurafsky, 1999; Patterson & Connine, 2001). In British English accents, the relevant literature suggests that the plosive variant is the high-status realisation; while the voiced variant is more frequently associated with working-class speech and an informal style together with being limited to relatively rapid speech (e.g., Hannisdal, 2006; Mathisen, 1999; Watt & Milroy, 1999; Wells, 1982; Woods, 1991), and that t-tapping is also found to be gender-sensitive as taps are more frequently used by males than by females (e.g., Hannisdal, 2006; Holmes, 1994; Watt & Milroy, 1999; Woods, 1991).

Deletion

Deletion –or elision– is another common reduction process that involves the omission of a sound that is otherwise pronounced in its canonical form (Hannisdal, 2006). It may be considered as an extreme form of lenition (Lass, 1984) and it is produced both within words and at word junctures, as in (47) below, respectively:

- (47) a. *landlord* /'ləndlɔ:d/ > ['lænɫɔ:d] *send letters* /send 'letəz/ > [sen 'letəz]
 b. *postcard* /'pəʊstkɑ:d/ > ['pəʊskɑ:d] *next cab* /nekst 'kæb/ > [neks 'kæb]

The examples above illustrate a very common elision in English: the elision of alveolar plosives in interconsonantal position (e.g., Cruttenden, 2014, p. 257). The voiced alveolar plosive /d/ is elided within words and at word boundaries (47a) and the same applies to the voiceless alveolar plosive /t/ within words (47b). In fact, the alveolar plosive /t/ is almost systematically elided in word-final syllables in some common phrases, such as (48) below:

(48) *don't know* /'dəʊnt nəʊ/ > ['dənəʊ], represented informally as *dunno*

Other frequent elisions in English involve the consonants /h/, /v/ and unstressed vowels. The elision of /h/ frequently occurs when /h/ is at the beginning of an unstressed syllable in a function word, as shown in (49) below (Brown, 2014, p. 103):

(49) *take him* /'teɪk hɪm/ > ['teɪk ɪm] *take her* /'teɪk hə/ > ['teɪk ə]

Regarding /v/, it is often elided words or phrases containing /v/ as in the preposition *of* or the verb *have* when the latter functions as an auxiliary verb, as shown in (50a, b) below (Brown, 2014, p. 103):

(50) a. *loads of money* /'ləʊdz əv'mʌni/ > ['ləʊdz ə'mʌni]

b. *should have done* /'ʃʊd əv dʌn/ > ['ʃʊd ə dʌn]

In the examples above, the high frequency of lexical items or even lexical collocations in which elision takes place may lead to the lexicalisation of the elision (e.g., Bybee, 2002).

Unstressed vowels are also frequently reduced or elided in English (e.g., Hooper, 1976). For instance, the unstressed vowel schwa /ə/ is particularly prone to elision, specifically between voiceless consonants (Brown, 2014), as in (51a, b) below:

(51) a. *potato* /pə'teɪtə/ > [p'teɪtə]

b. *perhaps* /pə'hæps/ > [p'hæps]

An unstressed vowel schwa can also be elided when followed by a syllabic consonant (e.g., *garden* /'gɑ:dən/ > ['gɑ:dən]). An interesting aspect related to this kind of vowel elision is the phenomenon of compression (Wells, 2000). When schwa is followed by a liquid, which is in turn followed by another weak vowel, the first vowel may also be elided and the whole lexical item undergoes a reduction in the number of syllables, such as in (52):

(52) *difference* /'dɪfərəns/ > ['dɪfrəns]

The phenomenon of compression, also called syllabic compression, is a common process in high-frequency words, in fast or casual speech which involves syllable reduction. The process refers to a general vowel reduction when vowels are not stressed, or they are in prosodically weak positions. For example, a weak vowel changing to a semivowel (e.g., *influence* /'ɪnfluəns/ > ['ɪnflwəns]), a syllabic consonant changing to a non-syllabic consonant (e.g., *doubling* /'dʌblɪŋ/ > ['dʌblɪŋ]), or the phenomenon called smoothing whereby a long vowel or diphthong undergo reduction (e.g., *scientist* /'saɪəntɪst/ > ['saɪəntɪst]) (Wells 2000, p. 165-166).

1.1.4. Insertion processes

Nathan (2008) claims that sounds cannot only be deleted but can also be inserted, and there seem to be two main reasons for the insertion of segments: “preventing clusters of consonants that violate syllable structure constraints in the language, and easing transitions between segments that have multiple incompatibilities” (p. 81).

Depending on the part of the word where the insertion occurs, the insertion process is given a different Greek-based term. The insertion of a segment (or syllable) at the

beginning of a word is called prothesis, or less commonly prosthesis (from Ancient Greek meaning ‘addition’); the insertion at syllable boundaries is called epenthesis; and paragoge (from Greek meaning ‘a leading past’) adds a sound (or syllable) at the end of a word (Blevins & Garret, 2004; Campbell, 2013). Another term, prothesis, consists in the insertion of a vocalic or consonantal sound or even a syllable at the beginning of a word, without altering either the meaning or the structure (Trask, 1996). Nevertheless, this phenomenon should not be confused with the addition of a prefix since the latter does change the meaning of the word.

Prothesis is a well-known process in the evolution from Latin into modern Spanish whereby /e/ was inserted in word-initial /sC-/ clusters, as shown in (53) below (Janda & Joseph, 2003; Nathan, 2008, p. 81):

(53) Lat. *speciālis* > Sp. *especial* [espe'θjal] ‘special’

This type of insertion process is not only found historically but also affects borrowings from English into Spanish in their usual pronunciation (e.g., Sp. *stop* /estop/ ‘stop (in driving)’) but also Spanish-accented English words. For instance, Charlie Brown’s dog *Snoopy* is pronounced as [esnupi] in Spanish by inserting an epenthetic [e] (Nathan, 2008, p. 49) since, as Monroy-Casas (2001) argued, Spanish L2-English learners often insert /e/ in their speech (e.g., *stay* > [es'tei]), by analogy with Spanish phonotactic rules.

Epenthesis refers to the insertion that occurs within a word, typically across syllable boundaries. An example of epenthesis is the insertion of schwa between a liquid and another consonant. This process takes place in non-standard English when the schwa is preceded by a voiceless dental [θ] and is followed by a liquid [r, l], as shown in (54) below (Nathan, 2008, p. 81):

(54) *athlete* > /'æθli:t/ vs. *ath[ə]lete* > ['æθ[ə]li:t]

Schwa is also commonly inserted in Standard British English in words such as *inferior* [ɪnˈfɪ(ə)ɪər] or *pirate* [ˈpaɪ(ə)ɪət]. Even though this process is perceived as a normal form of pronunciation, this is not possible, in particular, for the phoneme /l/ in monomorphemic forms even though the same syllable context is given; an example is the word *pilot* [ˈpaɪlət]. In some morphological complex forms, however, schwa can optionally be inserted for some speakers of British English, as in *healing* [hɪ(ə)lɪŋ] (Krämer, 2005; Wells, 1982).

Another relatively common case of epenthesis is intrusive /r/ in British English as well as in other non-rhotic accents around the world, in which a non-etymological /r/ can be inserted both word-internally or across boundaries in word-final coda position when it functions as a hiatus breaker (Mompeán & Mompeán-Guillamón, 2009), as shown in (55) and (56) below:

(55) *drawing* /ˈdrɔːɪŋ/ > *draw[r]ing* [ˈdrɔː[r]ɪŋ]

(56) *idea is* /aɪˈdɪə ɪz/ > *idea[r] is* [aɪˈdɪə[r] ɪz]

Intrusive /r/ was discussed above as a sandhi process and, more specifically, as an example of consonant sandhi (see 1.1.2.), whereby the phonological form of lexical items is modified at word boundaries. It was discussed alongside linking /r/, as processes or strategies used to break hiatus, a sequence of juxtaposed syllables resulting in two separate vowels alongside one another (Cox et al., 2014). However, while both phenomena can be considered as consonant sandhi processes –grouped under the label /r/-sandhi– intrusive /r/ can also be considered to be an epenthetic phenomenon because, unlike linking /r/, it is not justified by the spelling. This may explain why intrusive /r/ has a lower rate of occurrence than linking /r/ (Mompeán & Gómez, 2011; Mompeán & Mompeán-Guillamón, 2009; Pavlík, 2016).

A plausible explanation for this, as Kamińska (1995) pointed out, can be due to the speaker’s language awareness of spelling, since “the pattern of its distribution is somewhat disturbed by the inhibiting factor of spelling consciousness” (p. 101). Even though potential cases of intrusive /r/ are generally disapproved of as correct because of

speakers' spelling consciousness, many RP speakers have to make a conscious effort to avoid inserting such forms not supported by the spelling (Kamińska, 1995).

A similar phenomenon to that of intrusive /r/ is intrusive /l/ in some varieties of American English (Gick, 2002). Intrusive /l/ refers to the (variable) insertion of a historically unattested liquid that functions as a hiatus breaker between a morpheme-final non-high vowel and the following vowel, it may also operate either across or within words, as in (57) below:

(57) *drawing* /'drɔ:ɪŋ/ > *draw[l]ing* ['drɔ:[l]ɪŋ]

As is the case with intrusive /r/, it may also operate either across words, as in *saw it* /sɔ:ɪt/ > *saw[l] it* [sɔ:[l]ɪt] (Gick, 2002, p. 178).

Finally, paragoge, also called final-vowel epenthesis, implies a process whereby a non-etymological sound, usually a vocalic one is inserted, at the end of a word (Campbell, 2013). The process refers to a case of gestural strengthened release, in which a consonant release is strengthened and interpreted as a vowel (Blevins & Garret, 2004). As the specialised literature claims, paragoge is a frequently observed phenomenon in creole languages, as in (58) below (Holm, 1988; Plag & Uffmann, 2000):

(58) Eng. *big* > Sranan *bigi*

Paragoge has also been found in second language acquisition as well as loanword adaptations, as in (59) and (60) (e.g., Haspelmath & Tadmor, 2009; Tarone, 1980; Uffmann, 2007b; Wilner, 2003).

(59) Eng. *sack* > L1 Korean [sækə]

(60) Eng. *class* > Yoruba [kíláàsi]

As to why this type of process occurs in these languages or during the process of English learning acquisition, it is worth noting that it might probably be due to certain phonotactic restrictions in the syllabic structure of these languages. Languages whose syllable pattern is mostly CV, in contrast to CVC. It might also be due to a scarcity of available phonotactic combinations containing a consonant cluster in the given language, as in Spanish, which has a limited number of consonant clusters followed by a vowel, such as *fax* /faks/ or *sport* /espor(t)/ ‘casual (of clothes)’ (i.e., (CC)VCC).

In the same vein, in other languages, such as Japanese, there exists a phonotactic restriction of avoiding word-final consonants whereby a word-final vowel is added at the end of a loanword containing a word-final consonant. The process of paragoge is, therefore, in connection to how the syllable structure is reorganised during the second language acquisition process, which seems to be mainly conditioned by the syllable structure constraint of the L1 (i.e., Japanese speakers of English as a foreign language). The motivation appears to be the reorganisation of second language syllables into the L1 CV syllabic pattern, as in (61) and (62). Therefore, paragoge seems to be a fundamental epenthetic process in loanwords, since, as Ross (1994) identified, it can be considered to be a typical process in second language acquisition.

(61) Eng. *freezer* > L1 Japanese [furizā]

(62) Eng. *passport* > L1 Japanese [pasapoto]

1.1.5. Phonological approaches

The different connected speech processes discussed in the previous sections can be understood as happening in the context of spoken interaction as described by the Speech Chain Model (Denes & Pinson, 1993). More specifically, as pointed out above, they can be interpreted in terms of their articulatory, acoustic, or perceptual characteristics corresponding, respectively, to the speech production, transmission, and speech perception phases of the Speech Chain.

While a relatively simple definition of connected speech processes is useful for the description of such processes in phonetic and applied linguistic studies, the numerous

linguistic theories that have proliferated during the past 120 years have long been concerned with the processes that are collectively described here as connected speech processes and treated them in various ways depending on the theoretical stand of each approach. These phonological approaches typically reflect their view of human language (and cognition in many cases) at large, which is also used in the description of other levels of language beyond the phonological one (e.g., syntax, semantics, etc.).

The introductory chapter concludes with this section by examining the theoretical development of the progress of phonological theories. Figure 7 shows a list of some of the most outstanding phonological theories that have dealt –to various degrees– with connected speech processes, phonological paradigms and how the connected speech phenomena under investigation could be understood (see Clark, Yallop & Fletcher, 2007 for a thorough description of all phonological approaches). They are broadly classified as: (a) classical approaches –those between the late 19th and first half of the 20th century; b) modern approaches –those in the second half of the 20th century; and c) current trends –those developing mostly in the 21st century. Despite the many differences between approaches, they seem to share the common –related to Bloomfield’s definition of linguistics as a science (Bloomfield, 1933)– that linguistic (including phonological) work should be descriptive (and possibly explanatory) but not prescriptive.

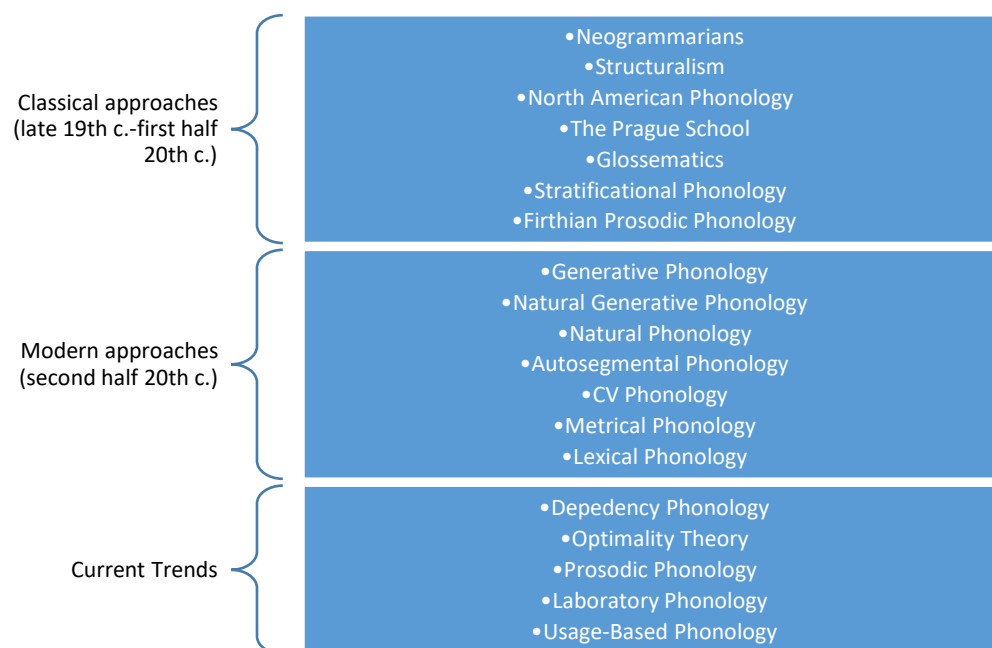


Figure 7. Outline of phonological approaches.

Subsequent paragraphs proceed through the different phonological paradigms and approaches. This section emphasises the essential nature of theoretical discussion in a scientific study of language.

The classical approaches are diverse, and their interests range from the interest in sound change of the Neogrammarians during much of the late 19th century to the focus of prosody of the Firthian school (see McMahon, 1994). The German school of Neogrammarians postulated that every sound change is regular and conditioned solely by its phonetic environment (Clark et al., 2007). However, such a view limited the description of the linguistic change to the sound level, and a widespread view during the first part of the 20th century arose.

One of the most influential classical approaches was Structuralism –or structural linguistics (see Almela & Marqués, 2016; Clark et al, 2007 for a detailed description of the structuralist paradigm), which understood language as a system whose elements and their oppositions are in relation to other elements within the system.

Structuralism was divided into two branches: its North American branch of (post-) Bloomfieldian phonology or phonemics (e.g., Bloomfield, 1933; Harris, 1951; Sapir, 1925) and its European branch, where the most influential figure was de Saussure -other later leading figures were Coseriu, Hjelmslev, founder of the Linguistic Circle of Copenhagen; Jakobson, founder of the Prague School; Trubetzkoy (1939/1969) and his functional definitions of the phoneme as an element in opposition.

Despite the two branches of this linguistic paradigm (North American and European Structuralism) developed simultaneously, they were far from being homogeneous since they followed slightly different directions. For instance, American Structuralism was less interested in studying meaning than European structuralism, while de Saussure, in contrast, made a paramount contribution to the linguistic science that enabled to differentiate the synchronic aspect of language from what belongs to the diachronic description (Almela & Marqués, 2016).

Structuralism could be defined as a systemic and structural approach to language and its fundamental idea is that language is understood as a system of interrelated elements which occur in relation to each other and that their relations are part of a whole, of a structure. By declaring that language is a system, the phonological choices are restricted and make sense by reference to the system itself only, which has long been

debated in the discussion of the “phonemic principle” or “phonological distinctiveness” (Clark et al. 2007, p. 394; see also Almela & Marqués, 2016).

The most debated dichotomies concerning structuralist linguistics have been those of *langue* and *parole*, *paradigmatic* as opposed to *syntagmatic* and *synchronic* in contrast with *diachronic* aspects of language, concepts that set the foundations of Structuralism. These concepts have been repeatedly questioned by other dominant paradigms and in modern phonology as well (Almela & Marqués, 2016). Within the structuralist view, the Prague School emphasised the difference between phonetics and phonology and gave importance to the oppositions among phonemes rather than the concept of the phoneme as a mental image. Trubetzkoy, for instance, also developed the concepts of neutralisation and archiphoneme (Clark et al., 2007).

Following subsequent paradigms and trends, another theory of language that exercised some influence on the development of phonology was Glossematics, which can be considered as a general theory of language developed by two Danish linguists: L. Hjelmslev and H. J. Uldall. Hjelmslev, for example, stated that the phoneme must not be described using a physical or psychological principle, but in terms of its function in language.

Stratificational phonology, a theory of language extensively influenced by glossematics, was developed in the USA in the 1960s. In terms of the Stratificational phonology approach, language is understood as a network of associations. That is, language is structured on different levels or strata, where phonemes are represented as points within a network. The innovative concept added by stratification phonology was the operationalisation of an inclusive network or multilevel relationships which not only included phonemic oppositions and allophonic realisations but also phonotactic rules and alternations (Clark et al., 2007).

Within the classical approaches to phonology, Firthian prosodic phonology, developed by J. R. Firth, understood the language system as reflections of paradigmatic oppositions among phonemes; therefore, a system would consist of phonemes in opposition. In contrast to Stratificational phonology, Prosodic phonology put into question the significance of the phoneme as a segmental unit (Halliday, 1978, 1985; Sampson, 1980; Clark et al., 2007).

From the late 19th century to the first half of the 20th century, diverse phonological approaches emerged that postulated different views of understanding the language system. In terms of the connected speech processes that were studied and understood during earlier traditions or trends in phonetics and phonology, Structuralism was the most influential paradigm for the most part of the 20th century, among the different phonological approaches mentioned earlier. However, structuralist phonologists did not pay much attention to connected speech processes, given their emphasis on discovering the underlying phonological ‘system’ of languages. For instance, Sapir (1925) emphasised the concept of studying sound structure taking into consideration a central mental reality, which would be part of the speaker’s cognitive organisation of the language.

Later, Bloomfield (1933) and Harris (1951) focused the study of sounds on a theory of the phoneme based completely upon observable linguistic data, from which contrastive minimal elements would be derived. The central concept of this theory was a phonemic representation based on the notion of bi-uniqueness (Anderson, 2001). As Anderson (2001, p. 11388) defines “the requirement that given either a phonetic or phonemic representation of an utterance in a given language, that could be converted uniquely into the other (disregarding free variation) without additional information”. North American structuralist linguists, such as Harris (1951), often referred to phenomena as junctures and boundaries, and even morphophonemic alternations, because their analysis would eventually refer back to their phonemic analyses, as in *the seat is loose* that could be pronounced as [ðə si:t 'ɪz lu:s] or [ðə si:r 'ɪz lu:s] (Harris, 1951, p. 39).

The beginning of the second part of the 20th century marked the decline of structural linguistics and the rise of Generativism, which was based on the idea of the hypothesised innate grammatical structures and rules. Among the so-called ‘modern approaches’, Generative Phonology stands out with a seminal work on English phonology based on a transformational-generative theory of language: *The Sound Pattern of English* (SPE; Chomsky and Halle, 1968), which represented a key landmark not only in phonological theory but also in the phonological description of the English language.

As Goldsmith (1976) argued, Chomsky and Halle’s (1968) Generative Phonology accounts for a set of rules that are organised sequentially, from which a surface representation is linked to the underlying (abstract) representation. The phonological component in the generative approach consists of a set of phonological rules that generate

In this insertion rule, a rhotic segment ([ɹ]) is not found in the underlying representation of a given lexical item (e.g., *more* /mɔ:/), but would be introduced if the preceding sound is a non-high vowel (e.g., *more of* [mɔ:ɹ ɒv]). In the deletion rule, /r/ would be considered to be in the underlying representation (e.g., *more* /mɔ:ɹ/) but would be elided if followed by a consonant (e.g., *more people* [mɔ: 'pi:pəl]) or pre-pausally (e.g., *I want more* [aɪ 'wɒnt mɔ:]).

Notice, in the examples immediately above, however, that the underlying representation of the rhotic, which can be conveniently represented as /r/, may surface as a rhotic of variable pronunciation (an approximant, a tap), but the latter may not differ significantly from the underlying specification of the rhotic.

Generative Phonology is also widely known for its degree of abstractness in phonological analysis, in many cases making underlying forms radically different from surface forms. This is one reason why, in the 1970s and 1980s a number of post-generative phonological approaches appeared. Most of them were also based on the notion of phonological rule, which bears witness to the influential role of Generative Phonology until the end of the 20th century. Some of these approaches include Natural Generative Phonology (e.g., Hooper, 1976), Natural Phonology, CV Phonology and Lexical Phonology, among others.

In Natural Generative Phonology (NGP), for example, the speakers construct phonological realizations only if the underlying forms are identical to the surface representations (Clark et al., 2007; Vennemann, 1971). Following this approach, the surface phonological representations are described by means of direct phonological alternations that can be defined in purely phonetic terms (Picard, 1979). In Natural Phonology (NP; Donegan & Stampe, 1979), a distinction is made between *(morpho)phonological rules*, which have no synchronic phonetic motivation, but are language specific, and are simply learned from other speakers in language acquisition, and phonetically motivated *phonological processes* (Trask, 1996), which are motivated by production and perception, or “are governed by forces implicit in human vocalization and perception” (Donegan & Stampe, 1979, p. 126). These phonological processes are evinced in a wide range of languages and they can be considered to be universal and innate –although the universality does not mean that they apply in all languages– only

that they are motivated in all speakers as they operate in ongoing speech production (Donegan & Stampe, 1979).

Within this framework of generative phonology, rules of phonology specify what combinations of phonological categories are well formed. The phonological structure is, therefore, realised through phonetic mechanisms, generally referred to as phonetic implementation (i.e., a set of rules whereby the surface structure is generated based on articulatory or phonetic representation). Rules of phonetic implementation, by contrast to phonological rules, relate the categorical representations of phonology to representations of a description of speech as a physical phenomenon. By doing so, phonetic implementation rules assign language-specific phonetic values to phonological features (Donegan, 2002).

In discussing phonological alternation (represented here with a colon) between pairs of items, a distinction can be made between phonetically unmotivated, (morpho)phonological rules, as shown in (64a), and phonetically motivated natural processes, as in (64b) (Donegan & Stampe 1979, p. 132-136):

(64) a. *haus* /haus/ : *häuser* /'hɔɪzə/

b. *tin* /tɪn/ : *tin can* [tɪŋ kæn]

The alternation in German between *haus* /haus/ ‘house’ and *häuser* /'hɔɪzə/ ‘houses’, known as vowel umlaut, is phonetically motivated insofar as the /au/ - /ɔɪ/ alternations has to be learned. In contrast, the alternation between /n/ and /ŋ/ as found in *tin* and *tin can*, respectively, is phonetically motivated as it is due to the immediate phonetic context –the alveolar nasal following a velar– resulting in assimilation.

Another example of a phonetically motivated process is tapping or flapping (Donegan & Stampe, 1979, p. 146), as in (65), where a voiced alveolar plosive has turned into a voiced alveolar tap, thus:

(65) *that apple* /ðæt æpl/ : [ðær æpɾ]

Tapping is an example of ‘lenition’, a type of phonological natural process representing a move towards articulatory simplicity enabling the vocal apparatus to make less articulatory effort. NP makes a distinction between ‘lenition’ processes, such as these, and ‘fortition’ processes, or those intensifying the salient features of individual segments making them more perceivable. However, not all natural processes are segmental, and together with fortitions and lenitions, NP also distinguishes ‘prosodic’ processes (Donegan & Stampe, 1979, p. 142).

Another interesting development within the ‘modern’ phonological approaches in the second half of the 20th century is the proliferation of ‘nonlinear’ phonology from the mid-1970s on. Before the 1970s, most phonological work assumed that phonological representations were strings of phonemes or bundles of distinctive features, as is the case of SPE. In reaction to the strong segmental orientation of classical Generative Phonology, a number of non-linear approaches posited a richer phonological structure with segments organised into a hierarchy of prosodic constituents, including the syllable, the stress foot, the phonological phrase, etc. (McCarthy, 2001), which allowed to address issues or problems that had either been ignored or treated unsatisfactorily in previous approaches.

These approaches included, among others, Autosegmental Phonology (e.g., Goldsmith, 1976), Metrical Phonology (e.g., Hayes, 1981) or CV Phonology (e.g., Clements & Keyser, 1983). The key point of Autosegmental phonology was that speech is segmentally organised in articulatory gestures whose temporal duration is not divided whereas CV Phonology, for example, introduced a new approach to syllable representation, expanding on the hierarchical approach developed earlier by Daniel Kahn (Kahn, 1976), which assumed a CV-tier in phonological representation constituted by C and V slots that are filled by segments (Clark et al., 2007). One of the advantages of the CV approach was that it allowed to include syllable structure information in phonological rules.

Kahn (1976) also pointed out that polysyllabic items in English need not have well-defined syllable boundaries (p. 34). Furthermore, he also concluded that ambisyllabicity is a conditioning factor for flapping to be produced in /t/ or /d/ contexts. For instance, in *hit Ann* /hɪt æn/ > [hɪr æn] for /t/ to be tapped, it must be tautosyllabic (p. 96), that is, /t/ must belong to the next syllable as well as Kahn’s Rule V (p. 53). This tautosyllabic condition of /t/ syllabically lined to the following vowel is applicable to

flapping or tapping both at word junctures and within words. Another example is /r/-sandhi, Khan (1976) pointed out that some traditional descriptions often claim that /r/ is deleted in a post-vocalic position (e.g., *car*, *card*); however, it is not deleted postvocally if followed by a vowel (e.g., *Mary*). What appears to be behind references to post-vocalic /r/ is the assumption that the rhotic which “follows a vowel and stands fully within that vowel’s syllable is subject to deletion” (p. 106). Post-vocally (e.g., *bar is*), /r/ is also lost unless it serves as a syllable-initial segment, that is, syllabically linked to a following vowel (p. 109).

CV phonology also underscores the importance of distinguishing between lexical contexts and contexts beyond the individual word. This tendency became more apparent at the end of the ‘modern’ period (second half of the 20th century) with the emergence of Lexical Phonology, developed by Kiparsky (1982) and Mohanan (1982), which was concerned with the notion of lexical and postlexical rules application. Its fundamental claim was that all morphological processes, and many phonological ones, are applied in the lexicon.

Lexical Phonology states that there are two types of rules: one set of phonological rules –lexical- that are applied in the lexicon, simultaneously with morphological operations (they require morphological information), while another set of phonological rules –postlexical- that are applied after syntactic rules (Kiparsky, 1982; Mohanan, 1982). According to this theory, morphological and phonological rules apply cyclically, that is, lexical phonological rules are reapplied prior to and after every morphological process (Kaisse & Shaw, 1985).

According to Mohanan (1986), one of the criteria to determine whether a phonological rule takes place lexically or postlexically is its application across words since “morphemes are concatenated to form words in the lexicon, but words are concatenated to form phrases in the syntax” (p. 9). Given this, connected speech processes happening at word boundaries may be considered to be postlexical rules. However, if a given process occurs both within words and across word boundaries, the rule application may be either lexical or postlexical, respectively. Thus, flapping in American English may be lexical when operating within words (e.g., *divini[r]y*) but postlexically across word boundaries (e.g., *I’ll see you [r]omorrow*) (Mohan, 1986, p. 10).

The importance of considering phonological contexts beyond the individual word as well as prosodic constituents was also underscored in modern trends from the mid-1980s due to the emergence of Prosodic Phonology (e.g., Nespor & Vogel, 1986). A phonological theory which should not be confused with Firthian Prosodic Analysis, which was an early school within the classical period and could be classified as either a late approach of the ‘modern’ period or as an early example of current approaches. Prosodic Phonology views phonological components as incorporating levels of prosodic constituency made up of different interacting subsystems governed by their own principles, such as the metrical grid, lexical phonology, autosegmental phonology and prosodic phonology (Nespor & Vogel, 1986, p. 1). Its main claim is that the prosodic constituency uses phonological and nonphonological information (Nespor & Vogel, 1986, p. 2). In addition, Prosodic Phonology identifies various domains, including the phonological word, the phonological phrase, the intonation phrase, and the phonological utterance (the widest domain span of phonological rules). These domains are hierarchically structured consisting of one or more constituents at the next level down (see Figure 8).

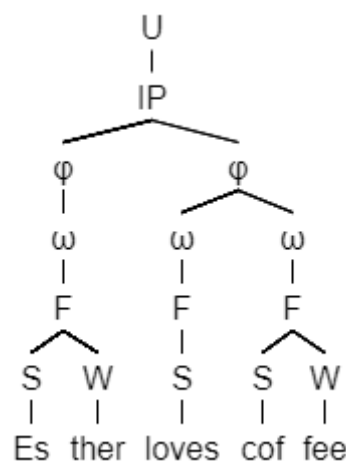


Figure 8. Hierarchical representation of the relevant domains in prosodic phonology.

The importance of these levels in accounts of connected speech processes becomes clear since the application of some of these processes may occur not simply across any word-boundary context but depending on the process, at different prosodic unit boundaries. Thus, /r/-sandhi in non-rhotic English may happen between prosodic words but it is

blocked at the phonological utterance level, which usually corresponds to the syntactic sentence. For instance, linking /r/ is applied within the word and across words, as in *a fai[r] idea*, but it is blocked at the utterance level, as in *Hi Pete*[r]!* or *There's my mothe*[r]. I've got two cats* (e.g., Gussenhoven & Jacobs, 2011, p. 252; Nespor & Vogel, 2007, p. 4).

While Prosodic Phonology may be considered to be a current approach insofar as it continues to be influential at present, other remarkable approaches developing mostly at the end of the 20th century and in the 21st century are Optimality Theory and Usage-Based Phonology. These approaches represent radical departures from previous approaches in different ways.

Optimality Theory (OT) was originally developed in the 1990s by Prince and Smolensky in a book manuscript (Prince & Smolensky, 1993), although the authors also published their revised work a decade later (Prince & Smolensky, 2004). One major difference between OT and the previous generative or post-generative phonological approaches is its reliance on the notion of constraints, not on phonological rules. OT addresses how the more practical forms of language arise from the optimal agreement among conflicting constraints. OT is mainly oriented to the output forms of language since its fundamental basis is to understand how the possible outputs are ranked (Clark et al., 2007; Cutillas, 2003).

According to McMahon (2002), OT expresses all phonological realisations in terms of constraints. While phonological rules address how underlying forms change in a particular environment, constraints set out what must or cannot happen (p. 62). A constraint is, therefore, a limit on what surface form or output the language can have.

In OT, the fundamental elements are markedness and faithfulness constraints. Markedness constraints ensure that language-specific differences are applied either to contrast or to give priority to cross-linguistically ideal forms, to the extent that they are physically and typologically justified; while faithfulness constraints ensure that surface forms are identical to the underlying forms so as to be selected as the ideal output or not. The main characteristic of the constraints is that they are both universal and that they can be violated.

Uffmann (2007) attempts to formalise a theory for optimal epenthetic consonants to explain why both glottal stops and glides are frequently used as epenthetic segments

alongside other strategies, such as intrusive /r/, across languages in distinct prosodic domains. While glottal stops are optimal, marginal epenthetic segments (i.e., inserted word-initially), glides are optimal, peak epenthetic segments (i.e., inserted in peak positions). Figure 9 exemplifies how markedness and faithfulness constraints relate to some of the connected speech phenomena mentioned above.

/lɔ: ɪz/	ONSET	*G _[-hi]	DEP(hi)	DEP	*V_V/lar	*V_V/r	*V_V/V
[lɔ:ɪz]	*!						
[lɔwɪz]			*!	*			*
[lɔwɪz]		*!		*			*
^{ɪz} [lɔrɪz]				*		*	
[lɔ:rɪz]				*	*!		

Figure 9. The ranking of faithfulness and markedness constraints for *law is* from Uffmann (2007a, p. 466).

For the input *law is* /lɔ: ɪz/, the markedness constraints ONSET (i.e., syllables must have onsets), *G_[-hi] (i.e., no non-high glides), DEP(hi) (i.e., prohibits the insertion of a feature (high) which is not in the input) indicate that the first three candidates are not optimal since they violate the constraints mentioned earlier. The faithfulness constraint DEP (i.e., prevents insertion) excludes all except the first one as an optimal candidate because of the insertion of glides, the approximant, or the glottal stop.

A crucial component of Uffmann's (2007a) r-insertion analysis refers to the OT markedness scale for intervocalic consonants which consists of a series of six ranked constraints, each of which penalizes a certain type of consonant or glide epenthesis in intervocalic position, as indicated in Uffmann (2007a, p. 463):

*V_V/lar >> *V_V/obs >> *V_V/nas >> *V_V/l >> *V_V/r >> *V_V/V

From left to right, the ranked constraints represent the following intervocalic segments: laryngeal, obstruent, nasal, [l], [r] or a glide represented as V, which are assumed to correspond to slots in the sonority hierarchy. Thus, the most harmonic segment between vowels is argued to be the most sonorous (glides) followed by the next highest slot in the hierarchy, that is, [r]. The optimal candidate is *V_V/r (i.e., [lɔ:rɪz]) because it is more harmonic than *V_V/lar and more optimal than the insertion of a glide (i.e., *V_V/V) because it violates two constraints (i.e., *G_[-hi] and DEP(hi)). Based on the sonority-based prominence scale (Prince & Smolensky, 1993), *V_V/r (i.e., [lɔ:rɪz]) is the only output that does not incur a fatal violation. As the [r] is ranked as the best possible or optimal output, Uffmann's (2007) view argues against the assumption that epenthetic processes, such as intrusive /r/, are not synchronically arbitrary.

An even greater departure from generative and post-generative analyses in the past few decades is the development of approaches called Usage-based Phonology (e.g., Bybee, 1999, 2001) or even Emergent Phonology (Archangeli & Pulleyblank, 2022). The term “usage-based” refers to Langacker's (1987) assumption that the linguistic system is built up from concrete usage events of particular language units, following Bybee (2006) who claimed that “grammar is the cognitive organisation of one's experience with language” (p. 711) (Behrens, 2009, p. 384-385). These approaches are part of a larger group of cognitive-functional linguistic theories that stand radically opposed to generative and post-generative approaches in their view of language structure and use (e.g., Geeraerts, 2010; Langacker, 1987, 1999; Tummers, Heylen & Geeraerts, 2005).

Despite differences across various approaches, their shared assumption is that language structure is not independent of the way language is used and that grammar is acquired from usage, and that the innate language-specific faculty driving the shape of adult grammar is minimal, with the development of language –including its phonology– relying instead on cognitive capacities of a general nature. Grammar and phonology emerge from the input encountered by a learner, so the study of language usage or ‘usage events’ (Langacker, 1987) becomes a priority, unlike in previous approaches given the traditional Saussurean distinctions between *langue* and *parole* (Saussure, 1916) or *competence* and *performance* (Chomsky, 1964).

One noticeable aspect of current phonological approaches is that they deal with the rich and complex patterns of intra-speaker and intra-lingual variation, generally been

ignored given the earlier focus on either the structural system (*langue*) or system of unconscious linguistic knowledge (*competence*). In a version of OT known as Stochastic OT (e.g., Boersma & Hayes, 2001), for example, the rankings of constraints for each utterance are determined probabilistically, deriving non-trivial quantitative predictions for within-speaker variation.

In phonological approaches that can be considered to be ‘usage-based’, the intrinsic variation found in language use is unproblematic. The abstract notion of unique underlying representations is unmotivated; such underlying representations would require some innate principle to ensure their postulation by a learner. According to this approach, every word, or even phonological unit, consists of a cluster of phonetic exemplars or variants that contain detailed information about the context in which they are produced (Bybee, 2001, 2013). This cognitive process is referred to as entrenchment involving repeated unit usage that leaves memory traces; consequently, the more often the same unit occurs, the more stable it becomes, which leads to routinisation and automatisisation of productions (Behrens, 2009, p. 386). As Bybee (2006) stated, a usage-based theory proposes that frequency has an impact on a speaker’s ability to recognise what is linguistically conventionalised and what is not (p. 711). However, learners store exemplars of the linguistic items in the input they receive but may develop generalisations based on those exemplars afterwards, which are constructed in a bottom-up fashion (in this way, phonology is ‘emergent’).

One key aspect in this process of concept of frequency of use (i.e., how many times a given item or pattern is encountered or used). Language structure emerges from the reinforcement of repeated features in language use (Almela & Marqués, 2016). Generally, greater frequencies lead to the abstraction and the greater ‘entrenchment’ of any specific pattern. Bybee (1999, 2001) assumes that repetition drives the automatisisation of phonological patterns due to the effect of high-frequency tokens and phonological structures are created based on speakers’ usage of language structures. Given this view, sound features of the connected speech processes reviewed above can be understood as the result of the speaker having processed and stored cognitively a given phonological unit in a particular way (e.g., *far away* as [ˈfɑːr əˈweɪ], *had a* as [həd ə], *don’t you* as [dəʊntʃu]) due to a high-frequency pattern of production.

CHAPTER II: The variables under investigation

Connected speech phenomena have long been the centre of many studies in the field of linguistics and phonetics/phonology. They range from non-empirical research that makes use of theoretical data and the analyst's impressions to empirical research that makes use of evidence-based data, including data obtained from experiments, and sociolinguistic or corpus-based studies. While each methodology has its advantages and disadvantages, a review of the impressionistic accounts, theoretical analyses as well as empirical ones that exist in relation to a given phenomenon may be crucial for a thorough understanding of the factors that affect its production (and perhaps perception).

The sections in this chapter provide a detailed description of these connected speech phenomena that constitute the focus of the present dissertation: linking /r/, t-tapping, and yod coalescence. A description of each phenomenon will be presented, together with an overview of some relevant previous research concerning phonetic, linguistic, sociolinguistic, and usage-based factors that may help to acquire a better understanding of their variation. Finally, some concluding remarks are made before moving on to the empirical studies of the dissertation in Chapter III.

2.1. Linking /r/

One of the connected speech processes discussed in the previous chapter is /r/-sandhi, which implies the pronunciation of an r-sound use of the phoneme /r/ across word or morpheme boundaries after a high vowel (typically word-finally) when the next word begins with a vowel sound, as in (67):

(67) *the door is open* [ðə dɔːr ɪz əʊpen] cf. *door* /dɔː/

In Chapter I, /r/-sandhi was classified as a consonant sandhi process (see 1.1.2.) since it refers to the phonological modifications of grammatical forms that are juxtaposed either at word boundaries or within words. It is a widespread phenomenon in non-rhotic varieties of English. It has been documented, for example, in many varieties across the

British Isles (e.g., Bauer, 1984; Broadbent, 1991; Foulkes, 1997a, b; Tollfree, 1999; Trudgill, 1974; Williams & Kerswill, 1999), in some non-rhotic areas of the United States (e.g., Halle and Idsardi, 1997; Nagy and Irwin, 2010), or the Southern Hemisphere, including South African English, Australian English and New Zealand English (e.g., Hartmann and Sabine, 2009; Hay and Sudbury, 2005; Lass, 1987; Trudgill, 1986).

It should be pointed out that the term /r/-sandhi is neutral as to the specific phonetic realisation of the rhotic consonant. The phoneme /r/ has a wide range of variation in terms of its phonetic realisation across varieties of English. In many varieties such as Received Pronunciation (RP), for example, it is currently a voiced alveolar approximant [ɹ] (Cruttenden, 2014), but an alveolar tap realisation (i.e., [ɾ]) was also common in previous decades (Fabricius, 2017). A voiced alveolar trill [r] is rarely a realisation of /r/ in English accents (see Carr 1999, p. 12).

However, the symbol /r/ is typically used in descriptions of English consonants for typographic simplicity, although some phoneticians -notably Australian- use the IPA symbol for the approximant to refer to the phoneme, i.e., /ɹ/ (e.g., Cox et al. 2014). Here and henceforth, the symbol /r/ will be used for consistency and because its phonetic realisation is not of central relevance in the present dissertation. The symbol /ɹ/ or any other notation will nevertheless be preserved in the authors' quotations.

As mentioned previously (see 1.1.2.), /r/-sandhi comprises two different phenomena in non-rhotic accents of English. The pronunciation of an etymologically and orthographically justified pronunciation in non-rhotic accents of English is known as linking /r/, as in *the door is open* in (67) above. It reflects a historical /r/ that disappeared from the pronunciation at some point in the history of the language (Fabricius, 2017), as in *door* in (67) above. This dissertation focuses, alongside the phenomena of t-tapping, and yod coalescence, on this type of /r/-sandhi.

The second phenomenon that can be described as /r/-sandhi, although out of the scope of the present study, is that of intrusive /r/, and it typically coexists with linking /r/ in the accents where the latter is found. It is characterised by the insertion of a non-etymological epenthetic /r/ that is not justified by the spelling either, as in (68) below:

(68) *law*[r] and *order* [lɔ:ɹən 'ɔ:də]

Despite etymological and orthographical differences, intrusive /r/ is claimed to have emerged by analogy with linking /r/ non-rhotic accents of English, since the phonetic contexts in which both types of /r/-sandhi appear are the same (Sóskuthy, 2013). Even though linking /r/ and intrusive /r/ have been typically modelled as the same synchronic process, some diachronic accounts (see 2.1.2.) argue that linking /r/ alternations existed before an epenthetic /r/ started to be used in sandhi contexts (Vennemann, 1972). For a more detailed account of the diachronic loss of /r/ and the emergence of /r/-sandhi processes see McMahon (2002) and Wells (1982).

As present-day orthography does not contain any <r> because, historically, it was non-existent, the pronunciation of the unetymological intrusive /r/ was stigmatised and perceived as a vulgarism (Crystal, 1984; Jones, 1909/1956; Wells, 1982), although its use seems to have been common even among radio and TV broadcasters (e.g., Brown, 1988b; Fox, 1978; Lewis, 1975/2014, 1977; Pring, 1976).

The accounts on /r/-sandhi are numerous in the specialised literature and are mainly: a) phonetic descriptions based on impressionistic approaches, b) theoretical accounts of the phenomenon within different phonological frameworks, and c) empirical or data-driven approaches. Given that accounts on /r/-sandhi are numerous but the approach is different, from mere impressionistic perspectives and theoretical frameworks to empirical research based on actual language data, the following sections provide a review of some of these studies.

2.1.1. Impressionistic accounts

Impressionistic accounts of linking /r/ are mainly of two types: phonetician's descriptions of their own speech or descriptions of the perceived usage of a given language group. The latter are commonly based on a supposedly homogenous speech community (Wells, 1962). Most of these accounts were developed within the British school of phonetics that flourished in the 20th century, which preceded more recent theoretical and empirical works (cf. 2.1.2. and 2.1.3.).

While impressionistic accounts may be limited in their scope, one relevant aspect thereof is that many authors have not only identified the process and described it, but they have also tried to identify or point to some of the factors that might condition its use.

According to Jones (1909/1956), for example, one factor is the presence or absence of a syntactic boundary between the words flanking the r-link site, as in (69), where *door* is at the end of the clause and *and* at the beginning of the following one, and therefore, are not part of the same syntactic constituent. The syntactic boundary may typically correspond with a prosodic boundary, signalled by a vertical line in the example.

(69) *he opened the door and walked in* /hi: 'əʊpənd ðə dɔ: | ənd 'wɔ:kt ɪn/

One problem with impressionistic accounts is that, since they are not based on extensive empirical data, they often offer different views on the relevance of a given factor from that found in other impressionistic (or theoretical/empirical) studies. According to McCarthy (1993), for example, /r/-sandhi production was not particularly sensitive to the syntactic constituency; however, Foulkes (1997) suggested that syntactic boundaries were likely to affect /r/-sandhi production because clause boundaries marked the division of two separate syntactic units, as in (69) above, preventing the production of potential /r/-sandhi contexts.

Jones (1909/1956) identified a second factor potentially having an impact on the use of linking /r/: the presence or absence of an orthographic <r>, which stands for the phoneme /r/, in the syllable likely to make the link, as in (70), where *roar* contains a potential linking /r/ syllable that begins with /r/.

(70) *a roar of laughter* /ə 'rɔ: əv 'lɑ:ftə/

The reason for avoiding linking /r/ when the syllable likely to make the link began with /r/, as in (70), could be motivated by the lack of preference for identical sounds in the same phonological environment (Mompeán & Mompeán-Guillamón, 2009). This is a phonetic factor and this type of factor, alongside syntactic ones, has also been discussed in impressionistic studies, as conditioning the variability in the production of /r/-sandhi, including linking /r/, although the examples provided are often cases of intrusive /r/. For instance, the insertion of an epenthetic /r/ seemed to be more frequent after /ə/, as in

drama[r] and *music* /'dra:mər ən 'mju:zɪk/, but less frequent after /ɑ:, ɔ:/, as in *law*[r] and *order* /lɔ:ɹ/ ən 'ɔ:də/ (e.g., Brown, 1988a; Cruttenden, 2014; Jones, 1918/1972; Wells, 1982).

Some impressionistic accounts have also suggested that the production of /r/-sandhi could also be affected by sociolinguistic factors, such as the formality of the communicative situation and the speaker's awareness of language use.

For instance, Lewis (1975/2014) claimed that “the r-link is generally made in non-hesitant, not particularly deliberate styles of delivery” (p. 39), and Brown (1988a) pointed out that /r/-sandhi was a feature of fluent, colloquial communicative styles or language use and “not so common in careful declarative style” (p. 145). However, Cruttenden (2014), Ramsaran (1978) and Wells (1982) observed that linking /r/ was common in all styles.

The absence of linking /r/ could also be due, according to some impressionistic studies, to the tendency among some speakers to avoid using intrusive /r/ in their speech for considering it to be vulgar, only “at the expense of eliminating linking /r/s too” (Wells & Colson, 1971, p. 75). In the cases where an intrusive /r/ is potentially rare, RP speakers might tend to make conscious efforts to avoid stigmatised forms (Cruttenden, 2014; Gimson, 1962/1980), by using other linguistic strategies such as laryngeal strategies (e.g., use of a glottal stop or creaky voice), or zero link (i.e., hiatus) (Wells, 1982; Wells & Colson, 1971). Linking /r/ would then be somewhat affected by this tendency, which could explain why its use does not seem to be categorical (see the review of empirical studies below). The use of glottalisation could be considered as an alternative to a linking /r/, although according to Brown (1988a), something of a misnomer since “its function is surely to keep the two syllables separate rather than to link them” (p. 145).

Finally, it should be pointed out that some impressionistic studies have not only described linking /r/ (or /r/-sandhi more generally) for an audience of linguists but from the point of view of the teaching and learning of English as a foreign language (EFL). Brown (1988a), for example, argued that linking /r/ and intrusive /r/ were taken into consideration as an important characteristic of English pronunciation, even though the intrusion of /r/ was quite common in RP. However, they should be considered taking into account the broader phenomenon of rhoticity and whether English accents and pronunciations models are rhotic or non-rhotic.

Since the adoption of a rhotic model of pronunciation may convey some advantages to EFL learners, Brown (1988a) concluded that linking /r/ and intrusive /r/ were not of great importance in English pronunciation teaching. Other authors have not questioned the use of a non-rhotic accent such as RP as the pronunciation model but views have differed as to whether linking /r/ and, especially, intrusive /r/ should be explicitly taught. O'Connor (1967) and Pring (1976) suggested that the intrusive /r/ pronunciation could be overlooked. Lewis (1975/2014) recommended teaching both linking /r/ and intrusive /r/ as long as the prosodic environment was appropriate to sound as natural as a native speaker.

2.1.2. Theoretical analyses

Theoretical analyses of linking /r/ (and to a lesser extent intrusive /r/) have tried to interpret the phenomenon within a given theoretical framework. Within the heterogeneous work that has addressed these phenomena, we can make a broad distinction between functional (e.g., function, distribution, etc.) and formal considerations.

From a functional point of view, both linking /r/ and intrusive /r/ have been claimed to perform the same linguistic function namely, breaking hiatus (Knowles, 1987; Hannisdal, 2006; Mompeán & Mompeán-Guillamón, 2009). However, as previously mentioned, other alternative hiatus breakers are a linking glottal stop and zero link (Wells & Colson, 1971). Hiatus-breaking strategies are considered to be an intrinsic characteristic of many of the world's languages given that "languages tend to favour constructions built from syllables that reinforce the alternation between consonants and vowels" (Bell & Hooper, 1978; Cox et al., 2014, p. 156).

/r/-sandhi would then be one of the strategies used in English to avoid hiatus, alongside others such as the use of glides (or semivowels), other liquids, and even glottalisation (e.g., glottal stop) (see Broadbent, 1991; Uffmann, 2007 for discussions). Moreover, given that both linking /r/ and intrusive /r/ have the same distribution patterns, they are often considered to be the same synchronic phenomenon, despite historical, orthographic and prestige differences (Heselwood, 2006; Mompeán & Mompeán, 2007; Mompeán & Mompeán-Guillamón, 2009; Wells & Colson, 1971).

A great deal of theoretical work on linking /r/ has been carried out within formal approaches to phonology, most notably generative and post-generative phonology. In this tradition, the most common theoretical approach on /r/-sandhi proposes that the phenomenon is the result of a deletion rule. This deletion rule assumes that underlying coda /r/s are present for all words that show /r/-/Ø/ alternations, and speakers only produce /r/ if it occurs in prevocalic environments; otherwise, /r/ is deleted before a consonant or a pause (e.g., Donegan, 1993; Gick, 1999; Giegerich, 1999; Harris, 1994; Mohanan, 1986). This approach, therefore, assumes that coda /r/s are present in speakers' mental representations word-finally but absent in the surface form if they are followed by a consonant or a pause.

An alternative approach explains the phenomenon of /r/-sandhi as a rule insertion. This rule assumes that no underlying coda /r/s are present, and that /r/-/Ø/ alternations occur because of the insertion of an r-sound after non-high vowels when followed by a vowel-initial word (e.g., Johansson, 1973; see 1.5.).

While explanations of /r/-sandhi as a deletion rule or an insertion rule seem incompatible, a third approach within the generative tradition has suggested that rule inversion coexists with rule deletion in the synchronic stage of the language, so elements of both rules are combined (Anttila & Cho, 1998; Kijak, 2009; McCarthy, 1991, 1993).

In general terms, these rule-based approaches, which are derived from generative phonology, assume that the deletion or insertion of /r/ occurs when the surface form is generated in the production of utterances and that these processes are not related to any historical process in the English language. Moreover, these approaches typically exclude the functions and usage of language. The same can be said of other recent accounts of /r/-sandhi in the framework of OT (e.g., Uffmann, 2007a).

Thus, any plausible inter-speaker or intra-speaker variation in the use of /r/-sandhi was excluded and not accounted for (Mompeán, 2022). Despite this, theoretical accounts outside the generative and post-generative traditions have appeared in the past few decades. These approaches focus specifically on language variation and use and do not assume that surface forms are derived or generated from underlying forms.

This more functional, usage-based phonology (Bybee, 1999, 2001, 2006) claims that /r/-sandhi is a connected speech process that can be observed in actual language use,

but that phonological representation of specific instances of linking /r/ or intrusive /r/ may be more or less lexicalised or ‘entrenched’ depending on factors such as frequency of use (e.g., Hay & Sudbury, 2005; Mompeán, 2022, Mompeán & Mompeán-Guillamón, 2009; Pavlík, 2016). In this view, “speakers store contextually different forms of the same lexical item rather than deriving them by rule from a single underlying form” (Heselwood, 2006, p. 93) and “the more frequent a collocation including a potential /r/-sandhi is, the more likely it is that /r/ will become lexicalised” (Mompeán, 2022, p. 27).

The production of /r/-sandhi could be understood as motivated by various factors (i.e., hiatus-breaking) but subject to lexicalisation, with high-frequency word collocations resulting in a higher degree of lexicalisation and low-frequency collocations in less lexicalisation. For instance, the use of /r/-sandhi might be lexicalised or restricted to very common collocations such as *for instance*, *after all* or *the idea of* (Jones, 1909/1956; Langacker, 1987; Mompeán & Mompeán-Guillamón, 2009). Variation in phonological tokens may result in a speaker having possibly stored a wide range of variation concerning a particular phonological unit with /r/-/Ø/ alternation.

2.1.3. Empirical studies

In order to fully comprehend how /r/-sandhi production is motivated, data-driven approaches, which are based on actual language data, have studied empirically /r/-sandhi in many varieties of English. The present section provides a review of some of these studies.

Empirical (or data-driven) approaches to the study of /r/-sandhi have become more frequent over the years. These help revise and update the theoretical approaches, by giving information about actual language usage and the factors that shape /r/-sandhi patterns of use. In fact, the upshot of various studies, which often use different methods, populations, and analyses, is that linking /r/ (or /r/-sandhi more generally) is by no means categorical in most (if not all) varieties of English, and that several factors or variables can have an effect on variation. The next section discusses them in more detail. In what follows, some of the empirical studies on /r/-sandhi in English accents are presented.

The first point to notice is that /r/-sandhi has been studied empirically in a wide range of varieties of the British Isles, including several studies on RP (e.g., Bauer, 1984;

Hannisdal, 2006; Mompeán & Gómez, 2011; Pavlík, 2016), as well as regional varieties, which include, among others, Newcastle and Derby (Foulkes, 1997a, b), South East London (Tollfree, 1999); or Tyneside (Watt & Milroy, 1999). Outside the British Isles, /r/-sandhi has been studied in the city of New York (Labov, 1972, 2006); AusEng (Cox et al., 2014) and NZE (Hay & Sudbury, 2005; Hay & Warren, 2002), South African English (Hartmann & Sabine, 2009).

The British variety that has received the greatest deal of attention from an empirical perspective is RP. In the first noteworthy study of this kind, Bauer (1984) conducted an auditory analysis of 37 RP speakers (18 male and 19 female) and recordings made from 1949 to 1966 of a written passage. The study included 276 occurrences of linking /r/, out of which 60.5% corresponded to actual realisations of linking /r/. The analysis indicated that although most occurrences corresponded to an approximant, some of the cases were produced as taps, which contradicted the claim by Jones (1909/1956) that linking /r/ realisations were usually produced with the tap variety. This investigation revealed no correlation found between social factors, such as age or gender, or intra-speaker factors, such as speech-tempo, in the use of /r/-sandhi. However, linking /r/ was found to occur frequently before an indefinite article or stressed pronoun.

A second relevant study was conducted by Hannisdal (2006), who studied six television newsreaders' speech and collected a large number of potential cases (n= 3,612). The data consisted of 30 speakers collected from three TV news channels: *BBC World*, *Sky News* and *ITV News Channel*. Hannisdal's study produced numerous findings. For example, she found that linking /r/, with a total percentage score of 59.8%, was much more frequent than intrusive /r/ although their distributional patterns were similar. Moreover, she indicated that /r/-sandhi was a complex variable phenomenon that correlated with several factors, mostly semantic, prosodic and articulatory factors. She also found men's tendency to use /r/-sandhi more frequently than women, although there was no significant gender pattern. She also indicated that /r/-sandhi displayed significant variation concerning the speed of delivery (cf. Lewis, 1975/2014), which may be directly in connection with the level of speech attention in newscasters. The relatively low ratio of linking /r/ production may indicate that the formality of speech context may affect variation in broadcast RP.

Another series of empirical studies was conducted by Mompeán and colleagues in the late 2000s/early 2010s on the speech of RP newsreaders. Mompeán and Mompeán-

Guillamón (2007), for example, investigated phonetic factors that might influence /r/-sandhi usage by conducting a study of 152 different RP newsreaders in 2004 and 2005. The findings indicated that the presence of /r/ in the syllable that would make the link did not seem to have a great impact on the production of /r/-sandhi. Their results also revealed that the percentage of production of linking /r/ represented less than two-thirds (58%) of all potential cases (n= 984) that might be explained by a careful, speech-conscious context of the news. In a follow-up study to Mompeán and Mompeán-Guillamón (2009) looked at structural, sociolinguistic and usage-based factors that might condition /r/-sandhi usage in a similar corpus. The results revealed a statistically significant difference in the rate of incidence in both linking /r/ and intrusive /r/. The findings also indicated no significant differences between males and females and that linking /r/ was more frequently produced in high-frequency collocations than in free word combinations. In a final study on the phenomenon of /r/-sandhi in this series, Mompeán and Gómez (2011) analysed the speech of 162 RP newscasters in a large-scale sample from the years 1999 to 2009. Their results revealed that linking /r/ was used 60.5% of the time in the whole corpus (n= 1,476). One specific focus of this study was on /r/-sandhi as a hiatus-breaking strategy. Given this, found that in contexts in which /r/-sandhi was not used (i.e., 39.5% of the corpus cases), a laryngeal strategy (i.e., the use of a glottal stop or creaky voice) was the most common way to prevent hiatus (i.e., 78.8% of the cases).

In another relatively recent study on /r/-sandhi on RP, Pavlík (2016) analysed the potential effect of phonetic grammatical and usage-based (word frequency and collocability) factors on /r/-sandhi usage. The study consisted in analysing the speech of 22 newscasters presenting short top-of-the-hour news bulletins, from whom 33 hours of speech material were recorded and 647 potential cases of linking /r/. The results indicated that word combinational frequency, following vowel phoneme, vowel type frequency, as well as stress pattern, influenced the production of /r/-sandhi.

Moving beyond the study of RP but still within the British Isles, Foulkes (1997a, b) analysed natural conversational dyadic exchanges between speakers (n= 32) from Newcastle upon Tyne and Derby. In these studies, Foulkes noted that older middle-class speakers from Newcastle, both males and females, used linking /r/ more regularly than intrusive /r/ in their speech, which was more common in the lower socioeconomic group's speech; although the social group variable seemed to play no role in Derby. The study on Newcastle English revealed that middle-class speakers, both males and females, used

linking /r/ more regularly than intrusive /r/, although in both cities, linking /r/ occurred more frequently than intrusive /r/ (Foulkes, 1977a, b). In addition to this, linking /r/ in Derby seemed to be disfavoured when the second hiatus vowel was stressed or when the hiatus coincided with a prosodic boundary (Foulkes, 1997b).

Tollfree (1999) conducted a study using Labovian interviews covering four years (1990-1994) and 90 speakers from the South East London region involving working-class speakers as well as middle-class speakers. Her study covered two groups of accents: South East London English (SELE) and regionalised forms of RP English or South East London Regional Standard (SELRs). The number of informants was 32 SELE speakers and 58 SELRS speakers. Tollfree (1999) concluded that the production of linking /r/ was virtually categorical following /ɔ:/, a:, ə, ə:/, although there seemed to be no significant sociolinguistic patterns associated with the use of /r/-sandhi.

The last empirical study described here on /r/-sandhi in British English is by Watt and Milroy (1999). These authors conducted a variationist analysis in Tyneside (Newcastle) describing data collected from a 26-hour corpus of conversational exchanges. The study involved 32 informants from working-class and middle-class backgrounds. Their study confirmed that both linking /r/ and intrusive /r/ were infrequent in Tyneside English if compared to other non-rhotic accents of English. The findings revealed that 80% of the instances in which linking /r/ was produced were associated with older speakers while linking /r/ was infrequent in younger speakers' speech with only 40% of the cases. No greater differences among social classes were found.

Empirical studies on /r/-sandhi in varieties of English outside the British Isles have been mostly conducted on the speech of Australian (AusE) and New Zealand (NZE) English. Cox and colleagues on AusE (Cox et al., 2014), for example, carried out acoustic and auditory analyses for an AusE corpus comprising 32 phrases produced in a sentence-reading task by 103 speakers. Their study examined the likelihood of inserting linking /r/ as a hiatus breaker across word boundaries. The results revealed that linguistic variables such as the prosodic context and speaking rates took prominence over speaker variables, such as age, gender, or sociolect. Their study concluded that linking /r/ was more frequently produced if the left-hand vowel was a non-high vowel and the following onset consisted of a weak or unstressed syllable.

Another series of empirical studies was conducted by Hay and colleagues in the early 2000s and 2010s on the speech of NZE speakers. Hay and Warren (2002) elicited potential cases of linking /r/ via three experiments with 16 speakers (The first experiment involved reading sentences containing potential environments for intrusive /r/, the second experiment involved the examination of the degree of intrusive /r/ in nonsense words, and the third experiment aimed to test the morphological productivity of suffixes to produce intrusive /r/). Their findings revealed that phonetic (i.e., preceding phoneme) and linguistic factors (i.e., type of affix) condition the use of intrusive /r/. In a follow-up study, Hay and Sudbury (2005) analysed /r/-sandhi in a corpus containing data from the 1940s and the years 1989 and 1995 of native NZE speakers born between 1860 and 1925. Their results revealed that back vowels favoured the use of linking /r/; however, they disfavoured the use of intrusive /r/. The findings obtained provide empirical data on the diachronic relationship between the emergence of /r/-sandhi and the decline of rhoticity in NZE. Their analysis revealed that intrusive /r/ seemed to have arisen even before rhoticity was completely gone. Their data also indicated that there was a major correlation between rhoticity and the speaker's sex, to some extent, as predictors for the production of linking /r/. No effect of the presence of another /r/ in the vicinity of the potential /r/-sandhi site was found. No influence of syllable length or word class was found either. High-frequency words after the potential /r/-sandhi site prevented the use of linking /r/. In a final study on the phenomenon of /r/-sandhi on NZE in this series, Hay and MacLagan (2012), found that lexical frequency had an impact on the production of /r/-sandhi, more precisely, high-frequency preceding words favoured /r/-sandhi in the speech of 27 NZE speakers born between 1900-1935 (*ONZE* project). The study reports data on the frequency of /r/ and the degree of constriction in terms of the acoustic properties of the alveolar approximant and frequency of use. The findings reveal that the more often a speaker produced linking /r/, the lower the F3 appeared to be, and that words that normally occur before vowels are more likely to trigger linking /r/ productions.

2.2. T-tapping

Another connected speech process discussed in Chapter I is t-tapping, a process whereby an alveolar plosive turns into an alveolar tap (see 1.2.3), as in (71):

(71) *I had a nice time* /aɪ hæd ə naɪs taɪm/ > [aɪ hæɾ ə naɪs taɪm]

In the previous chapter, t-tapping was classified as a lenition process, as it implies the weakening of the articulation or reduced use of the articulators (Burridge & Bergs, 2016) in the production of alveolar plosives, typically in intervocalic positions or when the following segment is a syllabic /l/ (Wells, 1982, p. 248), which become a tap (or flap) [ɾ]. This may happen across word boundaries, as in (71) above or word-internally (e.g. *putting* /'pʊtɪŋ/ > ['pʊɾɪŋ]), when it may be considered to be an allophonic variant.

T-tapping is a well-known feature of American English accents (e.g., Wells, 1982). In connection with this, the phenomenon is often referred to by American linguists as t-flapping, t/d-flapping, or simply flapping (e.g., Kahn, 1976; Turk, 1992). The term t-tapping, or tapping, is the one adopted in this study, regardless of any articulatory or terminological distinction between taps and flaps. It should also be pointed out that, despite its use of the letter 't' in its name, t-tapping applies not only to the voiceless alveolar stop [t] but also to the voiced stop [d] when replaced by a voiced alveolar tap [ɾ].

While t-tapping could be considered a well-established feature of connected speech in American English, it has not attracted the same deal of scholarly attention in varieties of English across the UK. However, t-tapping could be found in certain casual speech contexts in British English accents, from RP (e.g., Hannisdal, 2006) to Cockney (Wells, 1982, p. 250), and even in south hemisphere varieties, such as AuE and South African English (e.g., Hovarth, 2008; Trudgill & Hannah, 2008). The present dissertation contributes to the evidence on t-tapping in non-rhotic varieties of English such as RP.

As in the case of linking /r/, work on t-tapping is also of different sorts, ranging from impressionistic descriptions to theoretical accounts and data-driven analyses. In addition, this work is more or less substantial depending on the language variety being studied. The sections below provide a review of some of these studies.

2.2.1. Impressionistic accounts

Impressionistic accounts of t-tapping involved general approaches to its description and phonetic distribution. In the early 20th century, Haugen (1938) pointed out the pronunciation of an intervocalic allophone of /t/ in American English, which the author compared to the phonetic outcome of this allophone to British approximants (as cited in de Jong, 2011, p. 2711). Jespersen's (1928) earlier accounts involving British dialects suggested the substitution of a tap for /t/. Oswald (1943) classified three possible phonetic realisations or outcomes of t-tapping: a voiced stop and a tap, or a "flip", which was probably perceived as a coronal approximant (as cited in de Jong, 2011, p. 2711-2712).

While t-tapping is currently clearly identified in descriptive or impressionistic accounts of the phenomenon, it has not always been like this. One reason for this is that the voiceless coronal stop [t] is sensitive to undergo different processes given the effects of lenition, such as voicing [d], tapping [ɾ], and even glottalisation [ʔ] (Amos, 2007; Carr, 1991). As de Jong (2011) highlights, earlier studies failed to fully distinguish cases that recent research has described as t-glottalling in word-final position (e.g., *get here* /get hɪə/ > [geʔ hɪə]), which was confused with a lack of aspiration in connected speech segments. As Carr (1991) points out when describing the weakening of intervocalic /t/, one possible explanation for this is that "[i]ts metrical conditions are identical to those of glottalisation inasmuch as it too operates foot-internally, where the target /t/ is in the coda of the head syllable" (p. 46).

From the syllable point of view, earlier phonological accounts tried to identify more efficiently the precise prosodic environment, which resulted in a threefold operationalisation: taps are ambisyllabic (e.g., Gussenhoven, 1986; Kahn, 1976), taps are syllable final (e.g., Selkirk, 1982), and taps are non-foot-initial (e.g., Kiparsky, 1979) (see Turk, 1992 for a detailed account of these perspectives).

As for the phonetic context, the conditions for t-tapping to occur are diverse in the specialised literature although most descriptions of t-tapping point out two main features of the phonetic context where t-tapping occurs (de Jong, 1998). Firstly, the stop typically occurs in intervocalic position. Secondly, the stop must occur between a stressed and an unstressed vowel. Traditionally, t-tapping occurs in intervocalic contexts in which /t/ or /d/ are preceded by a stressed or unstressed vowel and are followed by an unstressed

vowel (e.g., *atom* /'ætəm/ > ['ærəm], *vanity* /'vænəti/ > ['vænəri]) (Shockey, 2003, p. 29; Riehl, 2003, p. 272), although the segment that precedes the tapped /t, d/ does not exclusively need to be a vowel, since Kahn (1980) also includes as possible segments such as glides and /r/ (e.g., *party* /'pɑ:ti/ > ['pɑ:ri]), and also syllabic sonorants /l, m, n, r/ that can follow the tap (e.g., *bottle* /'bɒtl/ > ['bɒrl]; *bottom* /'bɒtm/ > ['bɒrm]; *important* /ɪm'pɒtənt/ > [ɪm'pɒrənt]).

Regarding the stress patterns that govern the production of t-tapping, it has conventionally pointed out that the following vowel must be unstressed, while the presence or absence of stress on the preceding vowel is irrelevant (see Kahn, 1976; Riehl, 2003); however, as it has been pointed out that as taps involve a reduction process involving /t/ in intervocalic position, it needed not to be an adjacent stressed syllable, as in (72a, b) (Amos, 2007; Boberg, 2015; Shockey, 2003, p. 30; Ladefoged, 2001):

(72) a. *phonetics* /fə'netiks/ > [fə'neriks]

b. *property* /'prɒpəti/ > ['prɒpəri]

The stress of the syllable that precedes the tap segment seems, to some extent, to play no role, since taps can occur in both environments, following a stressed vowel (e.g., *utter* /'ʌtə/ > ['ʌrə]) or following an unstressed vowel (e.g., *obesity* /əʊ'bi:səti/ > [əʊ'bi:səri]) (Riehl, 2003, p. 272) within words. In these examples, t-tapping occurs word-internally but the same prosodic context (i.e., a stressed vowel followed by an unstressed or stressed vowel) may also allow the production of t-tapping across word boundaries, as in *get out of here* /'get aʊt əv hɪə/ > ['geɹ aʊt əv hɪə] (Shockey, 2003, p. 30).

2.2.2. Theoretical analyses

As is the case of linking /r/ (see 2.1.2.), theoretical analyses of t-tapping also tried to interpret the phenomenon based on certain theoretical assumptions or within specific frameworks. During the structuralist period, for example, Trager and Smith (1951)

considered that tapping was, for the majority of (American English) speakers, in free alternation with /t/. Moreover, for linguists from the structuralist tradition working with the notion of contrastive oppositions and contrastive phonemes, a weakened articulation of /t, d/ in certain contexts such as those of t-tapping means a neutralisation of phonemes, as in (73) (Boberg, 2015).

(73) *metal* and *medal* /'metl/ and /'medl/ > ['merl]

A number of analyses are also found within the framework of generative phonology. One of the earliest accounts in this framework is that of Selkirk (1972). She discusses the phonological context and provides the following notation for the rule (p. 197; the notation has been slightly adapted):

$$t, d \rightarrow r / V (/l, r/) _ [V]_{[-stress]}$$

Notice that Selkirk (1972) described an environment for t-tapping in American English where consonantal sounds, such as /l, n, r/ and even glides, are permissible to the left-hand environment of the tap as long as the following syllable is unstressed. Moreover, Selkirk (1972) extends the rule to take morpheme boundaries to explain cases such as *gotta*, *oughta*, or *hadda* (p. 198-199), which she reformulates (p. 198) as:

$$t, d \rightarrow r / V (/l, r/) _ (\#) [V]_{[-stress]}$$

In another noteworthy analysis, Kahn (1976) discusses t-tapping (or flapping) rule extensively. Like Selkirk (1972), Khan (1976) discusses the phonological context of t-tapping and considers it to be a historical weakening process that had remarkably affected all American dialects, but not British ones. His analysis also pays special attention to the required phonological environment for a tap to occur, particularly the requirement of an

unstressed vowel at the right-hand environment, as a full vowel might block the flapping rule (e.g., *latex* *['leɪrɛks] cf. *later* /'leɪtə/ > ['leɪrə]). Kahn's innovative contribution implies treating the t-tapping rule as a case of ambisyllabicity, that is, treating the tap (or flap), as belonging to both syllables.

Kahn's ambisyllabicity analysis is particularly interesting since there has been a certain degree of uncertainty on what part of the syllable taps would be produced. While Kenstowicz (1994), Selkirk (1982) and Wells (1982) stated that t-tapping occurred in syllable-final positions, Giegerich (1992) pointed out that taps occurred in syllable-initial position. As mentioned above, Kahn (1976) as well as Gussenhoven (1986) described the underlying stop /t, d/ as ambisyllabic. Kahn's (1976) CV-tier phonological representation allowed to include syllable structure information in phonological rules concluding that ambisyllabicity was a conditioning factor for tapping to be produced (see 1.2.5.).

Finally, t-tapping has also been analysed within more recent phonological approaches, such as Optimality Theory. Following McCarthy and Prince's (1993), the constraints needed for tapping to occur are O_{NSET} (i.e., every syllable must have onset), $A_{\text{LING-LEFT}}$ (i.e., left-hand environments must be aligned) and $F_{\text{INAL-C}}$ (i.e., words must end in a consonant). Therefore, a constraint-based analysis where O_{NSET} and $F_{\text{INAL-C}}$ constraints satisfy ambisyllabic taps at word boundaries, as in *sought Ed* /sɔ:t ed/ > [sɔ:r ed] in McCarthy and Prince (1993, p. 51). More recently, Seo (2020) reanalysed the constraint-based analyses conducted by McCarthy and Prince (1993) previously. Seo (2020) can occur under the following constraints: O_{NSET} , $A_{\text{ATTRACT-C}}$ (i.e., the head of the prosodic word attracts the onset of unstressed syllables (Kim, 2004, p. 414; as cited in Seo, 2020, p. 335), $N_{\text{OASP\#}}$ (i.e., aspiration is not allowed), $A_{\text{LING-RIGHT}}$ (i.e., right-hand environments must be aligned) in the following constrain ranking (Seo, 2020, p. 335):

$$O_{\text{NSET}} \gg A_{\text{LING-RIGHT}} \gg N_{\text{OASP\#}} \gg A_{\text{ATTRACT-C}}$$

2.2.3. Empirical studies

Empirical studies focusing on t-tapping have mostly focused on North American varieties, although the phenomenon has also been studied in Australian English (AusE), and to a lesser extent, in British English.

The English variety that has received the greatest deal of attention from an empirical perspective is American English. In one of the earliest studies at the end of the 1970s, Zue and Laferriere's (1979) study involved three males and three females, and 3000-word tokens, each embedded in a carrier phrase (i.e., *say _ again*), that were recorded on two different occasions. They observed taps to be acoustically short, varying in duration from 10 to 40ms, with little or no sign of a release burst. Their study did not find significant differences between the duration of taps when they represent an allophonic variant of /t/ in contrast with taps as an allophonic variant of /d/. However, tap duration seemed to correlate with the phonetic features of the preceding vowel. The results indicated that the probability of occurrence of taps was almost categorical. The results also showed a difference in the production of taps between male and female speakers. While men tended to produce more taps, women tended to produce canonical forms.

In the early 1980s, Egido and Cooper (1980) conducted three experiments to analyse the influence of syntactic boundaries on t-tapping. The methodology included 34 (exp1: 10; exp2: 14; exp3: 10) native speakers of American English and four pairs of test sentences that contained the same number of syllables and fillers per experiment. The overall results showed that speakers produced few taps in sentences where a word-final boundary constitutes the boundary between two clauses. In sentences that contained boundaries between two lexical items, taps occurred in 60%-70% of the instances. Pronunciations that contained a tap at word boundaries indicated neither a silent interval nor a release burst at the boundary, contrasting with word-final [t] productions.

Another study in this decade is that by Stone and Hamlet (1982), who analysed acoustics and tongue and jaw movements during the production of /d/ and [ɾ] in nonsense phrases involving the repetition of the syllable <da> in stressed and unstressed patterns. Four different acoustic patterns for /d/ in unstressed syllables were identified concerning tongue gestures, consonant jaw height and subsequent opening jaw acceleration: a) a

voiceless or partially voiced /d/; b) a voiced /d/ with a normally long closure duration; c) a short /d/, voiced or voiceless; d) a tap-like allophone. The patterns of jaw movements observed suggested that [d] and [ɾ] were not two different sounds that varied allophonically, but a continuum roughly represented by four acoustic patterns.

Empirical interest in t-tapping increased in the 1990s, with studies such as those by Turk (1992), de Jong (1998), and Gregory et al. (1999). Turk (1992) studied both tap and non-tap environments and the duration of the vowels preceding the alveolar tap. Turk (1992) investigated three stress environments: a) between an unstressed and a stressed vowel, b) between a stressed and an unstressed vowel; and c) between unstressed vowels. Her results revealed that all stop consonants were considerably short in stressed syllables and that vowels that preceded a tapped /d/ were significantly longer in duration than those vowels that preceded a tapped /t/ only when the vowel before the tap was unstressed.

A second study in the 1990s by de Jong (1998) analysed whether American English flapping could be understood as the result of consonant-vowel coarticulation and jaw movement profile using X-ray microbeam data (see also de Jong, 2011). The data comprised a small corpus of consonants in coda position before a vowel-initial word in different prosodic environments. The data indicated no systematic difference in actual jaw position during the occlusion for taps and stops. The difference between them was later explained (see de Jong, 2011) as being greater in terms of tongue body positions, which was more evident during the closure, rather than the difference caused by the proximity of neighbouring vowels.

Gregory et al. (1999) designed several experiments to test probabilistic effects on /t, d/ deletion, tapping in word-final position and durational shortening using 8,472-word tokens ending in <t> or <d> from the phonetically transcribed Switchboard corpus of American telephone conversations (see Greenberg, Hollenback & Ellis, 1996). As for t-tapping, the results showed that neither frequency nor collocation relationship were factors that made an impact on the production of tapping. The findings also revealed that the strongest the cohesive information between the target word and the next word, the highest the likelihood of t-tapping to occur.

The 2000s and 2010s also saw a continued interest in the empirical study of t-tapping, partly due to the increasing availability of spoken corpora, although experimental and laboratory studies are also increasingly common. Fukaya and Byrd (2005), for

example, examined the articulatory foundations of the distinction between taps and non-taps word-finally and the overlap of the consonant and vowel in tapping environments across phrase boundaries. Kinematic and acoustic data in phrase-final and phrase-medial positions were explored, together with falling and level stress contours. Three informants were analysed using a sentence-reading task and 360 potential tokens were examined. 110 tokens were recorded as containing tap realisations of the 358 transcribed tokens. The study showed a discrepancy between acoustic and articulatory durational patterns given that the acoustic duration of taps was generally shorter than non-taps realisations.

An example of the use of corpora in the study of t-tapping is Patterson and Connine (2001), who analysed the distribution of allophonic variants of medial /t/ (i.e., [t, ʔ, r]) and variation as a function of word frequency and vowel length before the tapping environment in a conversational speech corpus of American English (see Godfrey, McDaniel & Holliman, 1992), supplemented with an additional transcribed speech corpus (see Carterette & Jones, 1974). 57 females and 98 males were analysed. Out of 4,093 tokens, the tap variant was produced almost categorically. Their results further showed that there was a difference in production in terms of the lexical frequency of the carrier words since high-frequency words elicited more t-tapping than low-frequency ones. The results also showed that in terms of morphological structure, t-tapping was more likely to be produced in morphologically simple words than in morphologically complex words in both high-frequency and low-frequency words. Finally, the findings indicated that the preceding vowel for medial /d/ tended to be longer when compared to the preceding vowel for medial /t/ with a difference of 16 ms when both stops were taps. The authors considered differences in vowel length to be secondary cues since previous research had indicated that tap minimal pairs were perceived as homophonous (e.g., Zue & Laferriere, 1979).

In a follow-up study, Connine (2004) conducted a word identification experiment in which a tap or a [t] variant was embedded in word-nonword continua (e.g., *pretty-bretty*) and participants had to identify the lexical item being used. The results showed more identification responses forming a real word when the to-be-identified sound occurred in the more frequently experienced tap carrier. The results revealed that lexical representations did not exclude the tap variant from the lexicon when stored as a frequent language item (see Bybee, 2001). The findings indicated that listeners were inclined to

show more identification responses when pronouncing a word with a tap variant if the sound occurred in a recurrently heard lexical item containing an underlying tap.

Patterson and Connine (2001) and Connine (2004) exemplify the increasing interest in lexical and collocational frequency from the early 2000s, as sparked by the work of Joan Bybee (2001). In this line, Warner and Tucker (2011) examined the phonetic variability of stops (/p, k, b, g/) and taps in spontaneous and careful speech in 13 American English speakers. They aimed to determine whether the reduction was mainly a casual speech phenomenon. In doing so, the study examined variables such as word frequency as well as preceding stress, segmental environment, and the presence of word boundaries. Their results indicated a greater reduction in connected spontaneous speech between unstressed syllables. Even though reduction was not observed in high-frequency words, high-frequency phrases exhibited reduction.

In another series of studies in the 2000s and into the 2010s, Eddington and colleagues focused on t-tapping and issues related to its phonetic environment and the syllabification of the tap. In his 2007 study (Eddington, 2007), for example, the author analysed 3,719 tokens related to the allophones of /t/ from the TIMIT corpus (Garofolo, Lamel, Fisher, Fiscus, Pallet, & Dahlgren, 1993; Zue & Seneff, 1996). The analysis consisted in the examination of the phonetic environment on either side of /t/ and the stress on the preceding and following syllables for the different allophonic representations of /t/. The analysis showed that analogy in allophonic variation was extremely predictable even if one of the variables was not present. Eddington (2007) categorised the distribution of the allophones of /t/ following an explicitly computational model by Skousen (1989, 1992). In terms of allophony, the simulations conducted showed that the particular allophones of /t/ (e.g., [t, t^h, ɾ, ʔ, t̚]) may appear in a given phonetic environment based on analogous allophonic forms of stored patterns in the mental lexicon (Stevens, 1998). The study concluded that sounds are grouped by speakers into mental categories, a cognitive procedure that was sensitive to orthography.

Eddington and Elzinga (2008) conducted four experiments, the first and fourth of them using the CMU Pronouncing Dictionary and the TIMIT corpus, respectively; and two experiments assessing judgements of syllabification through a questionnaire for the second experiment and a listen-and-repeat task together with an identification task for the third experiment. The researchers explored the phonetic context of t-tapping by

analysing the stress patterns, the type of vowel phoneme that follows the tap production and syllabification. Their findings showed that taps were produced more frequently in unstressed syllables and that the following vowel phoneme had some degree of influence on its production and that [ɾ] was not favoured to be produced in onset or coda positions. These results seemed to refute other theoretical investigations that claimed syllable division as a conditioning factor (cf. Gussenhoven, 1986; Kahn, 1976). Their findings further indicated that stress and vowel quality constituted the major factors when the syllable was likely to make the tap appeared word-medially.

Finally, Eddington, Treiman and Elzinga (2013) examined syllabification as a meta-linguistic entity rather than as an abstract phonological one through a questionnaire containing an Internet-based study of 4,990 bisyllabic English words and 841 participants, who had to syllabify 125 items each. The words were presented with standard orthography with a quasi-phonological transcription of possible syllabification options. The bisyllabic words were extracted from an expanded version of the Hoosier Mental Lexicon (Pisoni, Nusbaum, Luce & Slowiczek, 1985). The data were analysed by using mixed effects of logistic regression to scrutinise the effect of diverse variables. The authors suggested that a fact that may have affected syllabification was the sonority of the medial consonant. Sonorants in CVC sequences were more likely to be placed in the coda of the previous syllable, whereas obstruents were recurrently placed into the following syllable and taps seemed likely to be placed in the coda of a syllable.

While most empirical studies on t-tapping have focused on American English, some research has also focused on other varieties where it is also found, with most studies falling within the scope of sociolinguistics (or sociophonetics) research. Regarding the Southern Hemisphere, the accents that have received the most attention seem to be AusE (e.g., Tollfree, 2001) and NZE (e.g., Holmes, 1994, 1995). Within the British Isles, various accents have also been investigated including London English.

Even though t-tapping has not been extensively studied in AusEng and NZE, there are a few early studies that evidence its occurrence in these varieties of English (e.g., Bell, 1977, 1984, 1991; Holmes, 1994, 1995). The data on the occurrence of t-tapping in intervocalic position in the speech of male NZE radio newsreaders provided by Bell (1977, 1984, 1991) asserted that t-tapping was a variable feature in NZE that was strongly motivated by sociolinguistic factors. The findings showed that speakers shifted from t-

tapping in commercial stations to more standard variants (i.e., stops) in more prestigious stations. The author claimed that the style shifting concerning t-tapping was not related to the speakers' social background, but it was audience-oriented. The data showed that t-tapping was favoured before morpheme boundary positions rather than across word boundaries.

Holmes (1994), for example, investigated the use and distribution of t-voicing intervocally and t-glottalisation of word-final /t/ in NZE by using data from the *Wellington Corpus of Spoken New Zealand English* (WCSNZE). In Holmes (1994), the study examines the correlation between the production of t-tapping and speakers' age (young and middle-aged), gender (female and male) and social class (middle and working class) in two styles of speech, in conversation and formal interview. In Holmes (1994), nine different linguistic constraints were explored (p. 202), which include *it*, *that* and *sort of* as independent lexical constraints. The analysis yielded a total amount of 60 tokens per speaker. Holmes (1994) pointed out the usage of [d] as "semi categorical" (p. 215; as cited in Hay & Foulkes, 2016, p. 305) in young-working-class males' speech, constituting a "change from below" (Labov, 1994, p. 78). Holmes (1994) found a semi categorical use of intervocalic t-tapping (80%) in the casual speech of younger male working-class speakers (/d/ apparently does not tap in NZE), is favoured between vowels of unlike stress, of which 8% correspond to tap realisations before unstressed syllables, and especially disfavoured between stressed vowels.

In a more recent study on NZE, Hay and Foulkes (2016) studied the phonetic distribution of word-medial intervocalic /t/ at word-level representation and how word-medial intervocalic /t/ is implicated in the course of sound change in NZE. The data correspond to the *Origins of New Zealand English* corpus (ONZE) (see also Gordon, MacLagan & Hay, 2007) from which 98 speakers and 2,679 tokens were analysed. Among other findings, the researchers found that lexical items such as *city* have shifted in pronunciation over time, from [t] to [d, ɾ]. This is an example of how frequent words led to sound change (see also Phillips, 2006). The authors also concluded that Individual words may have an impact on sound change (i.e., words favoured by younger speakers were produced with newer variants), and even the topic of the conversation might determine which variant is favoured, in the sense that topics about the past, elicit more traditional variants, according to experience-based representations by an association of the phonetic form and non-linguistic factors, such as the topic of discussion.

Tollfree (2001) conducted a descriptive phonetic account of AusE consonants. The data used for the study derived, in small part, from a previous survey of urban speakers conducted in the 1990s; and in larger part from ongoing fieldwork conducted in Australia as part of the project *Dimensions of Spoken Australian English* interviewing 55 adolescent informants. The results obtained were analysed for four contexts: intervocalic medial, intervocalic final, pre-consonantal final and pre-pausal final. Two speaking styles were also analysed: conversational or informal and word and sentence-list or formal. For each of the four contexts, 36 tokens were examined from the two groups of speakers: a group of lower socioeconomic background and a group of middle socioeconomic background, yielding 144 tokens per group, and an overall amount of 288 tokens. Tollfree's findings indicated that t-tapping was favoured in intervocalic word-final contexts (e.g., *lot of*, *get up*) and word-medial contexts as well (e.g., *bitter*, *mutter*). Several items were identified as categorically containing [ɾ] for some speakers (e.g., *beauty*, *theoretical*, *data*). For contexts in which the plosive /t/ is preceding a syllabic lateral or syllabic nasal, /t/ and the tap realisation were found; although [ɾ] was found to be almost categorical when preceding syllabic lateral in younger speech. The phonological distribution was conditioned by phonological, prosodic and morphological factors, while their frequency was governed in part by social factors, with the lower socioeconomic background group producing less t-tapping in intervocalic word-medial contexts and formal style except in the case of word-final contexts.

Within the British Isles, Amos (2007) analysed Mersea Island English (East Anglia). The study examined data from interviews with 16 speakers. The results revealed that traditional variants of /t, d/ were present across two different generations, but that older speakers favoured the tap variant in lexicalised expressions (e.g., *sort of*, *a lot of*), in word-medial position, and across word boundaries. Tap allophonic variants were on the decline in favour of a glottalized variant in the case of /t/. The results showed that older speakers only used the voiced variant [d] for /d/. Regarding gender, the study suggested that males produced more taps than females although they represented low percentages in comparison to the ratio of production for [ʔ].

Tollfree (1999) examined two different groups of accents, South East London English Accent (SELE speakers), and regionalised forms of RP English or South East London English Regional (SELER speakers). The number of informants amounted to 32

SELE speakers and 58 SELER speakers. The findings indicated that the production of t-tapping occurred mainly in two environments: intervocalically across a word boundary, especially in phrases that contained frequent lexical items (e.g., *but I, lot of*) and intervocalically word-medially (e.g., *getting* or *better*).

As mentioned above, Tollfree (1999) analysed a regional form of RP. However, other studies have looked at this standard, supposedly regionless accent of England more specifically. T-tapping has seldom been described as a feature of this variety of English, but its use seems to characterise present-day RP, often referred to as Standard Southern British English (SSBE) as well. In her pioneering study in this respect, Hannisdal (2006), indicated that t-tapping, as used by the six BBC television newsreaders that she analysed, showed greater acceptance in idiosyncratic terms in contrast to t-glottaling. In her research, the tap variant was much more frequent than glottal stops although their distributional patterns were very similar. As for gender, men tended to use taps more frequently than it had been long described in earlier descriptions of the accent. Furthermore, her investigation revealed that glottal stops were absent. The results also revealed taps have significant variation concerning style shifting, which may be directly in connection with the level of speech attention of the newscasters under study. Given this, t-tapping may be understood as a context-dependent variable. The rate of t-tapping production found in her investigation may indicate that taps might be favoured in certain phonetic environments when a balance between authority and informality is required on the part of the speaker, as in broadcasting.

Two more recent studies on t-tapping in RP are those by Badia Barrera (2015) and Alderton (2022). Badia Barrera (2015) studied the use of t-tapping and t-glottaling in 20 teenagers and 15 adults, all of them RP speakers. The data for the study was collected using informal interviews. As for taps, the results showed an increasing use in RP, mostly in informal speech, although the rate was low (i.e., 317 cases out of 5,248 tokens). Taps were found more frequently in word-final pre-vocalic contexts, and word-medial intervocalic contexts, with the lowest score in word-medial syllabic /l/ contexts. The study also revealed that taps were produced commonly in words with a medium frequency of occurrence, and word-medial position in high-frequency items, whereas lexical frequency in word-final position did not reveal any impact. Concerning social factors, tap production occurred more frequently in adults' speech than in teenagers' speech. As for gender, males tended to produce more taps than females. The highest rates belonged to speakers

who attended boarding schools, followed by informants who attended comprehensive schools, while the lowest rates belonged to informants who attended private non-boarding schools.

Alderton (2022) studied sociolinguistic variation in t-tapping by examining elite speech in adolescent RP speakers. 45 teenagers aged 16-19 from two schools in the county of Hampshire were interviewed. The data collection consisted in reading a short story called *The Boy Who Cried Wolf* (Deterding, 2006) as well as discussing life at school. The tokens were categorised into alveolar stop, alveolar tap, glottal stop, unreleased and other in three phonological environments (i.e., word-final /t/ preceding a vowel, word-medial /t/ preceding a syllabic /l/ and word-medial /t/ preceding a vowel). His study explored tap across several linguistic factors, such as the phonological environment, word type, word frequency, and number of syllables. Alderton (2022) concluded that the tap was commonly used among adolescents who had previously attended a private school. T-tapping occurred more frequently among boys in formal contexts but also they were more prone to use taps regardless of the style or task. Girls were more likely to make taps in conversation rather than when reading. The author also suggested that taps might be used to indicate an amalgamation of authority and informality, which may be the consequence of elite speakers showing a twofold image of themselves to maintain a privileged status while showing openness and a friendly image.

2.3. Yod coalescence

Yod coalescence implies an assimilation process whereby two segments, an alveolar plosive, /t, d/ and a palatal approximant /j/, fuse or coalesce, given the mutual coarticulatory influence that segments exercise on one another (see also 1.2.1). The result is a post-alveolar affricate, /tʃ/ or /dʒ/. Yod coalescence involves not only a change in the place of articulation but also a change in the manner of articulation. The articulation of the post-alveolars /tʃ, dʒ/ turns a plosive followed by an approximant into an affricate sound. Affricates imply a stop closure at the beginning and a constriction of the articulators typical of fricatives when the plosive is released. Yod coalescence is, in this view, an intermediate sound between a plosive and an approximant, because of consonant spirantization.

Yod coalescence may occur in three contexts: across word boundaries and word internally before an unstressed vowel, and word internally before a stressed vowel, as shown in (74a-c) below (e.g., Hannisdal, 2006; Lindsey, 2019; Ramsaran, 1990; Wells, 1982):

- (74) a. don't you /'dəʊn tju/ > [dəʊntʃu] would you? /'wʊd ju:/ > ['wʊdʒu:]
 b. situation /'sɪtju'eɪʃn/ > [sɪtʃu'eɪʃn] educate /'edʒukeɪt/ > ['edʒukeɪt]
 c. Tuesday /'tju:zdeɪ/ > ['tʃu:zdeɪ] duke /dju:k/ > [dʒu:k]

While yod coalescence is normally discussed in connection with plosive followed by /j/ sequences, some authors do not restrict the term yod coalescence to /t, d/ followed by /j/ but include the alveolar fricatives /s, z/ in their definition of yod coalescence (Cruttenden, 2014; Glain, 2012), as shown in (75a-b) below:

- (75) a. *in case you need it* /ɪn keɪs ju: 'ni:d ɪt/ > [ɪŋ keɪʃu: 'ni:d ɪt]
 b. *has your letter come?* /həz jɔ: 'letə kʌm/ > [həʒɔ: 'letə kʌm]

A more accurate phonetic definition for the phenomenon of yod coalescence could, therefore, be as follows: yod coalescence refers to an assimilation process of the consonantal clusters alveolar plosive followed by yod (/dj/, /tj/) and alveolar fricative followed by yod (/sj/, /zj/) resulting in postalveolar affricates (/dʒ/, /tʃ/) or postalveolar fricatives (/ʃ/, /ʒ/), respectively. This enlarged view of yod coalescence is adopted in the present dissertation.

2.3.1. Impressionistic and theoretical accounts

Yod coalescence has long had the attention of scholars interested in the history of the English language, particularly in the evolution from Middle to Contemporary English. Some impressionistic accounts (e.g., McMahon, 2002) have argued that /ɜ/ was probably the newest arrival to the English phonological inventory that likely emerged in Middle and Early Modern English from [zj] sequences, in words such as *vision*.

Regarding present-day English, yod coalescence was commented on by Daniel Jones at the beginning of the 20th century in his famous *An Outline of English Phonetics* (Jones, 1917/2011), followed by the work of his successor, A. C. Gimson in his *Introduction to the Pronunciation of English* (1962), revised and published at present by Cruttenden (2014). Yod coalescence seems to be currently a feature of many accents of English both within and outside the British Isles (Glain, 2012). In England, for example, the phenomenon of yod coalescence is quite common in RP English (Cruttenden, 2014), although some linguists were initially reluctant to include yod pronunciations in earlier descriptions of the accent as it was considered to be informal or sloppy (Ramsaran, 1990; Wells, 1990).

One of the authors that has commented on yod coalescence from a descriptive or impressionistic point of view is John Wells. Despite its initial lack of prestige or even stigmatisation, Wells (1994a) has claimed that “RP is drifting towards categorical coalescence” (p. 203). According to Wells (1994a), yod coalescence is well-established in the pronunciation of the clitic pronouns *you* and *your* in phrases such as *what you want* [wɒɜwɒnt] and in questions tags as in *don't you* as /'dʌʊntʃu/, although coalesced forms are avoided in a careful or mannered style, as yod coalescence is sometimes perceived as a Cockneyism. Yod coalescence in stressed syllables across morpheme boundaries (e.g., *tune* /tju:n/ > [tʃu:n]) is increasingly found (Glain, 2012; Lindsey, 2019; Wells, 1997) but generally perceived as non-RP, which is frequently used to compare RP to other varieties of English, such as Cockney or Estuary English. The latter was reported to be representative of middle-class speakers in the Southeast of England and placed linguistically somewhere between RP and Cockney (Bauer, 1994; Fabricius, 2000; Ramsaran, 1990; Wells, 1994a). In fact, despite their increased use in standard accents like RP in England and the suggestion in impressionistic accounts that uncoalesced

variants may be as traditional or prestigious forms, coalesced /tʃ, dʒ/ segments are commonly associated with non-standard accents, such as Estuary English (Hannisdal, 2006).

Impressionistic accounts are generally limited insofar as they only describe yod coalescence, the context where it appears, and point out perceived tendencies in its use in varieties of English. Similarly, theoretical accounts of yod coalescence are few and hardly ever contain analyses in terms of generative phonological rules, OT constraints, or any other phonological framework. Conversely, phonological analyses of yod coalescence derive mainly from analyses of the historical development of the English language, so yod coalescence is approached from a diachronic phonological perspective. This may be related to the fact that the palatal approximant /j/ or ‘yod’ has had different evolution paths in different accents of English. Yod coalescence is often discussed in connection with the phenomenon of yod dropping in the history of varieties of English and the appearance of this phoneme in some items of the GOOSE lexical set. In this respect, both yod coalescence and yod dropping in sequences involving the vowel /u/ are contemporary manifestations of a long articulatory tendency leading to the disappearance of /j/ (Cruttenden, 2008).

According to Wells (1982), the GOOSE lexical set contains two different groups in many English accents: those items that have /u:/ (e.g., *blue*, *noon*, *do*) and those that have /ju:/ (e.g., *new*, *tune*, *you*). These two groups are mostly the result of different sound changes (Amos, 2007; Beal et al., 2020; Glain, 2012; Wells, 1982 for some accounts). One relevant element of this evolution is that soon afterwards the cluster /Cju/ was completely established in the English inventory, and /j/ started to be dropped because of a stress-shifting process after some consonants such as other palatals, the liquid /r/ and in the sequence /Cl/, as in *chew*, *juice* or *blue* (Glain, 2012; Kwon, 2006). This stress-shifting process resulted in an early yod dropping process that implied the deletion of /j/ when preceding a coronal consonant, although /j/ was retained when followed by a labial or velar (Amos, 2007; Wells, 1982). Early yod dropping affected /Cju/ in different stages. Firstly, after palatals in words such as *juice*. Secondly, after /r/ in words such as *grew*. Thirdly, after consonants followed by /l/ in words such as *flew* (Amos, 2007; Clarke, 1993; Wells, 1982). The dropping of /j/ in pairs of words containing the sequence /Cju/

led to pairs of homophones in the majority of English accents (Glain, 2012), for example, *threw* and *through* (/θru/).

The yod phenomenon emerged and affected differently to the phonological inventories of the different accents of English. For instance, yod coalescence exhibits variability in British accents such as RP whereas yod dropping is common in American English. This yod dropping phenomenon in American English varieties affects the yod phoneme when preceded by a coronal consonant in words such as *new* (e.g., AmE /nu:/), although /j/ is retained after labials, velars and labiodentals, such as *beauty*, *cute* (Cruttenden, 2014; Glain, 2012; Wells, 1982).

Given the long tendency in the English language to turn /tj/ and /dj/ into /tʃ/ and /dʒ/, respectively (Wells, 1997), some varieties that preserved /j/ in these contexts, such as RP, may currently exhibit a higher degree of yod coalescence while other varieties, such as American English, have a lower degree. This may explain why American phonologists have largely ignored or insufficiently addressed yod coalescence from a synchronic phonological perspective.

2.3.2. Empirical accounts

As discussed previously, impressionistic, and theoretical ways tried to account for the emergence and description of yod coalescence. More recently, empirical approaches have investigated this phenomenon from data-driven perspectives, which are discussed in the present section. Several studies investigated yod coalescence in British accents (e.g., Altendorf, 2003; Amos, 2007; Beal et al., 2020; Glain, 2012, 2014; Hannisdal, 2006; Spurling, 2004; Tollfree, 1999), although the phenomenon has also been studied in American English (e.g., Glain, 2012), Canadian English (e.g., Chambers, 1998; Clarke, 1993, 2006) and NZE (e.g., Hay & Foulkes, 2016).

Empirical studies on yod coalescence in varieties of British English have often studied both yod coalescence and yod dropping. They have often compared two or more varieties of British English, often using RP as a reference variety for comparison. Regarding the methodology implemented in the research designs, some studies have used standard sociolinguistic methods while others have used historical corpora.

Altendorf (2003) examined yod dropping after the alveolar plosives /t, d, n/ and yod coalescence after /t, d/ in stressed syllables at morpheme boundaries in Estuary English. The study focused on young EE/RP girls' speech using Labovian's structured interviews. The results suggested that yod dropping is a low prestige or non-standard variant, commonly associated with working-class and middle-class speakers and infrequent in middle-class girls' speech. Speakers from middle and upper classes produced yod coalescence, which supported the statement that yod coalescence "seems to have been spreading socially into the upper middle classes" (p. 99), although with a low ratio of production.

In Spurling (2004), 12 informants from Ipswich (Suffolk) were analysed involving two data collection methods: a forty-five-minute interview and a twenty-five questions and statements reading list. The results showed that only young females favoured [ju] forms, while young male speakers favoured coalesced forms [tʃ, dʒ]. Yod coalescence realisations were categorical for young male speakers, whereas young female speakers produced yod coalescence forms in only 50% of the tokens. [u] was still widely prevalent among middle and old-age groups for both male and female speakers. The author concluded that young females favoured yod-less forms, while young males favoured coalesced variants. The findings showed that Ipswich English speakers only used the two ends of the continuum; they used either yod-less forms or coalesced variants, [u] or [tʃ, dʒ].

Amos' (2007) study aimed to examine variation concerning the presence or absence of palatal [j] (e.g., [ju]) and yod coalescence [tʃ, dʒ] in Mersea Island English in the speech of 16 speakers (old and young speakers) across 14 interviews. The results indicated that traditional yod-less realisations were on the decline in favour of a more standard [ju] variant. The analysis indicated that 62% out of 480 tokens corresponded to [ju] as the dominant form in /Cju/ sequences. Regarding yod coalescence, [tʃ] represented 51% and [dʒ] 31%. The results revealed a great deal of variation across generations and gender in Mersea Island English. While no yod coalescence tokens for older females were found; yod coalescence forms for [tʃ] and [dʒ] are found in older male informants. The yod coalescence rates of [tʃ] and [dʒ] in younger speakers reveal a quite high percentage

of production for both female and male speakers, being categorical in younger males' speech.

Hannisdal (2006) investigated yod coalescence word-internally before stressed vowels in the speech of 30 BBC newsreaders. The findings revealed that the coalesced variants /tʃ, dʒ/ were common in stressed syllables. The results showed that the unusually high amount of 46.4% suggested the introduction of already widespread yod coalescence forms into the RP inventory. The author concluded that the phenomenon was very well established within the RP inventory and, therefore, yod coalescence should not be excluded from the description of the accent.

While the previous studies mostly used standard sociolinguistic data collection methods, some other empirical studies have analysed data from dictionary entries. Beal et al. (2020), for example, conducted research on historical phonology by compiling data involving yod coalescence and yod dropping from 18th-century pronouncing dictionaries. The results showed some evidence of yod coalescence pronunciations before /ju:/, particularly concerning coalesced forms with /s/ and /z/ from the 17th century onwards; while yod coalescence of /t/ and /d/ was an infrequent realisation. Stigmatisation was found in the production of yod dropping in all contexts and of yod coalescence in stressed syllables due to the influence of several factors, such as stress, type of phoneme, rhoticity and other extralinguistic factors, such as frequency, in the dictionaries examined (Beal et al., 2020). The study concluded that there was also a tendency in the use of yod coalescence in the last quarter of the 18th century and that yod dropping variants were perceived to be a vulgar innovation by the end of that century.

Beyond the British Isles, some studies focusing on yod coalescence have looked at its use in North American varieties. Chambers (1998), for example, analysed changes involving pronunciation and lexical variants, such as yod dropping, among others, in Canadian speakers of different ages in a geographical region where standard Canadian English is in close contact with American English (i.e., the border in the Niagara Falls-Buffalo region). In this study, several variables that were undergoing change were examined by collecting data through surveying a large sample of male and female speakers whose ages ranged from adolescents to elderly speakers (935 Canadians and 80 Americans). The data was extracted from a geographical region of Canada where the standard dialect used by the linguistic community was in close contact with American

English. Chambers (1998) pointed out that the percentage of yod dropping was relatively high for the oldest respondents, although it was higher for the younger ones. Younger speakers favoured yod dropping pronunciation, while older speakers favoured yod-retained pronunciation. As attributing the cause of change to Americanisation is too simplistic, the author concluded that, even though yod dropping was still increasing in the speech of younger people in Canadian English, pronunciations with yod represented a minority, even for older people who were born in the 1920s.

Another series of studies focusing on North American varieties are those conducted by Glain (2012, 2014). Glain (2012) looked at yod dropping and yod coalescence in American English and British English in the transcriptions from two editions of *Longman Pronunciation Dictionary* (Wells, 1990, 2008). The author claimed, based on this transcriptional evidence, that “[i]n unstressed syllables, the change towards coalescence is already well-established” (p. 13) concerning the coalescence of /sju/ and /zju/ segments into /ʃu/ and /zu/, whereas in stressed syllables, it may be regarded as a change in progress. Coalescence of /tju/ and /dju/ segments in stressed syllables is currently widespread; while yod coalescence is virtually non-existent in stressed syllables in accents of English in which yod dropping is frequent, such as in General American (Glain, 2012).

Glain (2014) conducted a corpus-based study to explore yod coalescence in speakers from England, Scotland, and the USA. He used data from IDEA (*International Dialects of English Archives*). The data consisted of 216 recordings from *Comma gets a Cure* and 315 interview recordings. The author suggested that there was a widespread production of coalesced forms in England, Scotland and in the USA. Coalesced pronunciations were increasingly more noticeable with speakers who were born in the late 1960s or early 1970s. Scottish and USA speakers exhibited a higher rate of palatalised productions for scripted speech than English. Men used more palatalised pronunciations than women in both Britain and the USA. No significant differences were observed between scripted and unscripted speech. The study concluded that contemporary yod coalescence was a linguistic trend that started in the latter part of the 20th century due to the standardisation of non-standard forms of language.

2.4. Factors affecting linking /r/, t-tapping and yod coalescence

The present section provides a detailed description of the factors that, according to the impressionistic accounts, theoretical analyses, and data-driven studies may affect linking /r/ (or /r/-sandhi more generally) as well as t-tapping and yod coalescence. The four main types of factors are grouped below under four different categories: phonetic, linguistic, sociolinguistic, and usage-based. Providing an overview of the factors that may condition variation in the three phenomena under investigation in this thesis is important so that a more informed decision could be made regarding which variables are used in the empirical studies of this dissertation (see Chapter III).

Phonetic factors

Linguistic factors at the pronunciation level are generally referred to as phonetic factors. The phonetic factors include the type of vowel phoneme, syllable stress, and the prosodic environment of the potential syllables and words involving the production of the phenomena, and more particularly, the presence of another /r/ in the vicinity as well as the length of the syllable flanking the /r/-sandhi site in the case of linking /r/, and the syllable structure for t-tapping and yod coalescence.

Regarding linking /r/ production in terms of phonetic variables, a potential factor affecting /r/-sandhi variation is the quality of vowel flanking at the /r/-sandhi site. In this respect, previous research found no effect of the preceding vowel in the production of both linking /r/ and intrusive /r/ in RP (e.g., Mompean & Mompean-Guillamón 2009, Pavlík 2016); although Pavlík (2016) found that both linking /r/ and intrusive /r/ are more common in RP when followed by mid-central or close front vowels.

Regarding the vowel context where t-tapping occurs, Eddington & Elzinga (2008) also argued that [ɾ] tended to be placed in coda position of the syllable when preceded by short vowels, and in the onset of the syllable when preceded by a long vowel phoneme, and that tap or flap-associated phonemes are [ə, ɪ, i, ʌ, l].

Syllable stress and the prosodic domain have long been suggested as influential factors in the production of linking /r/. While Jones (1918/1972) claimed that linking /r/

is generally avoided when followed by an unstressed syllable; Allerton (2000) and Foulkes (1997b) argued that linking /r/ is likely to be absent before a stressed syllable, as in *your age* [jɔ:(r) 'eɪdʒ]. According to specialised literature, the possibility of introducing a linking /r/ when there is a pause across morpheme boundaries seems to be lower than if there is no pause and if the linking words belong to the same prosodic domain (e.g., Bauer, 1984; Jones, 1918/1972; Joseph, 1999). One possible explanation for this is that linking /r/ is often avoided before a clause boundary as it corresponds to an intonation unit, which usually coincides with a tone group boundary (Jones, 1909/1956).

Stress has also long been proposed as a conditioning factor in the specialised literature in the production of t-tapping (e.g., Borowsky, 1991; Giegerich, 1992; Kahn, 1976; Ladefoged & Johnson, 2015; Selkirk, 1972). While Cruttenden (2014) and Jones (1917/2011) suggested that t-tapping was commonly restricted in post-accentual positions and never followed by stressed syllables, in a remarkable analysis of the t-tapping phonological context, Kahn (1976) discussed the required phonological environment for a tap to occur, particularly the requirement of an unstressed vowel at the right-hand environment.

However, these prosodic environments seemed too restrictive as [r] pronunciations can be intervocally followed by a stressed or unstressed syllable either across morpheme boundaries or across word boundaries (e.g., *whatever* [wɒr'evə], *at all* [ət 'ɔ:l]). The preceding and following syllables could, therefore, be stressed or unstressed (e.g., Riehl, 2003).

As for yod coalescence production, the relevant literature points out that the stress of the syllable is an influential factor as well. The plosive /d/ might display yod coalescence in stressed and preceding stressed syllables; /t/ and /s/ might only produce yod coalescence word-initially; whereas /z/ is considered to produce yod coalescence in all syllable positions; however, yod coalescence is said to be more frequent with /s, z/ after stressed syllables rather than /t, d/ (Beal et al., 2020). While Cruttenden (2014) considered yod coalescence a frequent feature of fluent conversational speech operating either across word boundaries or across morpheme boundaries before an unstressed vowel, Ramsaran (1990) pointed out that, many speakers use coalesced and uncoalesced forms, even within the same conversation.

Glain (2014) pointed out a tendency towards the acceptance of coalescence in stressed syllables, in line with Hannisdal (2006), who argued that the number of occurrences of /tʃ, dʒ/ in stressed syllables is higher than estimated in the specialised literature. As Glain (2012, 2014) suggested, yod coalescence after /s, z/ is well-attested in unstressed syllables, but not as widespread in stressed syllables and does not occur as frequently as yod coalescence after /t, d/.

As for the distribution of the allophonic tap variant concerning the syllabic structure, it seems to vary considerably and it remains unclear (Eddington et al., 2013; Shockey, 2003; Riehl, 2003). For instance, Kenstowicz (1994), Selkirk (1982) and Wells (1982) suggested that t-tapping was produced in the syllable-final position, while Giegerich (1992) argued that a more suitable context was syllable initial position. However, Gussenhoven (1986) and Kahn (1976) claimed that taps were always ambisyllabic, and Hammond (1997) stated that ambisyllabicity seemed not to be an accurate distributional pattern for tapping, who proposed foot-domain rules instead of resyllabification.

In terms of syllable structure, both yod coalescence and yod dropping could be understood as yod cluster reductions (Glain, 2012; Hannisdal, 2006), that is, as articulatory mechanisms for reducing complex syllable onsets towards the preferred simple CV syllable structure (e.g., Hannisdal, 2006; Lutz, 1991). Yod dropping reduced the complex CCV onset by deleting the second consonantal sound, whereas yod coalescence reduced the complex CCV by fusing two neighbouring consonants (Hannisdal, 2006).

Finally, a particular phonetic factor that solely affects linking /r/ is the presence of another /r/ in the vicinity (before, after or at both sides) of the potential /r/-sandhi site (e.g., *roar of traffic* /'rɔ(r)əv/). While Hannisdal (2006) pointed out that /r/-sandhi is frequently avoided when preceded and/or followed by another /r/, other studies (e.g., Hay & Sudbury, 2005; Mompeán & Mompeán-Guillamón, 2009; Pavlík, 2016), did not observe any correlation. Some studies also pointed out that /r/-sandhi seems to be avoided if the syllable that would make the link begins with /r/, which concerns the tendency to avoid similar sounds in adjacent positions, as in *Victo[r]ia[r]* and *Albert* /vɪk'te:riə ən 'ælbət/ (e.g., Mompeán & Mompeán, 2007; Mompeán & Mompeán-Guillamón, 2009).

This may have its origin in a dissimilation process, that is, avoidance of two similar adjacent r-sounds that seems to be motivated by a lack of preference in natural languages for similar sounds in the same context (e.g., Ashby & Maidment, 2005; Brown, 1988a; Ladefoged & Johnson, 2015; Mompeán & Mompeán-Guillamón, 2009).

Linguistic factors

Linguistic factors that may affect the production of the three phenomena under investigation mainly refer to structural factors, which include lexical as well as grammatical factors. These include the lexical status of the word (i.e., lexical and function word, word-class combinations), the syntactic environment, and the morphological environment (i.e., word internal or word external in the case of linking /r/).

The type of lexical item that follows the r-link may also affect /r/-sandhi production. Linking /r/, for example, is inhibited before proper names and honorifics (e.g., *Mr*) to avoid ambiguities and misinterpretations (Hannisdal, 2006; Lewis, 1975/2014). Another linguistic factor that potentially influences /r/-sandhi production is the word class of the items flanking the /r/-sandhi site. Even though Hay and Sudbury (2005), Mompeán (2022) and Pavlík (2016) found little or no evidence of the effect of word class on /r/-sandhi production, Bauer (1984), Foulkes (1997) and Viollain (2014) noted that linking /r/ is more frequent after monosyllabic function words.

In terms of the lexical status of words containing a potential tap production, Carr (1991) pointed out that [ɾ] is often produced in function and content words (e.g., *but*, *not*), in some monosyllabic and bisyllabic verbs (e.g., *put*, *had*, *submit*), nevertheless, t-tapping does not occur if the verb is unstressed on the last syllable (e.g., *edit it*) that might be due to the presence of a foot boundary between /t/ and the following vowel.

Yod coalescence across morpheme boundaries is perceived as a type of lexical-incidental variation since words that exhibited yod coalescence across morpheme boundaries do not belong to everyday vocabulary, and therefore, reduction or coalescence from /tju, dju/ to /tʃu, ɟu/ does not occur frequently (e.g., Glain, 2012, 2014; Hannisdal, 2006).

In terms of the syntactic structure, /r/-sandhi is claimed to be less frequent at clause boundaries as well (e.g., *she opened the door and went out*) since syntactic boundaries mark the division of two different syntactic units (Foulkes, 1997b; Hannisdal, 2006; Jones, 1909/1956; Mompeán, 2022). In this respect, Hannisdal (2006) pointed out that, although the presence of clause boundaries did not prevent the production of /r/-sandhi significantly, the frequency of /r/ was lower at clause boundaries than in other environments.

The incidence of the syntactic environment is a crucial factor in the production of t-tapping as well. Egido and Cooper (1980) identified three linguistic factors that motivate the production of taps. Firstly, taps are not blocked because of pauses but by syntactic boundaries. Secondly, the occurrence of taps is noticeably reduced when the word boundary coincides with an embedded clause boundary. Thirdly, tapping is also avoided in the context of a syntactic deletion. Additionally, Parker and Walsh (1982) also suggested that taps are avoided if they are preceded by fronted adverbial phrases since they intensify the complexity and length of the syntactic structure.

/r/-sandhi has also been claimed to be sensitive to the morphological structure since it can be found both across word boundaries and at word-internal morpheme boundaries in affixed words (e.g., *feve[r]ish*, *draw[r]ing*) and compounds (e.g., *four-yea[r]-old*, *salsa[r]-evening*). For instance, linking /r/ before suffixes is considered to be categorical while intrusive /r/ typically occurs in affixed words, especially after nominal suffixes (e.g., *-ism*), adjectival suffixes (e.g., *-ish*, *-al*), adverbial suffixes (*-ly*) and verbal suffixes (e.g., *-ing*, *-ise*). Linking /r/ and intrusive /r/ in compounds exhibit some degree of variability (e.g., Hay & MacLagan, 2010; Hay & Warren, 2002; Mompeán, 2022; Mompeán & Mompeán-Guillamón, 2009). Hay and MacLagan (2010) and Mompeán and Mompeán (2009) found that intrusive /r/ was more frequent word internally in affixed words, such as *withdrawal*, than across word boundaries, whereas Hannisdal (2006) found that linking /r/ is more frequently produced within compounds than across word boundaries.

Concerning word class combinations, as Pavlík (2016) pointed out, the combination of lexical and content words may have an impact on the production of /r/-sandhi. For instance, Hannisdal (2006) found that the combination of two lexical words disfavours the use of linking /r/ in RP, while higher frequencies were found between

function words, which may be explained because function words are generally unstressed and carry little informational value.

Concerning t-tapping production, Gregory et al.'s (1999) results showed that neither frequency nor collocation relationship were factors that made an impact on the production of tapping whereas Patterson and Connine's (2001) results indicated that there was a difference in production in terms of lexical frequency of the carrier words since high-frequency words elicited more t-tapping than low-frequency ones. The results further showed that in terms of morphological structure, t-tapping was more likely to be produced in morphologically simple words than in morphologically complex words in both high-frequency and low-frequency words.

Sociolinguistic factors

The sociolinguistic aspects that may influence variability in the three connected speech phenomena usage refer to variables such as speech context, age, gender and social class, level of education, as well as accent-dependent factors, among others.

The speech context is a motivating factor since speakers tend to produce less linking /r/ in formal styles whereas the production of the r-link is usually made in not deliberate speech styles or in fluent colloquial contexts (Brown, 1988; Lewis, 1975/2014), although Ramsaran (1978) claimed that linking /r/ is a frequent phenomenon in all speech situations. Regarding the speed of the speech context, Lewis (1975/2014) stated that the absence of /r/-sandhi is not infrequent in both rapid and slow speech, however, Bauer (1984) found no effect of speech tempo on the production of /r/-sandhi.

Most research on /r/-sandhi in RP found no evidence of variation in /r/-sandhi being linked to age or gender (e.g., Bauer, 1984; Hannisdal, 2006; Mompeán & Mompeán-Guillamón, 2009; Pavlík, 2016). Age, however, is a factor that represents patterns of variation. The study by Bauer (1984) on RP found that younger speakers tended to use more intrusive /r/ while Foulkes (1997a, b) found that participants' age in Derby and Newcastle did not have any correlation between the use of linking /r/; however, the Newcastle data showed that older speakers use linking /r/ more frequently. Similarly, in 19th-century NZE, no correlation between date of birth and the use of linking /r/ was observed (Hay & Sudbury, 2005).

In terms of age factors, t-tapping is much more frequently used in younger male speakers -boys from a private-school background used t-tapping more frequently in Alderton (2022)- than in older generations when exploring the social dimension in NZE and RP (see Badia Barrera, 2015; Hannisdal, 2006; Holmes, 1994). Older informants seemed to favour the use of the tap variant over the glottal stop in lexicalised phrases (e.g., *sort of*, *(a) lot of*) (Amos, 2007). Likewise, Watt and Milroy (1999) suggested that [r] was significantly more common among younger speakers than older ones in Newcastle. Regarding gender differences, tapping was found to be more frequently used by men than by women (e.g., Holmes, 1994; Wyatt & Milroy, 1999; Woods, 1991). A pattern also observed in the pronunciation of medial /t/ in American English was that female speakers tended to articulate categorical forms and avoid forms related to casual speech processes (Zue & Laferriere, 1979). T-tapping was produced in a spontaneous or casual style and appeared to be restricted to moderately swift speech according to Hannisdal (2006) and Wells (1982).

Yod coalescence pronunciation has been attested to be commonly favoured by younger generations of speakers (e.g., Amos, 2007; Spurling, 2004) while older speakers showed a tendency to produce high rates of occurrences of yod coalescence forms only in word-medial position as in *situation*, avoiding the production of coalesced forms before stressed syllables, as in *Tuesday* (Amos, 2007). Conservative pronunciations containing uncoalesced forms are frequently perceived as more appropriate; while coalesced pronunciations are not only observed as incorrect but also more contemporary, given their association with younger speakers (see Glain 2012, 2014).

Regarding the speakers' level of education, the production of linking /r/ and intrusive /r/ is strongly associated with a phonetic production supported by the spelling, which may be related to the speaker's orthographic awareness of the language. As /r/ is not orthographically present, speakers tend to regard intrusive /r/ as a vulgarism (e.g., Jones, 1909/1956; Knowles, 1987; Lewis, 1975/2014; Wells, 1982).

In terms of prestige, yod coalescence stigmatisation on the part of the speakers may be related to the fact that the pronunciation is not supported by the orthography as forms that maintain phonological contrast carry greater awareness on the part of the speakers (Fabricius, 2000; Trudgill, 1986). While the high-status variant coincided with orthography, the stigmatised variant does not. Ramsaran (1990) argued a considerable

resistance for RP speakers to adopt yod coalescence pronunciations for considering them vulgar, while Wells (1997) considered yod coalescence common realisations in RP speakers' speech. This led to a long disagreement on whether yod coalescence forms belong to the RP speakers' inventory or not.

T-tapping is a recurrent pronunciation in many varieties of English to variable degrees. It is, for example, predominantly widespread in the English spoken in North America (see Eddington, 2007; Patterson & Connine, 2001; Shockey, 2003). It is also well-attested in Canadian, Australian and Irish English (e.g., Woods, 1991; Holmes, 1994; Trudgill & Hannah, 2008; Wells, 1982), but also fairly frequent at word boundaries in British English. Within the British Isles, Cockney is the only English variety where the voiced tap is a typical realisation of the intervocalic /t/ (Wells, 1982).

However, recent studies have shown that [ɾ] realisations are also attested in other varieties of English in the UK (see Foulkes & Docherty, 1999). Shockey (2003) further argued that even though it is often heard in high-frequency words (e.g., *British*), t-tapping is much less characteristic in RP. In terms of prestige, the voiced tap can be located in an intermediate social status between the stigmatised glottal stop and the prescribed plosive, which is used to evade the more socially marked pronunciation in different accents (e.g., Badia Barrera, 2015; Hannisdal, 2006; Trudgill, 1986).

Yod coalescence can be considered to be a social phenomenon as it exhibits great variability among the different varieties of English. The frequency of yod coalescence production observed in Altendorf (2003) may indicate that yod coalescence is a socially spreading variable in higher social classes in London, Colchester and Canterbury. Wells (1982) claimed that even in Popular London speech, there has been a switch towards yod coalescence in favour of yod dropping. In the southern hemisphere, yod variation is considered not to be geographically dependent, but rather a sociolinguistic matter in AusE and NZE (Burridge, 2008; Hovarth, 2008).

Usage-based factors

The phonetic, linguistic and sociolinguistic factors mentioned earlier may correlate with usage-based factors, which are mainly associated with the effects of frequency of usage,

such as morphological, lexical, collocational and vowel type frequency as well as lexicalisation or entrenchment.

Lexical frequency (i.e., the frequency of usage of a particular lexeme) and collocational frequency (i.e., frequency of sequence of words that co-occur more often) are determining factors. Particular high-frequency lexemes and set phrases might influence the incidence of occurrence of /r/-sandhi. As Mompeán (2022) pointed out, the general conclusion is that higher frequencies favour higher rates of /r/-sandhi. Some studies, such as Hay and MacLagan (2012) found that a high frequency of the previous word favoured the use of linking /r/ in NZE, in contrast to Mompeán (2022) who pointed out that the lexical frequency of the linking word did not have any influence on the linking /r/ production in RP.

Hay and Sudbury (2005) suggested that the use of linking /r/ is disfavoured if the word after the /r/-sandhi site had a high lexical frequency in AusE. These coincided with Mompeán (2022) results which suggest that the r-link is less likely the lexical frequency of the word after the potential /r/-sandhi site is high. Even though Pavlík's (2016) findings revealed no influence of word frequency on /r/-sandhi production, the frequency of the vowel type favoured the production of /r/-sandhi. Similarly, Hay and Sudbury (2005) found that /r/-sandhi is frequent if the following vowel has a medium lexical frequency in AusE whereas the frequency of production of /r/-sandhi was higher if followed by mid-central and close front vowels in RP, which are, as Mompeán (2022) pointed out, "the most frequent /rV/ sequences in connected speech" (p. 3). However, Mompeán (2022) found that vowel type frequency did not affect the use of linking /r/.

Concerning collocability and collocability frequency, some studies, such as Cox et al. (2014), Hannisdal (2006), Hay and Sudbury (2005), Mompeán and Mompeán-Guillamón (2009), and Navarro (2013) found that linking /r/ seemed to be commonly produced in frequent collocations than in free combinations of words. As Hay, Jannedy and Mendoza-Denton (1999) and Mendoza-Denton, Hay and Jannedy (2003) argue, the production of /r/ when the following segment is characterised as an infrequent word tends to have little or no effect on the /r/-sandhi production, while frequent word combinations exhibit greater phonetic variation. Some studies suggested that common collocations, such as *for instance* and *after all*, favoured the use of linking /r/ in RP (Jones, 1909/1956; Lewis, 1975/2014). Their findings revealed that "common phrases containing highly

predictable sequences of words are stored and accessed as a whole” (Mompeán 2022, p. 249).

Concerning word frequency and t-tapping production, Patterson and Connine (2001) also pointed out that two lexical dimensions, frequency and morphology, influenced the occurrence of taps in American English. Taps are predominantly present in high-frequency words; in contrast, low-frequency words exhibited a greater variability with fewer taps. As Connine (2004) suggested, high-frequency lexical items may show a greater lexical diffusion pattern concerning tap productions in American English rather than in British English, although Shockey (2003) stated that [r] realisations were produced by many speakers of RP, especially in high-frequency words.

As for yod coalescence, Glain (2014) pointed out that contemporary yod coalescence pronunciations are commonly affected by the lexical frequency principle, which suggests that high-frequency items were more likely to undergo yod coalescence (e.g., the clitic pronoun and possessive adjectives *you* and *your*). This might explain why yod coalescence involving the pronouns *you* and *your* seem to be well established in casual RP although some speakers consciously avoided using it in careful speech (Wells, 1994a).

2.5. Concluding remarks

The empirical studies have sometimes confirmed the observations of theoretical and/or impressionistic approaches but also revealed different factors affecting the phenomena patterns of usage.

The phonological factors include the quality of the vowel flanking the potential /r/-sandhi, tap and yod coalescence sites, the presence or absence of another /r/ in the vicinity in the case of /r/-sandhi as well as clause or tone boundary divisions and pauses in the case of /r/-sandhi and syllable stress for /r/-sandhi, t-tapping and yod coalescence.

Some linguistic factors have been discussed as motivating factors in the production of these three connected speech phenomena. Notably, the lexical status of words and certain word-class combinations (e.g., two function words).

Similarly, the production of linking /r/, t-tapping and yod coalescence seems to be widely conditioned by usage-based factors, such as lexical and collocability frequency (i.e., high-frequency lexical items, high-frequency word collocations).

The production of these three connected speech phenomena also seems to be widely conditioned by sociolinguistic constraints that can be found such as age, gender, socioeconomic status, as well as speaker's language awareness.

CHAPTER III: Objectives and methodology

3.1. Main objectives

The general aim of the current dissertation is to provide empirical evidence on the use of three phenomena of linking /r/, t-tapping, and yod coalescence in the non-rhotic variety of English known as Received Pronunciation (RP). In this respect, while the body of research conducted so far reveals some patterns concerning the usage of linking /r/, t-tapping and yod coalescence in RP; research concerning oral or spontaneous speech remains more limited as the majority of the empirical studies has generally focused, although not exclusively, on a BBC newscast context. In addition to this specific register, some studies involved informants reading sentences containing experimental targets or nonsense words; however, little is known about the actual production of connected speech processes in a spontaneous context. This study, therefore, aims to achieve a better understanding of linking /r/, t-tapping and yod coalescence by compiling a small-scale oral corpus from TV interviews.

The decision to conduct some research on the aforementioned connected speech processes is based on the assumption that conversational speech may reveal a wide range of phonetic data and detail, and because most of the empirical investigations conducted so far have focused on elicited or scripted speech. In contrast to elicited or scripted speech, which requires speakers' attention, connected speech represents a particularly rich form of language production, since speakers are not consciously aware of their own production (see Ash & Grossman, 2015). From the standpoint of natural conversation, exploring the variability in the production of each of the phenomena will contribute significantly to the existing empirical body of data.

As mentioned above, the accent chosen for the current studies is Received Pronunciation (RP). It is the most extensively described accent of British (e.g., Gimson, 1962/1980; Jones, 1917/2011; Wells, 1982), typically considered to be non-regional but with a strong presence in the southeast of England (e.g., Roach, 2004; Wells, 2000). Native speakers of RP belong to the middle or upper classes and the accent has traditionally been considered to be prestigious and the standard pronunciation variety in England since the 19th century, although its popularity has diminished in recent decades (e.g., Cruttenden, 2014; Ramsaran, 1990).

RP has been selected as the accent of study partly for various reasons. Firstly, it is an accent that is easily accessible in broadcasting materials, given its traditional use by the BBC since the early 1920s (Cruttenden, 2014, p. 77) and its de facto standard in England (Trudgill, 2001). This use allowed a larger margin in terms of the selection of a corpus of interviews where the connected speech processes under study could be investigated. Moreover, the accent has been widely described for decades and there is evidence that it has undergone substantial phonetic changes, particularly over the last half century or so (e.g., Wells, 1997), perhaps due to the influence of other accents of the South East of England and the London region (e.g., Rosewarne, 1984, 1994; Wells, 1994a,b). The patterns of variation that may be observed in the study of linking /r/, t-tapping, and yod coalescence in this dissertation, therefore, can be put in perspective against this background of studies (impressionistic, theoretical, or empirical). A more recent empirical account of those processes in RP will also shed some light on linguistic, sociolinguistic, usage-based and phonetic aspects that may have a bearing on the three phenomena under investigation.

The idea of compiling a small-scale corpus relates directly to the empirical approach of this dissertation and the assumption that empirical methods help researchers revise and update their knowledge of specific phenomena and their theoretical approaches. A small-scale corpus based on spontaneous or connected speech provides a set of speech data occurring naturally in conversation. The compilation of an oral-based corpus will make it possible to gain understanding of multiple phenomena occurring in spontaneous discourse although, in this study, only changes across word boundaries will be addressed (see 1.1.). As mentioned above, the dataset will try to provide a better understanding of the effect of different linguistic, sociolinguistic, usage-based and phonetic factors on the three connected speech processes under investigation across word boundaries.

In addition to using a corpus-based methodology, a functional, usage-based approach may also appear to be appropriate to the analysis of connected speech processes once the specific results are obtained. Language data such as the one analysed in this dissertation is not linguistic material produced *in vacuo* and decontextualized from its original setting but *in situ*, anchored to the context it was produced in. Therefore, a number of variables have to be taken into account (e.g., sociolinguistic, usage-based, etc.). For example, frequency often has an impact on linguistic forms and their storage and

processing (see Behrens, 2009; Bybee, 2001). As Pavlík (2016) pointed out, conducting an investigation following a more functional, usage-based approach allows to model the correlation of linguistic phenomena in terms of patterns of usage, token-frequency, categorization and language variation.

Within this more functional approach, words in the lexicon consist of a cloud of phonetic exemplars or phonetic variants that contain detail information about the recurrent context in which they appear (e.g., Bybee, 1999, 2001). As storage implies a network of connections between related items, following this usage-based framework, mental phonetic images are stored in a speaker's knowledge of the language, which operates by the influence of factors such as the type of token and collocation, lexical and collocability frequency (e.g., Bybee, 1999, 2001). These factors may result in lexicalization or lexical entrenchment, which refers to a cognitive routinization process of the linguistic units and structures based on repetitive events in language use (e.g., Bybee, 2001; Langacker, 1999; Mompeán, 2022).

Based on previous research, the present dissertation attempts to describe what correlation exists among the three phenomena mentioned above and phonetic (e.g., stress patterns), linguistic (e.g., lexical status of the word), sociolinguistic (e.g., gender) and usage-based factors (e.g., token frequency). In doing so, this dissertation intends to contribute to the growing empirical body of data by providing an account of linking /r/, t-tapping, and yod coalescence in contemporary RP. The present doctoral dissertation aims to shed light on the influence of the factors on the connected speech processes in actual (conversational) language use (see 2.4.).

3.2. Research questions and hypothesis

The investigation conducted in this dissertation addresses the following general research questions (RQ) to find out what the current status of linking /r/, t-tapping and yod coalescence is in contemporary RP and what factors motivate their variation:

RQ1. How frequent is the use of linking /r/, t-tapping, and yod coalescence in conversational language use?

RQ2. What factors condition variation in the production of linking /r/, t-tapping, and yod coalescence?

Based on the specialised literature, the hypotheses (H) entertained are the following. Regarding RQ1, the hypotheses entertained are as follows:

H1. Each process will be used variably (i.e., non-categorically).

H2. There will be significant differences in their rate of production.

These hypotheses are based on findings from previous empirical studies, which indicate that the processes are not only variable but also that their rate of production differs (e.g., Amos, 2007; Badia Barrera, 2015; Connine, 2004; Glain, 2012, 2014; Mompeán, 2022; Mompeán & Gómez, 2011; Pavlík, 2016). The possible variation found in the use of each phenomenon will be discussed below in relation to the possible variants for each phenomenon alongside more specific hypotheses (e.g., glottalisation, not hiatus, will be more common when /r/-sandhi is not used at potential /r/-sandhi sites).

Regarding RQ2, the main hypotheses entertained imply that patterns of variation will be motivated by a number of linguistic (lexical), sociolinguistic, usage-based, and phonetic factors:

- a) Lexical status (content vs. function) of the words involved at the production site
- b) Speaker gender (male vs. female)
- c) Lexical and collocation frequency
- d) The stress pattern at the production site

The possible effect of these factors on the processes under study is likely based on the specialised literature reviewed above (see Chapter II). For example, high-frequency lexemes and word combinations might lead to higher rates of linking /r/, t-tapping and

yod coalescence, which will point to a lexicalisation of /r/ in high-frequency contexts and the automatised tap productions and yod coalescence in high-frequency contexts (e.g., Bybee, 2001, 2006).

However, the specific impact of each factor in linking /r/, t-tapping and yod coalescence usage is less clear to predict and it will be addressed in each independent study. For instance, in terms of sociolinguistic factors, gender is not expected to be relevant, or at least, not as relevant as other factors, for the production of linking /r/ (e.g., Hannidal, 2006).

In contrast, gender is expected to be relevant for the production of t-tapping and yod coalescence as females tend to produce variants that are more standard and fewer low-status forms than males (e.g., Coates, 2004; Hannidal, 2006). Moreover, even if the same variables have an impact on two or the three phenomena, the specific contribution of each variable may be different. Regarding t-tapping variation, for example, phonetic factors are predicted to have a greater effect than linguistic or sociolinguistic factors.

3.3. Methodology

The following subsections describe the methodology of the studies in this dissertation. The methodology comprises data collection, speaker sampling and profile, data collection and codification as well as statistical methods. The subsections are distributed as follows. The first subsection describes the speakers along with a subsection about the speakers' profile and the second and third subsections describe the data used for the present study and the procedure of collecting and coding of the data. The last subsection contains information on the statistical tests.

3.3.1. Speakers

The current dissertation focuses on the speech of six speakers interviewed at *The Andrew Marr Show* (see 3.3.2. below, henceforth, the corpus of spontaneous speech in the interviews obtained from the *Andrew Marr Show* will be conveniently referred to as AMS corpus). The reason why they were chosen was that the interviewees were mostly political

figures, and given their accent (i.e., RP), they represented a homogeneous group of speakers. For this study, the number of speakers per gender was carefully selected to compile a corpus as homogenous as possible. This led to six speakers, equally chosen by sex; three males and three females, so men and women are equally represented.

For the present dissertation, the selection of informants is based on one linguistic criterion: all of them should be RP speakers whose speech exhibits certain phonetic criteria, such as the following: non-rhoticity, no diphthong shift and monophthong differentiation of /a; æ/ and /ʊ, ʌ/, no g-dropping or h-dropping, except for weak forms, among others (see Wells, 1982, 1994a, see also 3.3.2.).

A relatively homogenous sample

The six speakers selected for the present study are public figures who participated at some point during the history of the TV show *The Andrew Marr Show* (<http://www.bbc.co.uk/programmes/b0080bbs>), a talk show broadcast on BBC One from the mid-2000s to the early 2020s (see description below in the section on data). The speakers represent a relatively homogenous group in terms of social class, education background, as well as profession at the time of data collection.

They were selected under the criterion of being UK political figures and RP speakers. On the other hand, they are (or were at the time of conducting the study) all politicians and members of the House of Commons in the UK, so they have (or had) the same social status at around the same period.

The speakers were identified as RP speakers according to classical descriptions of the accent (e.g., Gimson, 1962/1980; Wells, 1982). To ensure homogeneity in the sample, as mentioned earlier, certain criteria were applied: all the informants were native speakers of British English and showed no regional features in their speech; their accent is non-rhotic, the diphthongs /eɪ/, /aɪ/, /aʊ/ are not ‘Cockneyfied’, their speech presents foot-strut split (e.g., Wells, 1982, 1994a).

The interviewer and TV presenter Andrew Marr, who is also an RP speaker, was not included as a potential subject for the investigation because he did not represent the criterion of being a UK political figure. While it is likely that politicians (or any other

speaker) are relatively speech-conscious at the time of public appearances, no professional BBC staff such as Andrew Marr were considered to analyse the connected speech processes under investigation in conversational settings.

The individual speakers

Biographical information about the speakers may be useful as representatives of the linguistic community of speakers under investigation. Even though it is sometimes difficult to find reliable sources concerning information about the speakers' profiles and occasionally bibliographical information about them is limited, several websites have been used to present truthful facts regarding their geographical background and education.

The short bibliographies presented in this section enable to account for the speakers' backgrounds. Biographic data were obtained from the UK Government (<https://www.gov.uk/>), the Encyclopaedia Britannica (<https://www.britannica.com/>) and the UK Politics (<https://www.politics.co.uk/>).

The speakers are George Osborne, MP former Chancellor of the Exchequer (retrieved from https://commons.wikimedia.org/wiki/File:George_Osborne_0437.jpg); Theresa May, MP former Prime Minister and Home Secretary (retrieved from https://commons.wikimedia.org/wiki/File:Theresa_May_in_Tallin_crop.jpg); Ed Balls, MP former Shadow Chancellor (https://commons.wikimedia.org/wiki/File:Ed_Balls.jpg); Yvette Cooper, MP former Chief Secretary to the Treasury and Work and Pensions Secretary who is serving as Shadow Home Secretary (https://commons.wikimedia.org/wiki/File:Official_portrait_of_Rt_Hon_Yvette_Cooper_MP_crop_2.jpg); David Cameron, MP former Prime Minister (https://commons.wikimedia.org/wiki/File:Prime_Minister_David_Cameron_-_official_photograph_%288947770804%29_%28cropped%29.jpg), and Harriet Harman, MP former Deputy Labour Leader (https://commons.wikimedia.org/wiki/File:Official_portrait_of_Rt_Hon_Harriet_Harman_QC_MP_crop_2.jpg) –see Table 1 below, as for the speakers' images, they were obtained using Creative Common licensed images.

They were selected under the criterion of being UK political figures and RP speakers. They were identified as RP speakers according to classical descriptions of the accent (e.g., Gimson, 1962/1980; Wells, 1982). Andrew Marr, the TV presenter, has not been included as a potential subject for the investigation because he did not represent the criterion of being a UK political figure but a BBC employer or representative.

At the time of the analysis, the corpus compiled for the present dissertation spanned a period of seven years of study (i.e., from 2010 to 2017) and contained 89 interviews. Table 1 below presents a summary of the information of the speakers analysed in this dissertation such as their gender (M/F), year of birth (YoB) and the speaker's age at the time of production (SaP).

Table 1. Information about speakers: gender, year of birth (YoB) and speaker's age at the time of production (SaP).

	Speaker	Gender	YoB	SaP (years old)
	George Osborne	M	1971	39-43
	Theresa May	F	1956	54-61
	Ed Balls	M	1967	43-47
	Yvette Cooper	F	1968	42-48
	David Cameron	M	1966	44-45
	Harriet Harman	F	1950	60-67

George Osborne was born in London on the 23rd of May 1971 and was educated at London public schools and Oxford University. He is a former British Conservative Party politician who works now as a newspaper editor. He entered No 11 Downing Street after the 2010 general election as Chancellor of the Exchequer and on the 8th May 2015, he was appointed First Secretary of State. He served as Chancellor of the Exchequer in David Cameron's Cabinet from May 2010 until July 2016. He served as the First Secretary of State from May 2015 until July 2016.

Theresa May was born on the 1st of October 1956 in Sussex and grew up in Oxfordshire. She is a British politician who started her career at the Bank of England. She was educated at both grammar and comprehensive schools. She attended the University of Oxford. She served as Home Secretary from May 2010 until July 2016 as part of the Coalition Government after the 2010 General Election. She served as Prime Minister from 13th July 2016 until 24th July 2019. In 2002, she became the first woman to chair the Conservative Party and in 2016, she became the second woman Prime Minister of the United Kingdom in British history after David Cameron.

Ed Balls was born on the 25th of February 1967. He is a British politician. He attended schools in Norwich and Nottingham and studied at Oxford University. After spending a year as a Kennedy scholar at Harvard University, he worked for the Financial Times newspaper. He served as Shadow Chancellor for the Labour Party from 2011, after Alan Johnson's resignation, until 2015 when he lost his seat in Parliament. He left politics, became a visiting professor at King's College London, and held a research fellowship at Harvard Kennedy School.

Yvette Cooper is a British politician. She was born in Inverness on 20th March 1968, and she attended Oxford, Harvard and The London School of Economics. She has been a member of the Labour Party since she was a teenager. She served as former Chief Secretary to the Treasury and Work and Pensions Secretary, and she is serving as Shadow Home Secretary.

David Cameron was born on 9th October 1966 in London. He is a British politician who served as Leader of the Conservative Party in December 2005, after working in business and government. He studied at Eton College and attended Oxford University. He became Prime Minister after the 2010 general election when he started a Coalition Government with the Liberal Democrats, as Britain's first Coalition Government since World War II. He served as Prime Minister from May 2010 until July 2016.

Harriet Harman was born on 30th July 1950. She is a British politician who served in the Labour Party and set up the first Parliamentary Labour Party Women's Group. She attended York University and worked as a solicitor at Brent Law Centre in 1974. She became the Leader of the Opposition in 2015.

Despite the fact that Harriet Harman is the oldest speaker, all the participants can be grouped in the same age group (i.e., middle-aged speakers) which is characterised for

being linguistically fairly unchanging (see Chambers, 2002; Hannisdal, 2006; Hudson, 1996). Under the criterion of being UK political figures and having non-regional features, Yvette Cooper, who is the only informant of Scottish origin, was not excluded from the analysis because she does not display rhoticity or any other Scottish phonetic features. In any case, all speakers have been classified as RP speakers in the present dissertation as all of them have in common that their accents lack obvious regional characteristics. In this respect, the non-localizability of their accent represents present-day RP, which is heard from younger groups of speakers of an educated middle-class origin whose pronunciation cannot be associated with a precise area (e.g., Collins & Mees, 2003; Mompeán & Mompeán-Guillamón, 2009).

Taking into consideration the facts stated earlier, the informants of the present dissertation constitute a socially homogenous community of speakers. They have similar background and education, their professional occupation is the same, and they share the same social and linguistic status.

3.3.2. Data

For this dissertation, one essential task was to find and gather large amounts of sound recordings in which a consistent conversational environment could be observed. In doing so, several criteria were applied to the data search: a conversational environment in which speakers provided continuous stretches of speech in a context as spontaneous as possible; the speakers had to be RP speakers -even though a certain amount of inter-speaker variation could be found, no regional features had to be found in their speech; the name of the political figure had to be identifiable; furthermore, in order to avoid potential bias in the conversational context selection a delimiting criterion was applied: the speech context had to be as similar as possible as to maintain consistency in the sample.

This resulted in the selection of a chat show as the source of data given its availability on the Internet as well as being characterised by a systematic interview format. The data were extracted from *The Andrew Marr Show* (<http://www.bbc.co.uk/programmes/b0080bbs>), a talk show broadcast on BBC One from 2005 to 2021, a type of TV programme that offered uninterrupted intervals of speech in the shape of a dialogue between the interviewer and the interviewees. The show was a

one-hour-long British television programme that was broadcast on Sunday mornings. British journalist Andrew Marr (1959-) presented the TV programme -although other TV presenters replaced him on a few occasions. He interviewed political figures about political and social events, such as immigration, public finances, UK politics, legislation, etc. Generally, the guests featured in this programme each week were one senior member of the UK Cabinet, a representative member of the Opposition, a famous non-political guest and, sometimes, other celebrities or well-known journalists, cultural icons or the UK's most influential commentators. This has allowed to minimise the impact of social variation by using data from speakers who all belonged to the same social status and/or profession.

Furthermore, the setting in which the interviews took place was always the same programme set and the interviews were always conducted in the same turn-taking method that has enabled to maintain consistency for data sampling. Andrew Marr and each interviewee are sitting comfortably in a chair in the studio facing each other. This setting allows both speakers to use gestures. Figures 10 and 11 show an example of the studio setting where the interviews took place and the interaction between Andrew Marr and two of the six speakers chosen for this study on two different occasions, George Osborne (on 02nd September 2012) and Theresa May (on 30th April 2017) - Andrew Marr with George Osborne (top) and Theresa May (bottom) at *The Andrew Marr Show*.



Figure 10. Andrew Marr with George Osborne at *The Andrew Marr Show*. Source. Creative Commons (<https://www.bbc.co.uk/programmes/b01mh0ds>).



Figure 11. Andrew Marr with Theresa May at *The Andrew Marr Show*. Source: Creative Commons (<https://www.bbc.co.uk/programmes/b08plldm>).

In addition, researching speech through mass media offers several advantages, the most obvious of which is accessibility, as this type of data is easy to access and, to some extent,

easy to collect as well (Hannisdal, 2006). Another major advantage relies on the fact that using TV audio material is, in general, of a high technical quality and, thus, its recording conditions typically reduce the background noise that normally appears in other spontaneous settings such as street interviews. This high technical quality may facilitate subsequent auditory and acoustic analyses.

Another benefit of collecting data from this type of source is that informants provide continuous stretches of speech. Nevertheless, significant disadvantages can be found when compiling a balanced corpus. Some of the major drawbacks are the inconsistency of the speakers' appearances and the irregularity in the duration of the interviews since not all of the interviews are equally long and not all the speakers are interviewed the same amount of times.

While *The Andrew Marr Show* was broadcast from 2005 to 2021, the data for the current study were only considered from 2010 on. This is because the transcripts of the interviews began to be accessible from 2010 on, which ruled out some data. In addition to this, neither all the video files nor all interview transcripts -before 2010- were accessible to be collected for the investigation.

The present study, therefore, examines the earliest interviews that were accessible at the time of conducting the research (2016) until the year when the data collection was completed (2017). This yielded a total amount of 89 interviews that were analysed corresponding to the years from 2010 to 2017, corresponding to the earliest and most recent years of production (YoP) in the corpus from the *Andrew Marr Show*.

The total number of words and duration of the interviews were calculated for the whole corpus using word processing and audio software programs. The shortest interview lasts 01:40 minute (it is a sofa chat interview: a more relaxed yet structured interview that was sometimes conducted on the programme) while the longest interview lasts 29:06 minutes. The shortest transcript amounts to 397 words whereas the longest transcript amounts to 6,065 words. The 89 interviews amount to ca. 240,000 words and 18 hours, 45 minutes, and 33 seconds of audiovisual material.

By isolating the participants' speech from the interviewer's speech (and other interviewers' speech when Andrew Marr did not present the programme), the analysed material amounts to ca. 163,400 words. The total word count excludes interview titles, headings, speakers' names repeated as an indication of their dialogues, and other written-

down specifications that are not part of the dialogue, such as *over, Marr tries to interject, On Sunday 28 February Andrew Marr interviewed [...], interview ends*. Appendix I includes all the interviews analysed in the current study.

Table 2 summarises the information about the data collected per speaker and the corpus as a whole including their cataloguing in terms of gender (M1/F1) for subsequent, simpler quantification of the data, the year of production (YoP) of the interviews, any non-available interview (NA) (i.e., either the transcript or the audio file were accessible at the time of conducting the analysis or the speaker was not interviewed that year), the number of interviews (NoI) which were analysed for each speaker, and the total amount of words (W) per speaker in the corpus.

Table 2. Information about speakers: gender, year of production (YoP), non-available interviews (NA), number of interviews (NoI) which were analysed for each speaker, and the total amount of words (W) per speaker.

Speaker	Gender	YoP	NA	NoI	W
George Osborne	M1	2010-2014		15	29,061
Theresa May	F1	2010-2017	2011	16	36,735
Ed Balls	M2	2010-2014		18	29,129
Yvette Cooper	F2	2010-2016		14	17,335
David Cameron	M3	2010-2011		7	26,728
Harriet Harman	F3	2010-2017	2016	19	22,244

To achieve a sizeable amount of data, an (arbitrary) minimum of 300 tokens per phenomenon and speaker was established. However, not all the speakers had the same chance to produce the same number of tokens because they were not given the same

amount of interview time and/or were not interviewed the same amount of times. Given this, the 300-token threshold was not reached in certain cases given the different phonetic environments required for the different connected speech phenomena to occur, which makes some phenomena more frequent than others. The most notable example is yod coalescence, represented in this sample with a different number of tokens, in contrast to linking /r/ and t-tapping. In those cases in which the 300-token threshold was not reached, the threshold of tokens was established using the number of tokens collected from the speaker whose threshold corresponded to the lowest of all, which enabled to collect the same threshold of tokens for the rest of the sample.

The sample, therefore, consists of 1,800 potential cases for linking /r/ and t-tapping, and 960 potential cases for yod coalescence. The total amount of potential cases analysed is 4,560. Table 3 below summarises the total amount of tokens per phenomenon and speaker.

Table 3. Information about speakers: the total amount of tokens per phenomenon and speaker.

Speaker	Linking /r/	T-tapping	Yod coalescence
George Osborne	300	300	160
Theresa May	300	300	160
Ed Balls	300	300	160
Yvette Cooper	300	300	160
David Cameron	300	300	160
Harriet Harman	300	300	160
All speakers	1,800	1,800	960

3.3.3. Procedure

This section describes how the potential cases of the phenomena under investigation were obtained from the data, how the data were coded, and how they were analysed both auditorily and acoustically. The same procedure was followed for each informant and interview.

Identification of potential cases

A preliminary step to the analysis of potential contexts for linking /r/, t-tapping and yod coalescence was to find the potential cases in the interview data described above. The procedure for the data collection involved several steps. For each interview, its transcription was downloaded from the programme website as a PDF document, then the PDF document was transformed into a Word document. The transcripts were further checked manually against the recordings until a final version of the transcript was obtained. Once the final version was available, typical orthographic contexts were explored in the transcript of the interviews through word search tools.

Regarding linking /r/, the searches for potential contexts across word boundaries concerned the combination of the letters <r> and <re> followed by vowels and the letter <h>, with a space in-between to check word boundaries. That is, the combinations contained the following searches: <r/re> followed by <a/e/i/o/u/h>. Vowel letters typically stand for a vowel phoneme so they were an obvious choice. The letter <h> was also considered since some word-initial <h> in English are silent letters (e.g., *hour*, *honour*, etc.) and, above all, because many function (or grammar) words (e.g., *have*, *has*, *had*, *him*, *her*, etc.) are often pronounced with a weak form involving the elision of /h/. Additionally, acronyms, such as MP (e.g., *labour MP*) and the number ‘8’ were also examined since they begin with a vowel in their phonological makeup.

Regarding t-tapping, the searches for potential contexts across word boundaries comprised the combination of the letters <t>, <tt>, <d> and <dd> preceded and followed by vowels or /h/ for the same reasons as those explained above for linking /r/, with a space in-between to check word boundaries. That is, the combinations contained the following searches: <t/tt/d/dd> followed and preceded by <a/e/i/o/u/h>. Word-initial alveolars (e.g.,

monosyllabic verbs and nouns, such as *tax* and *deal*) have been excluded from the analysis as they are never tapped.

Regarding yod coalescence, the searches for potential contexts across word boundaries comprised the combination of the letters <t(e)>, <d(e)>, <s(e)>, <ss>, <ce>, <z(e)> and <zz> preceded by the letters <y>, <u> and <eu> with a space in-between to check word boundaries. That is, the combinations contained the following searches: <t/te/d/de/s/ss/se/ce/z/zz/ze> followed by <y/u/eu>.

The morphological structure of potential cases has been taken into consideration as well (word-external vs. word-internal). Orthographic cases in which the morpheme boundary is external have been included as potential contexts, while cases of internal morpheme boundaries, such as affixed words (e.g., *roaring*, *sitting*, *factual*), or without morpheme boundaries (e.g., *carry*, *lady*, *Estuary*) have been excluded. However, some tokens where the potential connected speech phenomenon may occur at internal morpheme boundaries, as in the case of compounds, have been exceptionally included as potential cases since there exist distinct degrees of lexicalisation due to the usage of many technical terms in the AMS corpus. This is also linked to the uncertainty of whether such compounds were, in essence, real compounds or a relatively frequent collocation, as variation is commonly found in this type of tokens and phenomena (e.g., linking /r/ exhibits some degree of variability in compounds, in contrast to the production of linking /r/ in affixed words (before suffixes), where it is considered to be categorical).

The search for potential orthographic cases was as inclusive as possible insofar as it took into account phenomena that might be easily skipped in a broader identification of the cases based on the orthographic contexts for potential cases described above. Thus, contractions were searched as well (e.g., *there's/there're*) since they might have been used by the transcribers but not by the interviewee and vice versa, and they can alter the number of potential cases.

As the transcripts had been previously checked manually against the recordings, if discrepancies were found between the written and spoken versions in terms of contractions, they were edited in the corpus. For example, any word or phrase that was different in the audio/video concerning the text transcription. All orthographic searches were done taking into consideration punctuation marks such as dots, commas, etc. paying attention to whether these marks actually stood for silent pauses or, instead, represented

none and could therefore also be included as potential cases (e.g., *year*, *and*; *that*. *And*; *later*, *I*; etc.).

Similarly, all the orthographic searches were done taking into consideration orthographic pauses (e.g., , and .), paying attention to whether they were real phonetic/orthographic pauses or not. Finally, since the data were obtained from live conversations, unintentional repetitions were occasionally used, so contexts that involve unintentional repetition or stuttering of word combinations were also considered if a potential case or cases could be found (e.g., *but I but I*).

Auditory and acoustic analyses

After searching all potential contexts in an interview, an initial auditory analysis was conducted to check whether the speakers' potential cases as determined by the orthographic searches actually corresponded with the required phonological contexts described in the literature. Additionally, while conducting the auditory analysis those cases that constitute potential orthographic contexts but may not be possible phonological tokens were identified as well (i.e., orthographic cases whose pronunciation does not correspond to a potential phonetic case, such as *number one*, *other European*, *that happening*, *that one*, *chancellor has*, etc.). In *number one*, for example, the letter <o> of *one* stands for /wʌ/, starting with an approximant consonant (/w/), so the orthography is misleading in this respect. Similarly, the word *European* begins with approximant /j/, so the digraph <eu> at the beginning of the word also begins with a phonological consonant.

In addition to this, those cases where no punctuation mark was used but there was an actual pause were discarded as well (e.g., *And you*; *for a*; *but you*; *but after*; etc.). A pause was considered to exist where at least 100 ms of silence between segments, determined acoustically (Mompeán, 2022). As a case in point, although the combination *that in* might seem to imply no pause, in an interview with David Cameron (2nd May 2010) the speaker actually used a pause, as exemplified in (76):

(76) [...] I want to explain to people is *that in* making these decisions [...] (00:58)

Similarly, if a punctuation mark was used but implied no pause (e.g., *but, Andrew; And, yes; this. You;* etc.), the potential cases were included in the final set of potential cases to be analysed. By way of example, although the combination *but, if* might seem to imply a pause, in an interview by Yvette Cooper (17th July 2011) the speaker actually used no pause, as exemplified in (77):

(77) [...] she'd be able to do that. *But, if* so, she needs to say so [...] (03:17)

Once all the orthographical and phonological suitable potential contexts had been identified, the final set of potential cases was analysed auditorily for the second time. To conduct the auditory and acoustic (see below) analyses, the interviews had been previously downloaded from the Internet as video (.mp4). The videos of the interviews were converted into audio files (.wav) to conduct the auditory analyses and the video (.mp4) version was only used when visual support was considered useful. The acoustic analyses required an analysis of the audio (.wav) files.

When inspecting the potential cases auditorily, the occurrence of the potential connected speech phenomenon was categorised as occurring or not occurring based on perceptual judgement, leaving the analysis of unclear cases to speech analysis software for disambiguation (Milroy & Gordon, 2003). Auditory analysis was combined with spectrographic analysis as the next step for ambiguous cases using the speech software to determine the presence/absence of the phenomena in unclear cases. Similarly, tokens that were not clear enough for reliable cataloguing, either because the speakers in the interview interrupted each other when talking or due to the background noise, were excluded from the analysis.

As in other studies (e.g., Hannisdal, 2006; Mompeán, 2009), the reason why auditory analysis was used in the first place is primarily because the decision as to whether the connected speech phenomena has been produced is simple in general terms. Furthermore, instrumental analysis of the entire corpus would result in an impractical procedure. Despite decisions being carefully made regarding the presence or absence of instances of the phenomena described earlier, auditory analysis of the data was occasionally considered to be unsatisfactory and spectrographic software was used for

clarification. The acoustic analysis has proved helpful for ambiguous tokens so as to provide a greater level of precision.

As mentioned above, the instrumental analysis procedure involved converting the .mp4 video files that had been downloaded from the Internet into .wav audio files using a free audio converter programme to make them readable for the speech software. The software used was *Praat*, a free software tool for speech analysis developed at the University of Amsterdam (www.fon.hum.uva.nl/praat/) used to conduct spectrographic analysis for ambiguous cases.

As most phonetic phenomena, linking /r/, t-tapping, and yod coalescence display gradient features and variability. As a case in point, the use of glottalisation is common when linking /r/ is not used as opposed to pure hiatus (i.e., sequence of two vowel sounds with normal/modal voice) and in /t/ contexts in the case of t-tapping (i.e., t-glottaling). This use of glottalisation has been described as an alternative hiatus-breaking strategy to that of pure hiatus (Mompeán & Gómez, 2011) for the former, and as a stigmatised counterpart for the latter. However, the use of glottalisation occurs on a continuum from laryngealisation (or creaky voice) to a full glottal closure, with intermediate stages such as creaky-voiced glottal approximants [ʔ] (Mompeán, 2022). In defining glottalisation acoustically, the region in the speech signal is characterised by irregular spaced pulses, which are often accompanied by other parameters, such as stop gap, low or even absence of F0, breathiness, low amplitude, or even a combination of these (Redi & Shattuck-Hufnagel, 2001).

Despite a certain degree of variation is found, the connected speech phenomena under investigation have been analysed as binary variables (i.e., they are either used or not). Following this methodology, a categorical decision was adopted as to the occurrence or absence of the phenomena. However, when the auditory analysis was insufficient, acoustic cues were inspected on the corresponding spectrograms. For example, one relevant feature that reveals whether the speaker has produced the r-link is the acoustic characteristic of the phoneme /r/. Rhotic approximants (e.g., a postalveolar approximant) are identified on a spectrogram with the presence of a low F3 (below 2000 Hz) that may be related to its retroflex articulation, and a low F2, showing a curly-shaped fall of energy (e.g., Pavlík, 2016; Mompeán, 2022).

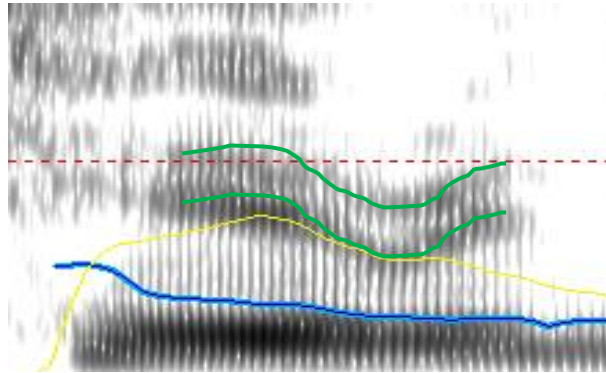


Figure 12. An example of linking /r/ in *here is* by Theresa May (31st October 2010). Low F2 and F3 in green, intensity in yellow and pitch in blue.

Figure 12 shows an example of linking /r/ in the heterosyllabic context *here is* from the AMS corpus by Theresa May in an interview that was on 31st October 2010. The criteria as to the presence of linking /r/ production, as previously mentioned, lowered F2 and F3 (in green), relatively stable pitch (in solid blue) as well as intensity contours (in solid yellow) are illustrated.

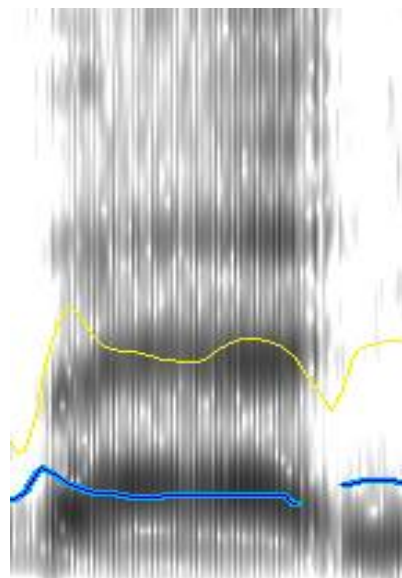


Figure 13. An example of the absence of linking /r/ in *there and* by Harriet Harman (31st July 2011). No formant transition, intensity in yellow and pitch in blue.

Figure 13 shows an example of the absence of linking /r/ production in *there and* from the AMS corpus by Harriet Harman in an interview that was on 31st July 2011. The criteria as to the absence of linking /r/ production is reflected by the lack of lowered F2 and F3 formants.

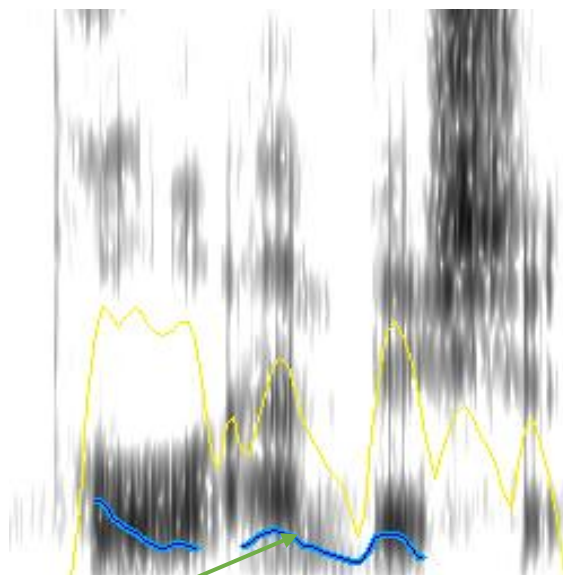


Figure 14. An example of laryngeal strategy (i.e., creaky voice) in *poor ambition* by George Osborne (31st January 2010). Laryngeal strategy indicated by green arrow, intensity in yellow, and pitch in blue.

As for the criteria to determine the presence or absence of laryngeal strategies such as creaky voice (Figure 14 above), certain acoustic cues were determinant: decreased acoustic intensity, lowered F0, and irregular spaced pulses (see Mompeán & Gómez, 2011; Mompeán, 2022).

Taps are reflected on the spectrogram as a relatively short production that generally lacks a release burst. Acoustically, taps are observed to be acoustically short, varying in duration from 10 to 40ms, with little or no sign of a release burst (Shockey, 2003; Zue and Laferriere, 1979). The main difference between the production of taps and stops /t, d/ is measurable according to their duration. The acoustic characteristics of stops are visually clear, with /t/ having a longer VOT than /d/ (Zue & Laferriere, 1979). Depending on the tongue tip movement, taps will result in a variety of acoustic realisations. The closure could be complete or partial. If there is partial closure, turbulence will sometimes be generated at the constriction giving the appearance of a voiced

fricative. (Zue & Laferriere, 1979). An average tap is 26 ms, while plosives are longer since an average [d] is 75 ms long and an average [t] is 129ms long (Pándi, 2014). A tap-like pattern, according to Stone and Hamlet (1982), displays a momentarily reduction in the intensity of both the preceding and following vowels, which is occasionally accompanied by a small burst. A categorical tap is normally recognised on the spectrogram because of the presence of a voiced bar but without formants, with a burst accompanied by a large drop in intensity (Warner & Tucker, 2011).

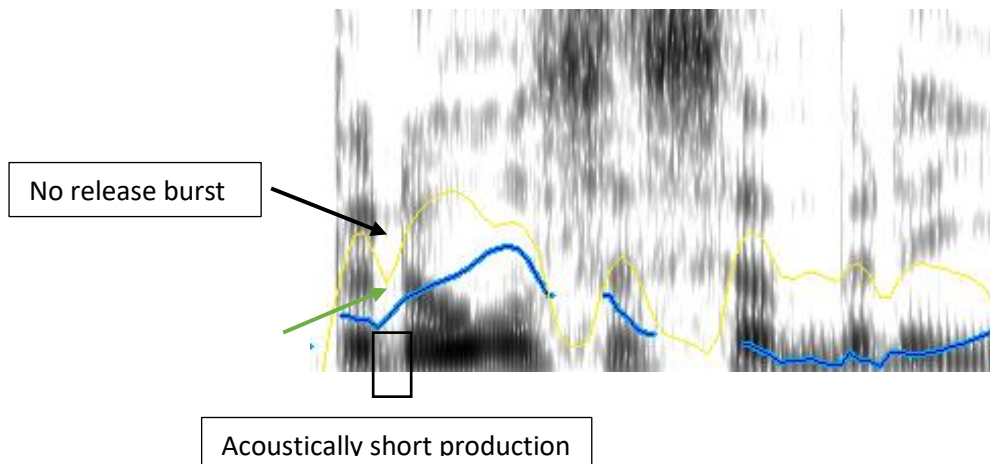


Figure 15. An example of /ɾ/ in *but also* by Ed Balls (20th June 2010). Tap region indicated by green arrow, intensity in yellow and pitch in blue.

Figure 15 shows an example of t-tapping production in *but also* from the AMS corpus by Ed Balls in an interview that was on 20th June 2010. The criterion as to the presence of a tap production on a spectrogram is reflected by their relatively short duration, with no release burst.

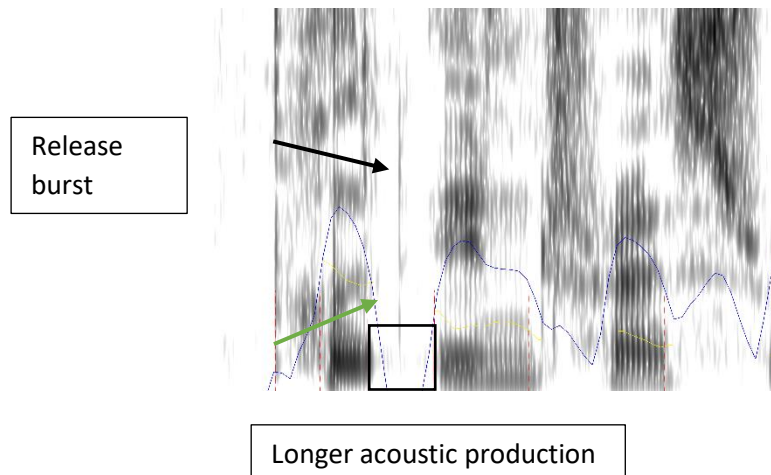


Figure 16. An example of /t/ in *out of* by George Osborne (31st January 2010). A voiceless plosive is indicated by a green arrow, intensity in yellow and pitch in blue.

Figure 16 shows an example of /t/ production in *out of* from the AMS corpus by George Osborne in an interview that was on 31st January 2010. The criteria as to the presence of the voiceless plosive on a spectrogram is reflected by a longer acoustic duration, with a release burst.

Even though there also exists a gradient realisation in /t/ contexts from the standard end of the continuum (i.e., /t/) to the non-standard end of the same continuum which was analysed (i.e., /ɾ/), it was also possible to identify another distinctive variant (i.e., /ʔ/). While the occurrence of the voiceless plosive and the tap variant was the main focus of the analysis, the subsequent acoustic analysis identified the occurrence of t-glottaling when taps were not applied.

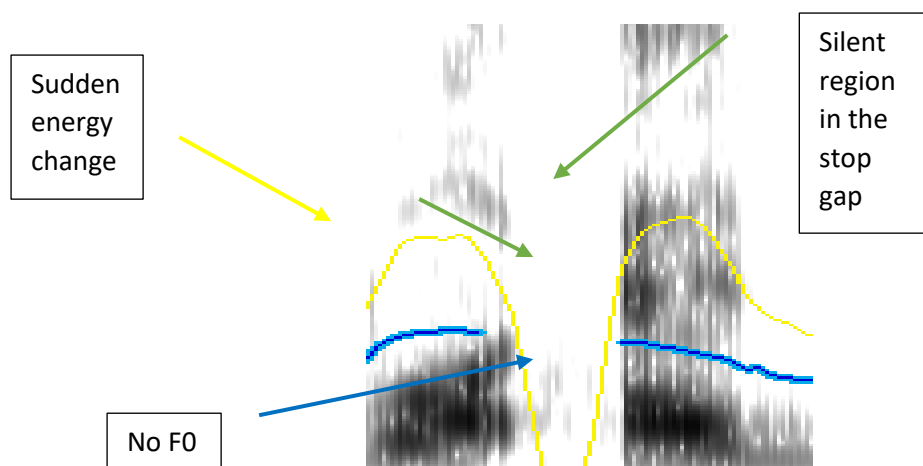


Figure 17. An example of [ʔ] in *what Ed* by Harriet Harman (7th July 2013). The glottal stop is indicated by a green arrow, intensity in yellow and pitch in blue.

Figure 17 shows an example of glottal stop production in *what Ed* from the AMS corpus by Harriet Harman in an interview that was on 7th July 2013. The criteria as to the presence of the glottal stop was made following certain acoustic cues, such as the presence of a stop gap or silent region on the spectrogram, absence of F0 and sudden energy change (see Mompeán & Gómez, 2011).

Regarding yod coalescence, it is worth taking into consideration that in terms of articulation, the closure phase of the affricates /tʃ/ and /dʒ/ is similar to the closure of a plosive (/t/ and /d/, respectively) but the friction that is produced after the release of the closure last longer than the noise burst that marks the release of a plosive and, in the case of the voiceless plosive /t/, any aspiration that may be present after the burst. In contrast to stops, the closure for the English affricates [tʃ, dʒ] is made well back on the alveolar ridge, and the lower cut-off point of energy is located at about 1800-2000 Hz (Fry, 1979). In English, while the place of articulation of affricates is post-alveolar, affricates are characterised by the combination of two manners of articulation: they start as plosives (in /t, d/ contexts), or alveolars (in /s, z/ contexts), and end as fricatives (i.e. [ʃ, ʒ]), which are produced at a single place of articulation with a slow release resulting in distinct fricative noise. This single place of articulation is reflected in the acoustic cues on a spectrogram (Beňuš, 2021).

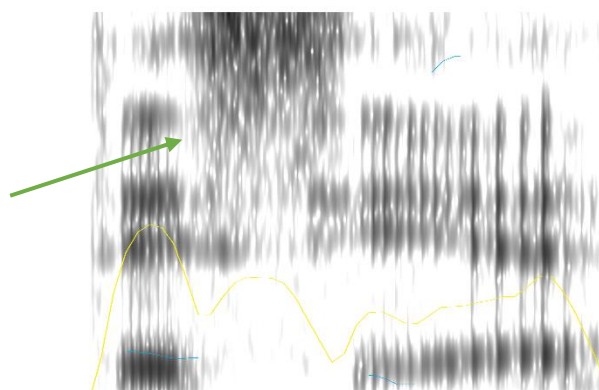


Figure 18. An example of /j/ in *this year* by George Osborne (31st January 2010).
Affricate indicated by green arrow, intensity in yellow.

Figure 18 shows an example of yod coalescence of /s/ followed by /j/ in *this year* from the AMS corpus by George Osborne in an interview that was on 31st January 2010. The criterion as to the presence of yod coalescence on a spectrogram is reflected by a slow release resulting in fricative noise or turbulence.

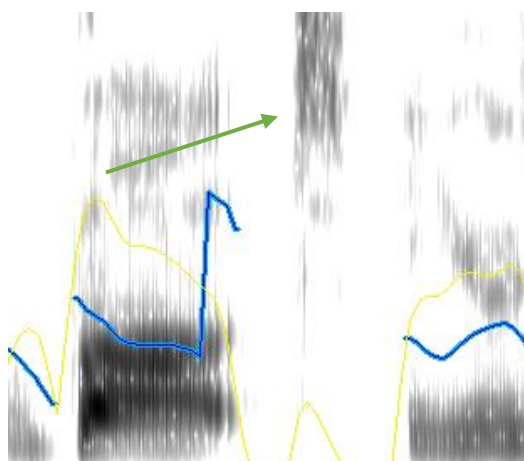


Figure 19. An example of /t/ followed by /j/ in *but you* by Theresa May (24th April 2016).
Plosive release and aspiration are indicated by a green arrow, with intensity in yellow.

Figure 19 shows an example where yod coalescence does not apply in *but you* from the AMS corpus by Theresa May in an interview that was on 24th April 2016. The criteria as to the absence of yod coalescence on a spectrogram is reflected by the release burst typical of alveolar plosives.

For those cases in which linking /r/ or t-tapping is not applied, the production of glottalisation has been identified. Even though different realisations within the glottalisation continuum can be found, the varying degrees of glottalisation have not been differentiated in the present study. Regardless, typical acoustic cues for glottal stops are the following: stop gap, absence of F0, and sudden energy change. In terms of acoustic cues, the presence of a full glottal closure (as in a glottal stop), for example, involved a stop gap, the absence of F0, and a sudden energy change while the presence of creaky voice implied decreased acoustic energy, irregularly spaced pulses and a lowered F0 (Mompeán & Gómez, 2011).

Variables and data coding

Once the final set of potential cases was obtained, these cases were coded for several variables. The decision was made to examine a relatively limited number of variables yet comprehensive in their scope, including phonetic (stress pattern), linguistic (word class and word class combination) and usage-based (lexical token frequency, collocation frequency) variables. Sociolinguistic factors, such as gender or socioeconomic status, were controlled in this study so they are not discussed at this point.

Phonetic factors

The most notable phonetic factor that seems to play a role in all three connected speech processes under study relates to the stress pattern of the syllables flanking the connected speech site, so this factor was the one analysed in the current study.

Stress pattern. The stress pattern created between the syllables at both sides of the potential connected speech process sites was coded as a binary option: stressed/strong (S) vs. non-stressed or unstressed/weak (U) for the syllable before and after each side. This

yields four possible combinations across the potential contexts: a) stressed-stressed (SS), b) stressed-unstressed (SU), c) unstressed-unstressed (UU); and d) unstressed-stressed (US).

The stressed or unstressed feature of the syllables analysed was coded as determined by the specific contextual prominence of the syllable in its intonation group. In most cases, lexical and rhythmic stress coincided but when they did not, the actual prominence of the syllable was the one coded for the analysis.

Linguistic factors

Two linguistic variables were analysed in relation to the words at both sides of each connected speech process site: word class and word class combinations.

Word class. Each lexical item at either side of the potential site for a given connected speech phenomenon was classified as either a content (C) or function (F) word. Content words are those that possess semantic content and contribute to the meaning of the sentence in which they occur, typically nouns, main verbs, adjectives, and most adverbs. Conversely, function (or grammar) words have little lexical meaning and express grammatical relationships among other words and they are typically prepositions, pronouns, auxiliary verbs, conjunctions, articles, particles, etc.

The content vs. function word status of items flanking the potential contexts was determined given the specific contextual function of each item. For instance, the verb *to be* can function as a content item (i.e., *to be* functions as main verb) or as a function word (i.e., *to be* functions as auxiliary verb). Another example is *there*, which can function as a content word (i.e., *there* functions as an adverb) or as a function word (i.e., *there* functions as the existential pronoun as in *there is/are*). For this study, most nouns, verbs, adjectives and adverbs were classified as content words, whereas prepositions, conjunctions, pronouns, determiners and auxiliary verbs were considered function words.

Word class combinations. Word class combinations were coded as containing a binary option as well: content (C) and function (F) types of words, in both the left-hand and right-hand lexical item of the word combination that constitutes a potential case. This yields four possible combinations across the potential contexts for the three connected

speech phenomena under investigation: a) content-content (CC), b) content-function (CF), c) function-function (FF); and d) function-content (FC). As is the case of individual words and their word class described above, the specific word class combination was determined given the specific contextual function of each item.

Usage-based factors

Frequencies of the use of the items at potential connected speech process sites were obtained with the help of online corpora software *Sketch Engine* (www.sketchengine.eu/). Apart from frequency data from the corpus compiled for this dissertation, frequencies were also obtained from the British National Corpus (BNC 2014 Spoken corpus), which contains almost 10.5 million words. The tools available from the online corpora software made it possible to compare the data of the corpus compiled against the data in the BNC 2014 Spoken corpus. For this study, two usage-based, frequency-based variables were coded: lexical token frequency and collocation frequency.

Lexical token frequency. Token frequency, as described by Bybee (2001), refers to the frequency of occurrence of a unit of language in running text. This was calculated for both the potential site of the given phenomenon and the following or preceding word using the median. The lexical frequency of the lexical tokens was also calculated for the BNC Spoken corpus. Use was made of the online corpora software *Sketch Engine*.

Collocation frequency. This was calculated for both words constituting the word collocation using the median. The collocational frequency was also calculated for the BNC Spoken corpus. For this study, only the collocation frequency was measured for bigrams. The collocation frequency of trigrams or larger n-grams was not analysed.

Even though frequency is coded differently across studies, for example, using a binary criterion (e.g., Hay & MacLagan, 2010) or transforming raw frequencies into other measures (e.g., Piercy, 2012), other studies used the median, the mean, and the maximum, which is used to indicate the percentage of tokens that are above the midpoint (e.g., File-Muriel, 2010). Some studies (e.g., Bybee, 2002) use the median as a cut-off point because it allows to split the data into two: the data above the midpoint of the distribution (i.e., to be considered as frequent or more frequent) and the data below the midpoint of the distribution (i.e., to be considered as infrequent or less frequent). In this study, for lexical

token frequency and word collocation frequencies, the median or middle value was used. Items above the midpoint of the distribution were coded as having a higher frequency or frequent while those items below the midpoint of the distribution were coded as having a lower frequency or infrequent.

3.3.4. Statistical analysis

The main goal of a corpus-based study is to find relations between the variables under investigation. Graphs and contingency tables are sometimes insufficient to describe effectively how linguistic phenomena perform. In this respect, statistical tests provide procedures to quantitatively assess and interpret such relations (Cantos-Gómez, 2013).

The quantified data for the three connected speech phenomena under investigation were statistically analysed and the findings were tested for significance. In doing so, the statistics programme used was *R* version 4.1.2. (*R Core Team 2021*) to test the data against the linguistic constraints under study when appropriate. The data was computed by grouping the tokens according to the phenomena, speakers and/or factors, when necessary, to test the different research questions.

Measures for central tendency, such as the median, indicate the representative behaviour of a given data set. In any given set of scores, the median indicates the midpoint of the distribution. For instance, when computing the number of tokens and word collocations, the median or central score of the distribution divides the data into two equal subgroups to determine the more and less frequent items (see 3.3.3.). Despite the mean being the preferred measure for central tendency in most cases, the median is used in the present study because the mean is not reliable with extreme scores (Cantos-Gómez, 2013).

The data were mostly tested using the Pearson chi-square test, also called X^2 , as it is the most common test in linguistic research. The chi-square test is a non-parametric method that allows to investigate the independence or interdependence of the distribution. While this test may reveal distributional patterns, it might also skew the results in cases where speakers' contributions are unequal; for this reason, the number of tokens and type of speech are controlled in the present study (Cantos-Gómez, 2013; Levon, 2010). The test compares the observed frequency with the expected frequency if the factors had no

effect on its distribution. It is a suitable procedure to compare two variables within the same sample, as long as the variables are categorical (e.g., the presence vs. absence of linking /r/ regarding gender) (Cantos-Gómez, 2013; Levon, 2010). For instance, to test whether the data report statistically significant results on the frequency of production of the three connected speech phenomena under investigation, a chi-square test was applied, to be precise, a three-sample test for equality of proportions without continuity correction.

Nevertheless, it is only advisable to use this test if the numbers in the contingency table are not small, so the expected frequency in any cell is not below five; otherwise, the test will not be reliable (Butler, 1985). Whenever the chi-square test could not be applied because the expected frequency in one or some of the cells is below five, Fisher's exact test for count data, or simply *p*, was applied.

The Pearson correlation coefficient, often abbreviated as *r*, is a type of measure of the degree of linear relationships between variables. A Pearson's coefficient, which is based on the chi-square test, is also applied to find the degree of association between variables. It takes on any value between plus and minus one (i.e., $-1 \leq r \leq +1$). If the value of one variable increases, the value of the other increases too, and vice versa (Cantos-Gómez, 2013). Cramer's V test, or simply *V*, measures the strength of association between two nominal categorical variables in a contingency table and it is normally used to confirm the degree of the association according to Pearson's coefficient (Levshina, 2015).

Correlation matrixes are used to represent and compare succinctly sets of linguistic data (e.g., word class combinations and stress patterns). A mere glance at them allows to notice the dependency between the variables. A correlation matrix is made by calculating multiple Pearson's coefficients for different variables. It assesses the degree of influence that the linguistic dependent variables under investigation have on the production of connected speech phenomena. The matrix reports lineal associations among variables between -1 and +1, in which -1 indicates a negative correlation among variables, whereas the closer the correlation to +1, the stronger the association between the two variables is (Sheskin, 2020).

CHAPTER IV: Results and discussion

The present chapter deals with the results and discussion of the studies presented in Chapter III on the three connected speech phenomena under investigation (linking /r/, t-tapping and yod coalescence), which were auditorily and acoustically analysed and examined from a quantitative perspective. As described in Chapter III, the potential contexts where those phenomena could occur in a corpus of TV interviews were identified and coded for certain variables (i.e., AMS corpus), the actual realisation of each potential case noted, and statistical tests conducted to answer the research questions formulated in this dissertation study (see 3.2.).

This fourth chapter is divided into two sections: the results (4.1.) and the discussion of the findings (4.2.). In the first part of the chapter, the first three sections describe the results of each connected speech phenomenon studied. This is followed by a presentation of the general results of the three studies. In the second half of the chapter, the discussion of the results is presented. The discussion is divided into three sections concerning the findings of the three phenomena under investigation. The results of the different phenomena are discussed in terms of the factors which were coded and subsequently analysed (i.e., linguistic, sociolinguistic, usage-based and phonetic factors). These sections are followed by a general discussion concerning the three phenomena.

4.1. Results

This section presents the results of each study dealing with the specific phenomena as well as the general results obtained from the three studies. The research questions (3.2.) are also reproduced here for convenience:

RQ1. How frequent is the use of linking /r/, t-tapping, and yod coalescence in conversational language use?

RQ2. What factors condition variation in the production of linking /r/, t-tapping, and yod coalescence?

4.1.1. Study 1: linking /r/

Regarding RQ1, the results indicate that linking /r/ is a relatively frequent connected speech phenomenon in the spontaneous speech analysed in this dissertation. The results obtained for linking /r/ show that 1,320 cases were produced out of 1,800 potential cases. This corresponds to 73.3% of all cases while 480 of the potential cases that were not produced correspond to 26.7% (see Figure 20 below).

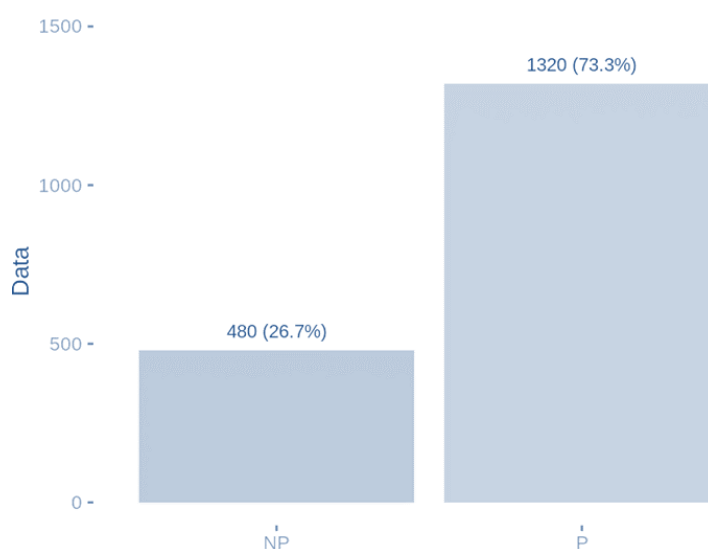


Figure 20. Bar plot by cases produced and not produced for linking /r/.

While the presence or absence of linking /r/ was considered to be categorical, the analyses also revealed aspects of how hiatus is resolved in the potential cases, namely the use of a strategy such as laryngelisation. Thus, the production of laryngeal strategies was also analysed to determine whether speakers tend to break the hiatus by using laryngeal strategies (i.e., glottal stops, creaky voice) and how frequent not breaking the hiatus in connected speech is. The results indicate that, while linking /r/ was used in 73.3% (n = 1,320) of all cases, a laryngeal strategy (glottal stop [ʔ] or creaky voice) was used in 13.8% (n = 248) and zero linking (hiatus) 12.9% (n = 232), as shown in Figure 21.

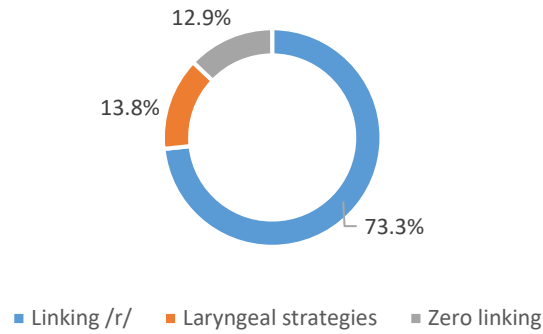


Figure 21. Linking strategies.

Turning now to RQ2, the main hypotheses entertained, as previously described (see Chapter III), imply that patterns of variation will be motivated by several linguistic, sociolinguistic, usage-based, and phonetic factors.

Sociolinguistic factors

Concerning sociolinguistic factors, the ratio of production of linking /r/ was divided by speaker and gender. All the speakers have a production above 150 cases, which constitutes a percentage of production higher than 50% of the overall potential cases per speaker. Three speakers (GO, EB, DC) have a production of linking /r/ between 55% and 69%, while the other three speakers (TM, YC, HH) have a production of linking /r/ above 80%. The average percentage is 73.3%, which indicates that only three speakers are above the mean (see Table 4 below).

Table 4. Cases produced and not produced divided by speakers for linking /r/.

Speaker	Cases produced	%	Cases non produced	%
George Osborne - Male 1	200	66.7	100	33.3
Theresa May - Female 1	244	81.3	56	18.7
Ed Balls – Male 2	177	59.0	123	41.0
Yvette Cooper - Female 2	259	86.3	41	13.7
David Cameron - Male 3	196	65.3	104	34.7
Harriet Harman - Female 3	244	81.3	56	18.7

The results show that males produced 573 cases out of 900 potential cases, which corresponds to 63.7%; while women's production resulted in 747 cases out of 900 potential cases, which corresponds to 83% (see Figure 22 below).

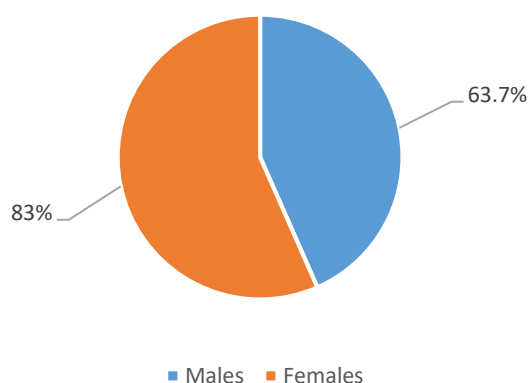


Figure 22. Linking /r/ production by gender.

With regard to one of the main hypotheses entertained for the second research question, i.e., that the production of linking /r/ would be motivated by gender, the results obtained indicate that for linking /r/ the ratio of cases produced by females is slightly higher in

comparison to the cases produced by males. To find out whether the difference across genders is statistically significant, a Pearson's chi-squared test with Yates' continuity correction was applied. The test reveals a relationship between the production of linking /r/ and gender ($X^2(1) = 85.0, p < 0.05$). This means that the difference found between the cases produced and gender is statistically significant. It could be concluded that the relationship between these variables may not be due to chance. The Cramer's V test confirms that the association is weak ($V = 0.2$). These results confirm the research hypothesis regarding sociolinguistic factors addressed for the second research question: gender seems to play a role in linking /r/ production in spontaneous speech.

Usage-based factors

Concerning another main hypothesis entertained in the second research question, i.e., that the production of linking /r/ would be motivated by usage-based factors, such as lexical and collocability frequency, statistical tests were conducted to determine the degree of association between linking /r/ and the lexical frequency and collocation frequency both the AMS corpus (see also Tables 17 and 18 below) and the reference corpus (BNC Spoken) (see also Tables 19 and 20 below). In light of the previous research hypothesis, a correlation matrix that assesses the degree of influence that the lexical and collocation frequency has on the production of linking /r/ is presented (Figure 23 below).

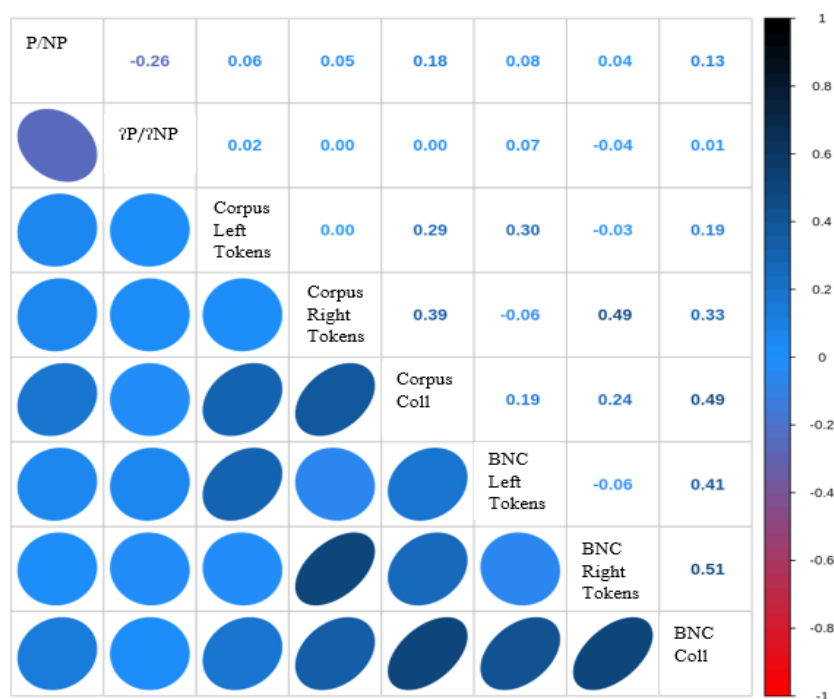


Figure 23. Correlation matrix for linking /r/ tokens. Actual cases produced, laryngeal strategies, and tokens (left and right-hand tokens and word collocations) in the AMS corpus and BNC Spoken corpus.

The results obtained indicate a weak association between the production of linking /r/ and the lexical frequency of the left-hand and right-hand tokens of each potential linking /r/ context and the lexical frequency of each word collocation computed as frequent in the AMS corpus. The strongest association among the variables is found between the production of linking /r/ and collocational frequency ($r = 0.18$) and for the data obtained from the BNC Spoken corpus ($r = 0.13$). There is also a strong effect between the production of linking /r/ and the left-hand lexemes computed as frequent in both the AMS corpus and the reference corpus ($r = 0.6$; $r = 0.8$, respectively), a pattern of association that is repeated for the right-hand lexemes computed as frequent in both corpora ($r = 0.5$; $r = 0.4$), although the degree of association is less strong.

To find out whether the data obtained report statistically significant results on the relationship of token frequency and the realisation of actual linking /r/ cases both in the AMS corpus and the BNC Spoken corpus, a Pearson's chi-squared test with Yates' continuity correction was applied. The test reveals a relationship between the production of linking /r/ and lexical frequency and the production of linking /r/ and collocational

frequency in both corpora ($X^2(1) = 160.0, p < 0.05$; $X^2(1) = 429.2, p < 0.05$; and $X^2(1) = 429.5, p < 0.05$; for left-hand tokens, for right-hand tokens and collocations, respectively). This means that the relationship found between the lexemes and word combinations computed as frequent in both the present study and in the BNC Spoken corpus and linking /r/ production is statistically significant. The results reveal that the relationship between these variables may not be due to chance, but due to an influence of the degree of frequency of lexemes and word combinations affecting the production of linking /r/ in connected speech.

The Cramer's V test indicates that the association is weak for the left-hand lexemes and moderate for the right-hand lexemes and word collocations ($V = 0.3$; $V = 0.5$; $V = 0.5$, for the left-hand items, the right-hand items and collocations, respectively). Therefore, the hypothesis is confirmed: there seems to exist a relationship between the production of linking /r/ and lexical and collocational frequency, and there exists a correlation between the production of linking /r/ and lexical and collocational frequencies in both the AMS corpus and the reference corpus.

Linguistic and phonetic factors

In relation to other main hypotheses entertained in the study that are related to the second research question, i.e., that the production of linking /r/ would be motivated by linguistic factors, the data were coded in terms of word class (C: content; F: function) combinations (i.e., CF, CC, FF, FC).

As a representative sample of the word-class collocations found in the study for potential cases of linking /r/, the ten most frequent segments within the AMS corpus were selected according to their production (see Table 5 below). The most frequent word collocation is *there are*, which has a frequency of 120 cases out of 1,800 potential cases for linking /r/ with a production of 112 linking /r/ cases.

Table 5. Ten most frequent tokens for linking /r/.

Word-class collocations	No /r/	/r/	Total
<i>there are</i>	8	112	120
<i>there is</i>	7	95	102
<i>for example</i>	0	61	61
<i>for a</i>	10	49	59
<i>number of</i>	2	44	46
<i>here in</i>	5	20	26
<i>for us</i>	4	20	24
<i>are in</i>	6	15	21
<i>our economy</i>	2	19	21
<i>their own</i>	6	13	19

The quantitative analysis shows that linking /r/ is categorical or almost categorical in set phrases formed by FC, CF and FF word class collocations (see also Table 16 below). Some instances of these word-class collocations are *for example* (100%) and *our economy* (90.5%) formed by an FC word class collocation, *number of* (95.7%) formed by a CF word class collocation, and *there is* (93.1%) and *there are* (93.3%) formed by a FF word class collocation. Some other word-class collocations that exhibit a high ratio production are *for us* (83.3%) and *for a* (83.1%) formed by an FF word class collocation. Also, *here in* (80%) and *are in* (71.4%) formed by a CF word class collocation, and *their own* (68.4%) formed by an FC word class collocation.

According to the results, the word-class collocation in which potential linking /r/ sites are most likely to appear corresponds to FF word collocations, although FC word combinations represent a categorical linking /r/ site. CC word class collocations have the lowest ratio of production. The highest score of a CC word class collocation is the phrase *year ago* that appears just 9 times.

An interesting subset of potential cases of linking /r/ where the stress pattern may be relevant is that of sites involving the deletion of /h/ in the weak form of a function word after the potential linking /r/. For this subset, the total amount of potential tokens

amounts to 13, which include the tokens *there have*, *minister has*, *year has*, *empire have*, *ever have*, *hear him*, *labour have*, *never have*, *sector have*, *there he*, *where he*, *number have*. The auditory and acoustic analyses assert that eleven cases contain linking /r/ out of the potential 13 cases found in the sample. The results are, however, not solid enough to test any impact of /h/ deletion on linking /r/ production due to the scarcity of data. Therefore, the hypothesis entertained in the second research question in relation to the presence or absence of /h/ deletion and another /r/ in the vicinity in the production of linking /r/ cannot be confirmed.

Turning now to phonetic factors, the data were also coded for factors such as the stress pattern of the syllables at both sides of the potential linking /r/ context (i.e., SS: stressed-stressed; SU: stressed-unstressed; UU: unstressed-unstressed; US: unstressed-stressed) and the presence or absence of /r/ in the syllable likely to make the link.

Another phonetic factor that may affect the production of linking /r/ refers to potential cases where the syllable likely to make the link begins with /r/. In the case of the presence of another /r/ in the vicinity, the total amount of tokens amounts to six, which include tokens such as *fairer I*, *fairer at*, *error a*, *terror or*. The auditory and acoustic analyses suggest that all of the word collocations except for *terror or* contain linking /r/. As the tokens represent such a small proportion within the whole AMS corpus, the results are not conclusive enough to test any impact of another /r/ in the vicinity of linking /r/ production due to the scarcity of data. Therefore, the hypothesis entertained in the second research question concerning the presence or absence of another /r/ in the vicinity in the production of linking /r/ cannot be confirmed.

Even though some linguistic and phonetic factors cannot be tested for significance due to the scarcity of data, statistical tests were conducted concerning the data for the linguistic (lexical) factor of word-class collocations and the phonetic factor of stress patterns found in the study (see 3.3.4.). To assess which word-class collocation and stress pattern has the greater degree of influence regarding the production of linking /r/ (see also Tables 16 and 21 below), a correlation matrix (Figure 24 below) that assesses the degree of influence that the linguistic and phonetic factors have on the production of linking /r/ is presented.

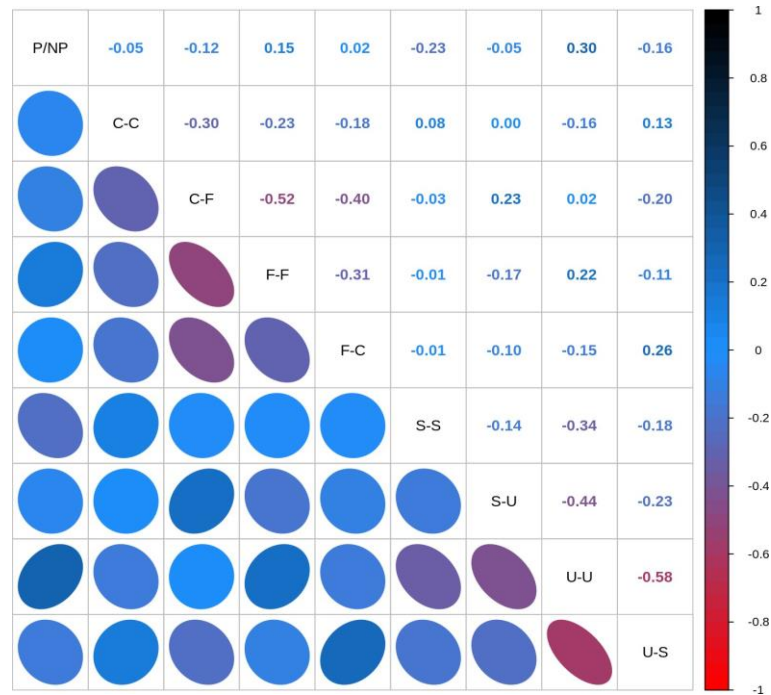


Figure 24. Correlation matrix for linking /r/ production and word-collocations and stress patterns combinations for linking /r/ in the AMS corpus.

The results indicate that the strongest degree of association between the production of linking /r/ and the type of word-class collocation is found when both lexemes are function words ($r = 0.15$). In contrast, the weakest degree of association is when the word combinations are formed by a content word followed by a function word ($r = -0.12$). The degree of association between the production of linking /r/ and an FC word combination is small ($r = 0.02$). CC word combinations and the production of linking /r/ show a negative correlation ($r = -0.05$). The strongest degree of association is found between the production of linking /r/ and word combinations containing unstressed syllables, although the degree of association is moderate ($r = 0.30$). The association is, however, negative between the production of linking /r/ and the rest of the stress pattern combinations ($r = -0.23$; $r = -0.05$; $r = -0.16$, for SS, SU and US, respectively).

The associations found between the production of linking /r/ and word-class collocations and stress patterns are negative, which indicates that the effect shown is negligible. Nevertheless, to find out whether the data obtained report statistically significant results on the relationship of word-class collocation and stress patterns regarding linking /r/ production a Pearson's chi-squared test was applied to both factors.

The test reveals a relationship between word-class collocation and stress patterns and linking /r/ production ($X^2(1) = 427.5, p < 0.05, V = 0.2$; and $X^2(1) = 1050.3, p < 0.05, V = 0.4$; for word-class collocations and stress patterns, respectively). This means that the relation found between the word combinations and stress patterns is statistically significant. The results reveal that the relationship between these variables may not be due to chance but to an influence of the word combinations and stress patterns affecting the production of linking /r/ in connected speech. Therefore, the hypotheses are confirmed: there exists a relationship, although not very strong, between the production of linking /r/ and word-class collocations and stress patterns.

4.1.2. Study 2: t-tapping

RQ1 questions how frequent t-tapping is in conversational language use. Given this question, the results obtained for t-tapping show that 526 cases were produced out of 1,800 potential cases. This, therefore, corresponds to 29.2% of all potential cases, whereas 1,274 of the potential cases not produced corresponds to 70.8% (see Figure 25 below). T-tapping, therefore, is not a highly frequent phenomenon in the data analysed but its frequency is also not neglectable.

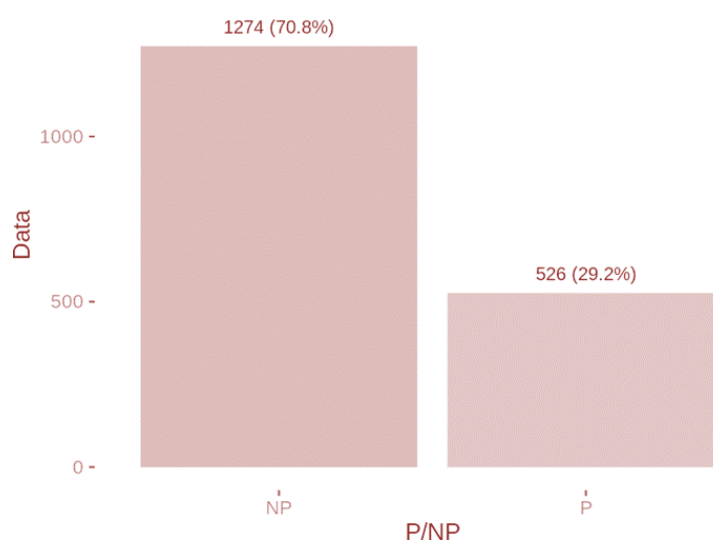


Figure 25. Bar plot by produced and not produced cases for t-tapping.

The results indicate that the use of t-tapping in the AMS corpus is far from categorical. Given this, an analysis was also conducted of the cases where t-tapping was not employed. Two strategies are possible in such cases: the use of alveolar plosives ([t, d]) or a voiceless glottal stop in the case of /t/ (but not /d/). This analysis is also particularly relevant since the traditional variant in RP is the use of an alveolar plosive (/t/ or /d/) but the use of voiced taps is now a feature of spontaneous speech but glottal stops may also occur as a realisation of /t/ across word boundaries (e.g., *what is* [wɒ? ɪz]).

The production of t-tapping was divided into tokens containing <t> and tokens containing <d>. Out of 1,326 tokens containing <t>, 410 were produced as taps (22.8% of the general tap production); while 116 tokens containing <d> were produced as taps out of the 474 tokens at potential t-tapping sites (6.4% of the general tap production). As Table 6 below shows, these data yielded 29.2% of the overall tap production (526 cases produced out of 1,800 potential cases) over glottal stops, which occurs in 10.6% of the potential cases (135 cases produced out of the 1,274 remaining potential cases). Out of 916 tokens quantified in which <t> was not produced as a tap, 135 tokens were identified as containing glottal stops (14.7% of the potential cases where glottal stops could have been produced in /t/ contexts). The production of alveolar plosives /t, d/ is 89.4% (1,139 cases out of the 1,274 remaining cases). In all, alveolar plosives /t, d/ are the preferred realisation when taps or glottal stops are not used, since out of the 1,800 potential t-tapping sites, 1,139 tokens (63.3%) are found when t-tapping or glottal stop are not applied.

Table 6. T-tapping strategies divided into categories.

<t> > /r/	n	%	<d> > /r/	n	%	<t> > ?	n	%
NP	916	50.9	NP	358	19.9	NP	781	43.4
P	410	22.8	P	116	6.4	P	135	7.5
Total	1,326	73.7	Total	474	26.3	Total	916	50.9

To answer RQ2, the results regarding the patterns of variation that motivate the t-tapping production according to a number of linguistic, sociolinguistic, usage-based, and phonetic factors are mentioned next.

Sociolinguistic factors

Concerning the production of t-tapping in terms of sociolinguistic factors, the data were divided by speaker and gender. All the speakers have a production below 150 cases, which constitutes a percentage of production lower than 50%. Four speakers have a consistent production of t-tapping between 20% and 30%, while two speakers have a production of t-tapping above 35%. The average percentage is 29.2%, which indicates that only three speakers are over the mean (see Table 7 below).

Table 7. Number and rate of potential cases produced in the AMS corpus divided by speakers (t-tapping).

Speaker	Cases produced	%	Cases non produced	%
George Osborne - Male 1	108	36.0	192	64.0
Theresa May - Female 1	67	22.3	233	77.7
Ed Balls – Male 2	67	22.3	233	77.7
Yvette Cooper - Female 2	80	26.7	220	73.3
David Cameron - Male 3	115	38.3	185	61.7
Harriet Harman - Female 3	89	29.7	211	70.33

The results show that males produced 290 cases out of 900 potential cases, which corresponds to 32.2%; while women's production resulted in 236 cases out of 900 potential cases, which corresponds to 26.2% (see Figure 26 below).

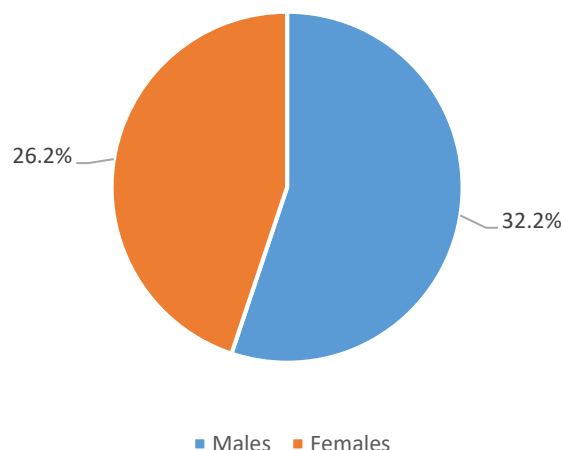


Figure 26. T-tapping production by gender.

With regard to one of the main hypotheses entertained in the second research question, i.e., that the production of linking /r/ would be motivated by gender the results indicate that the cases produced by males are slightly higher in comparison to the cases produced by females. As Table 8 illustrates, 26.2% of production corresponds to female speakers, while the male speakers' ratio corresponds to 32.2%.

Table 8. Overall cases produced and not produced divided by gender for t-tapping.

P/NP Males	n	%	P/NP Females	n	%
NP	610	67.8	NP	664	73.8
P	290	32.2	P	236	26.2
Total	900	100.0	Total	900	100.0

In order to find out whether the data obtained report statistically significant results between the production of t-tapping and gender, a Pearson's chi-squared test with Yates'

continuity correction was applied. The test reveals a relationship between the production of t-tapping and gender ($X^2(1) = 7.6$, $p < 0.05$, $V = 0.1$). This means that the difference between the cases produced and analysed by gender is statistically significant. The results reveal that the relationship between these variables may not be due to chance, but due to an effect of variation in the use of connected speech phenomena in spontaneous speech. The Cramer's V test confirms that the association is extremely weak ($V = 0.1$). Therefore, the hypothesis is confirmed, which indicates that there exists a relationship between the production of t-tapping and gender.

It is also interesting to see the relationship between gender and the use of glottal stops as an alternative to taps in potential cases of t-tapping. As Table 9 illustrates, the results show that the production of glottal stops when t-tapping is not used at potential sites is extremely low for both groups, females and males.

Table 9. Glottal stop production for t-tapping.

<i>Glottal Stop</i>	Female	Males
NP	852	813
P	48	87

Even though the production of t-glottaling is scarce for males and females, a Pearson's chi-squared test with Yates' continuity correction was applied. In order to find out whether the data obtained report statistically significant results between the production of glottal stops and gender. The test reveals a relationship between the production of glottal stops and gender ($X^2(1) = 11.6$, $p < 0.05$, $V = 0.1$). This means that the difference found between the cases of glottal stop and gender is statistically significant: males seem to use the glottal stop more than females.

Usage-based factors

With regard to another main hypothesis entertained in connection with the second research question, i.e., that the production of t-tapping will be motivated by usage-based factors, such as lexical and collocability frequency, statistical tests were conducted to

determine the degree of association between t-tapping and usage-based factors in both the AMS corpus (see also Tables 17 and 18 below) and the BNC Spoken corpus (see also Tables 19 and 20 below). In light of the previous research hypothesis, a correlation matrix that assesses the degree of influence of the usage-based factor under investigation on the production of t-tapping is presented (Figure 27 below).

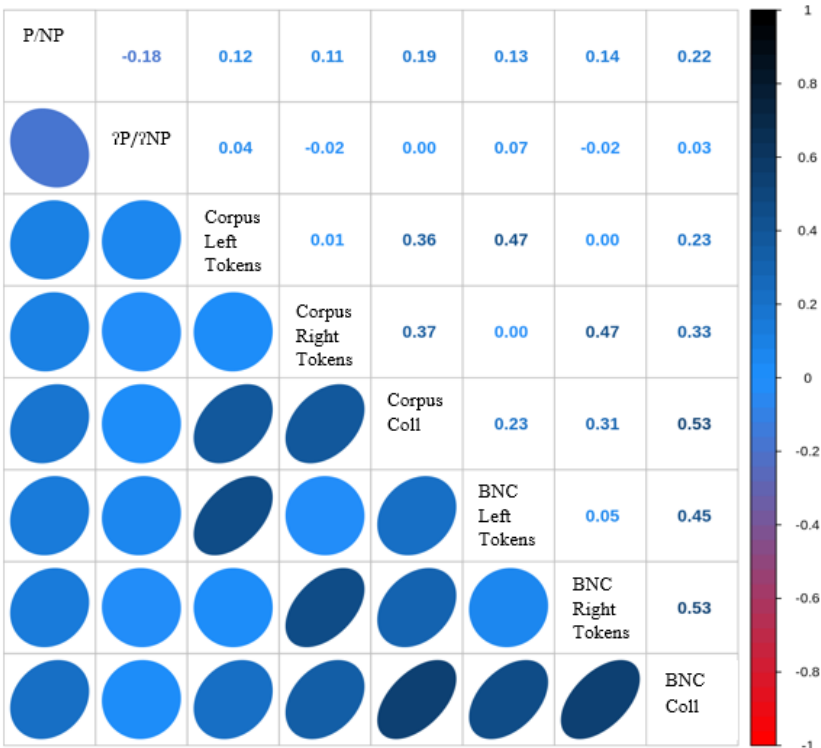


Figure 27. Correlation matrix for t-tapping tokens. Actual cases produced, laryngeal strategies, and tokens (left and right-hand tokens and word collocations) in the AMS corpus and BNC Spoken corpus.

The results obtained indicate that there exist strong associations in the production of t-tapping in relation to the factors analysed. Particularly, the degree of association between the production of t-tapping and the left-hand and right-hand tokens computed as frequent both in the AMS corpus and in the reference corpus is similar in strength ($r = 0.12$; $r = 0.11$; $r = 0.13$; $r = 0.14$; respectively). Similarly, the association is strong between the production of t-tapping and the word combinations (i.e., CC, CF, FF, FC) computed as frequent in both the AMS corpus and the BNC Spoken corpus ($r = 0.19$; $r = 0.22$, respectively).

Statistical tests were conducted to find out the relationship between the production of t-tapping and usage-based parameters (lexical frequency and collocational frequency) in both the AMS corpus and the BNC Spoken corpus. The Fisher's exact test for count data was applied for the left-hand items; while a Pearson's chi-squared test with Yates' continuity correction was applied for the right-hand items as well as for the collocations to compare the frequent and infrequent items in both corpora. The tests reveal a relationship between lexical and collocational frequency in both corpora (i.e., $X^2(1) = 404.9$, $p < 0.05$; $X^2(1) = 394.8$, $p < 0.05$; $X^2(1) = 504.1$, $p < 0.05$; for left-hand tokens, for right-hand tokens and collocations, respectively). This means that the relationship found between the lexemes and word combinations computed as frequent in the present study and the BNC Spoken corpus regarding t-tapping production is statistically significant: high-frequency lexemes and word collocations seem to have an impact on the production of t-tapping.

The Cramer's V test confirms that the association is weak for the left-hand lexemes and moderate for both the right-hand lexemes and for collocations ($V = 0.5$; $V = 0.5$; $V = 0.5$, for the left-hand items, the right-hand items and collocations, respectively). Therefore, the hypothesis is confirmed, which indicates that there exists a relationship between the production of t-tapping and usage-based factors.

Linguistic and phonetic factors

In relation to other main hypotheses entertained in the study that are related to the second research question, i.e., that the production of t-tapping will be motivated by linguistic factors and phonetic factors, the data were codified in terms of word-class collocations and stress patterns.

As a representative of the sample tokens concerning word-class collocations in the t-tapping sample, the eleven most frequent tokens within the AMS corpus were selected according to their production (see Table 10 below), from *that is* to *(a) lot of* and *what is*. The most frequent word collocation for t-tapping is *that is*, which has a frequency of 88 cases out of 1,800 potential cases with a production of 38 t-tapping cases.

Table 10. Ten most frequent tokens for t-tapping.

Word-class collocations	No tap	tap	Total
<i>that is</i>	50	38	88
<i>I don't</i>	40	25	65
<i>it is</i>	47	17	64
<i>but I</i>	10	41	51
<i>we don't</i>	24	12	36
<i>what I</i>	19	16	35
<i>part of</i>	27	3	30
<i>out of</i>	27	3	30
<i>set out</i>	23	5	28
<i>(a) lot of</i>	19	8	27

The quantitative analysis shows that t-tapping is not categorical but likely to occur in set phrases formed by FC (e.g., *set out* [ser aʊt]), CF (e.g., *part of* [pɑːr əv]), and FF (e.g., *but I* [bʌr aɪ]) word collocations. A few tokens represent two-word combination categories because of the contexts in which they appear. For example, *that is* appears 88 times, 13 of those tokens correspond to FF word collocations containing 4 tap tokens, while 75 of those tokens function as an FC word class collocation containing 34 tap tokens. Another example is *it is*, which appears 64 times. Three of those tokens function as an FF word class collocation containing one tap token, while 61 of those tokens function as an FC word class collocation containing 16 tap tokens.

FF word class collocation examples are: *I don't* (38.5%), *but I* (80.4%), *we don't* (33.3%), *what I* (45.7%), *out of* (10%). CF word collocations examples are *part of* (10%), *set out* (17.9%), and *lot of* (29.6%).

The word-class collocation in which potential t-tapping sites are most likely to appear corresponds to FF word collocations, whereas CC word class collocations have the lowest ratio of production. The highest score of a CC word class collocations is the phrase *right approach* which appears just 6 times.

Statistical tests were conducted for the type of word collocations and stress pattern combinations to assess which word-class collocations and stress patterns have the greater degree of association regarding the production of t-tapping (see also Tables 17 and 18 below). In light of the previous research hypotheses (i.e., linguistic and phonetic factors will motivate the production of t-tapping), a correlation matrix that assesses the degree of influence that the linguistic and phonetic factors under investigation have on the production of t-tapping is presented in Figure 28 below.

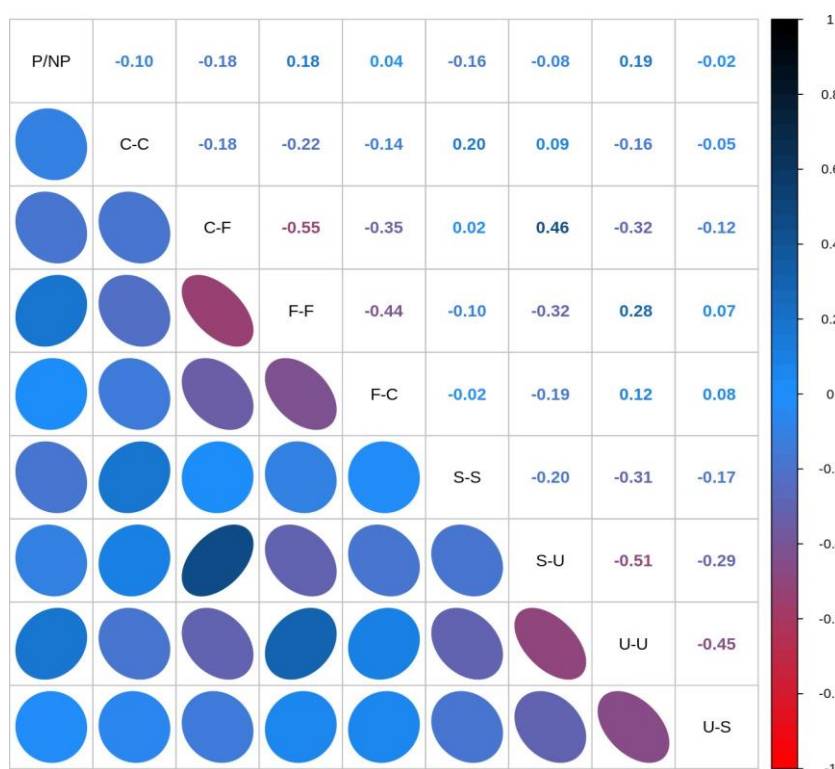


Figure 28. Correlation matrix for t-tapping production and word-collocations and stress patterns combinations for t-tapping in the AMS corpus.

The coefficient values related to the production of t-tapping and the word collocations CC and CF are below 0.1, which indicates a negative correlation among these variables ($r = -0.10$; $r = -0.18$, respectively). In contrast, the association between the production of t-tapping and the word collocations FF and FC shows a stronger association ($r = 0.18$, $r = 0.04$, respectively). FF word collocations show a stronger association than FC word collocations. The association between the production of t-tapping and stress patterns reveals that the strongest degree of association is found in word collocations in UU

environments, in both the syllable likely to trigger the tap and the following syllable (i.e., $r = 0.19$). However, the association is negative between the production of t-tapping and the rest of the stress patterns ($r = -0.16$; $r = -0.08$; $r = -0.02$).

To find out whether the data obtained report statistically significant results on the relationship of word-class collocations and stress patterns a Pearson's chi-squared test was applied to both factors. The test reveals a relationship between word-class collocation and stress patterns regarding t-tapping production ($X^2(1) = 601.7$, $p < 0.05$, $V = 0.3$; and $X^2(1) = 598.3$, $p < 0.05$, $V = 0.9$; for word-class collocations and stress patterns respectively). This means that the relation found between the word collocations and stress patterns is statistically significant. These data report that the relationship between these variables may be due to an influence of the word collocations and stress patterns affecting the production of t-tapping in connected speech. Therefore, the hypotheses are confirmed, which indicates that there exists a relationship between the production of t-tapping and word-class collocations and stress patterns.

Finally, the use of t-tapping was also considered in connection with the prosodic environment of the words on either side of the potential cases. It should be born in mind, however, that t-tapping can only occur when the lexical item after the potential site is unstressed in the specific prosodic context.

4.1.3. Study 3: yod coalescence

The results obtained for yod coalescence show that 311 cases of yod coalescence were identified out of 960 potential cases. This corresponds to 32.4% in the AMS corpus, whereas 649 were not produced (67.6%; see Figure 29 below).

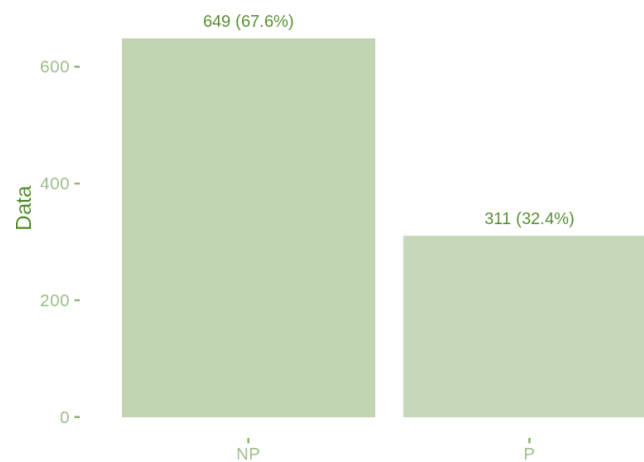


Figure 29. Bar plot by cases produced and not produced for yod coalescence.

Regarding the main hypotheses entertained for RQ2, several factors (linguistic, sociolinguistic, usage-based, and phonetic) also motivate the variable production of each connected speech phenomenon, as in the case of the other two processes analysed above (linking /r/ and t-tapping).

Sociolinguistic factors

Concerning the production of yod coalescence, the data were divided by speaker and gender. All the speakers have a production below 80 cases, which constitutes a percentage of production lower than 50% of the overall potential cases for each speaker. Five speakers have a production of yod coalescence between 25% and 40%, while only one speaker has a production of yod coalescence above 40%. The average percentage is 32.4%, which indicates that only three speakers are above the mean (see Table 11 below).

Table 11. Overall yod coalescence production divided by speakers.

Speaker	Cases produced	%	Cases non produced	%
George Osborne - Male 1	42	26.3	118	73.8
Theresa May - Female 1	62	38.8	98	61.3
Ed Balls - Male 2	66	41.3	94	58.8
Yvette Cooper - Female 2	40	25.0	120	75.0
David Cameron - Male 3	44	27.5	116	72.5
Harriet Harman - Female 3	57	35.6	103	64.4

The results show that males produced 152 cases out of 480 potential cases (31.7%), while women's ratio of production resulted in 159 cases out of 480 potential cases (33.1%) (see Figure 30 below).

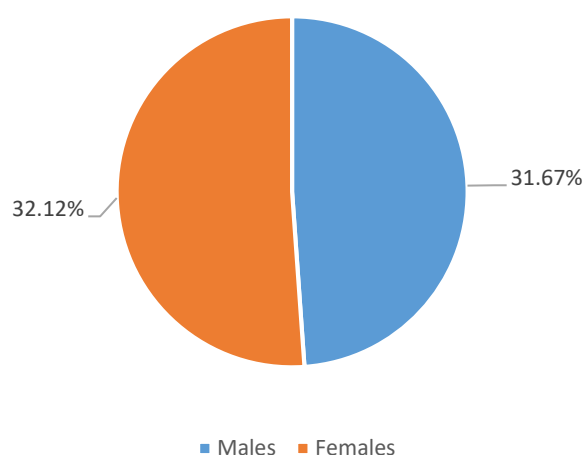


Figure 30. Yod coalescence production by gender.

With regard to one of the main hypotheses entertained in the second research question, i.e., that the production of yod coalescence will be motivated by sociolinguistic factors, such as speaker gender, the results obtained indicate that for yod coalescence, the rate of production of yod coalescence is similar in both gender groups. As Table 12 illustrates,

the percentage of production of 33.10% corresponds to the female speakers, while the percentage of production for the male speakers is 31.70%.

Table 12. Cases produced and not produced divided by gender for yod coalescence.

P/NP Males	n	%	P/NP Females	n	%
NPP	328	68.33	NP	321	66.88
Total	152	31.67	P	159	33.12
	480	100.00	Total	480	100.00

To find out whether the data report statistically significant results, a Pearson's chi-squared test with Yates' continuity correction was applied. The test reveals no relationship between the production of yod coalescence and gender ($X^2(1) = 0.17, p > 0.05, V = 0.016$). This means that the difference found between the cases produced and gender is not statistically significant. The relation between these variables, as the data report, may be due to an effect of variation in the use of connected speech phenomena in spontaneous speech. The Cramer's V test confirms that the association is weak for these variables. Therefore, the hypothesis is not confirmed, which indicates that there exists no relationship between the production of yod coalescence and gender.

Usage-based factors

With regard to another main hypothesis entertained in the second research question, i.e., that the production of yod coalescence will be motivated by usage-based factors, such as lexical and collocability frequency, statistical tests were conducted to determine the degree of association between yod coalescence and lexical as well as collocational frequencies in both the AMS corpus (see also Tables 17 and 18 below) and the BNC Spoken corpus (see also Tables 19 and 20 below). In light of the previous research hypothesis, a correlation matrix that assesses the degree of influence that the usage-based

factor under investigation has on the production of yod coalescence is presented (Figure 31 below).

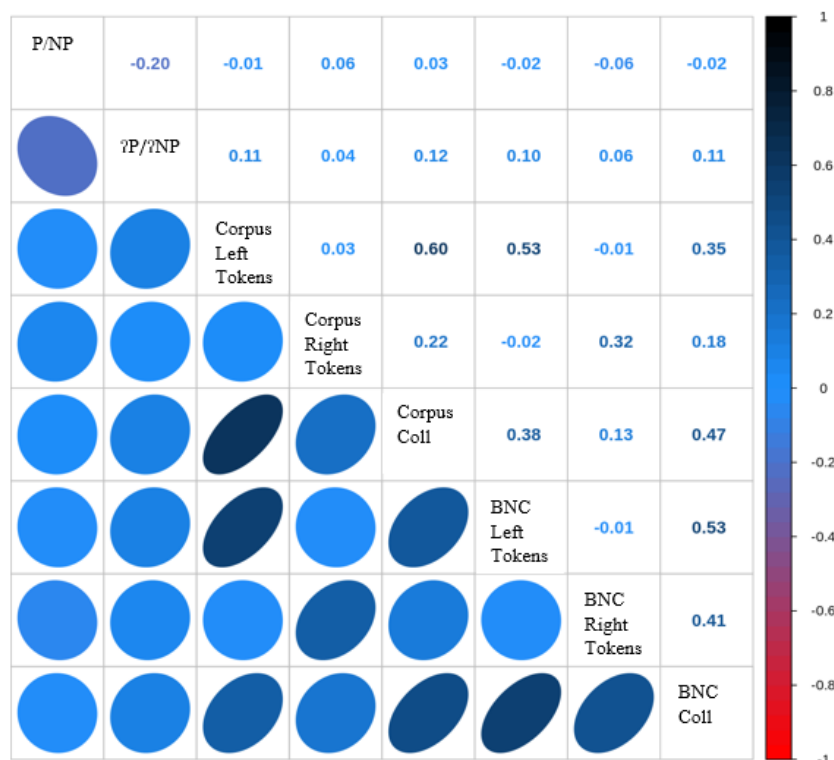


Figure 31. Correlation matrix for yod coalescence tokens. Actual cases produced, laryngeal strategies, and tokens (left and right-hand tokens and word collocations) in the AMS corpus and BNC Spoken corpus.

The results indicate that there exist almost no association between the production of yod coalescence and the left-hand and right-hand tokens or the word combinations computed as frequent in the AMS corpus. A similar pattern is given for the association between the production of yod coalescence and the left-hand and right-hand tokens as well as the word combinations computed as frequent in the BNC Spoken corpus. The correlation matrix shows that the association is negative between the production of yod coalescence and the lexical and collocability frequency found in the reference corpus ($r = -0.02$; $r = -0.06$; $r = -0.02$; for left-hand, right-hand and word-class collocations, respectively). However, the only negative association is found between the production of yod coalescence and the lexical frequency of left-hand tokens in the AMS corpus ($r = -0.01$). The association between the production of yod coalescence and the lexical frequency of right-hand tokens

and word-class collocations is positive ($r = 0.06$; $r = 0.03$; left-hand items and word-class collocations, respectively), which shows a stronger association in the case of the left-hand tokens than in word-class collocations.

Statistical tests were also conducted to find out whether there is a relationship between lexical and collocational frequency in both corpora. A Pearson's chi-squared test with Yates' continuity correction was applied for the left-hand items and the collocations, whereas a Fisher's exact test for count data was applied for the right-hand items. The tests reveal that there is a statistically significant relationship between lexical and collocational frequency in both corpora ($X^2(1) = 261.09$, $p < 0.05$; $p < 0.05$, $V = 0.321$; and $X^2(1) = 210.81$, $p < 0.05$, $V = 0.471$, for left-hand tokens, for right-hand tokens and collocations, respectively). The Cramer's V test indicates that the association is moderate for the left-hand lexemes and for collocations ($V = 0.527$; $V = 0.321$; $V = 0.471$, for the left-hand items, the right-hand items and collocations, respectively). Therefore, the hypothesis is confirmed, which indicates that there exists a relationship between the production of yod coalescence and usage-based factors.

Linguistic and phonetic factors

With regard to other main hypotheses entertained in the second research question, i.e., that the production of yod coalescence will be motivated by linguistic and phonetic factors, motivated by linguistic factors and phonetic factors.

By way of illustration, the ten most frequent segments within the AMS corpus were selected according to their production (see Table 13 below). The most frequent word collocation is *that you*, which has a frequency of 85 cases out of 960 potential cases for yod coalescence with a production of 10 yod coalescence cases.

Table 13. Ten most frequent tokens for yod coalescence.

Word-class collocations	No yod	yod	Total
<i>that you</i>	75	10	85
<i>and you</i>	65	10	75
<i>as you</i>	6	37	43
<i>but you</i>	35	4	39
<i>this year</i>	18	13	31
<i>what you</i>	23	4	27
<i>last year</i>	7	19	26
<i>next year</i>	6	12	18
<i>because you</i>	6	12	18
<i>is you</i>	7	6	13

The quantitative analysis shows that yod coalescence is not categorical and that it is likely to occur in set phrases formed by FF, FC, CC and CF word class collocations. There is one token that represents two-word combination categories because of the contexts in which it appears: *is you*, which appears 13 times. Twelve of those tokens function as a CF word class collocation containing six yod tokens, while one of those tokens functions as an FF word class collocation containing one yod token. FF word class collocation examples are: *that you* (27.33%), *and you* (24.12%), *as you* (13.83%), *but you* (12.54%), *what you* (8.68%) and *because you* (5.79%). An FC word combination example is *this year* (9.97%). The CC word class collocations are *last year* (8.36%) and *next year* (5.79%).

The word-class collocation in which potential yod coalescence sites are most likely to appear corresponds to FF word combinations, although FC word combinations also represent potential yod coalescence sites. CC word-class collocations have the lowest ratio of production.

Statistical tests were also conducted concerning the type of word combinations and the stress patterns combinations found in the study (see 3.3.3.) to assess which word-class collocation and stress pattern have the greater degree of association regarding the production of yod coalescence (see also Tables 16 and 21 below). In light of the previous

research hypotheses, a correlation matrix (Figure 32 below) that assesses the degree of influence that the linguistic and phonetic factors under investigation have on the production of yod coalescence is presented.

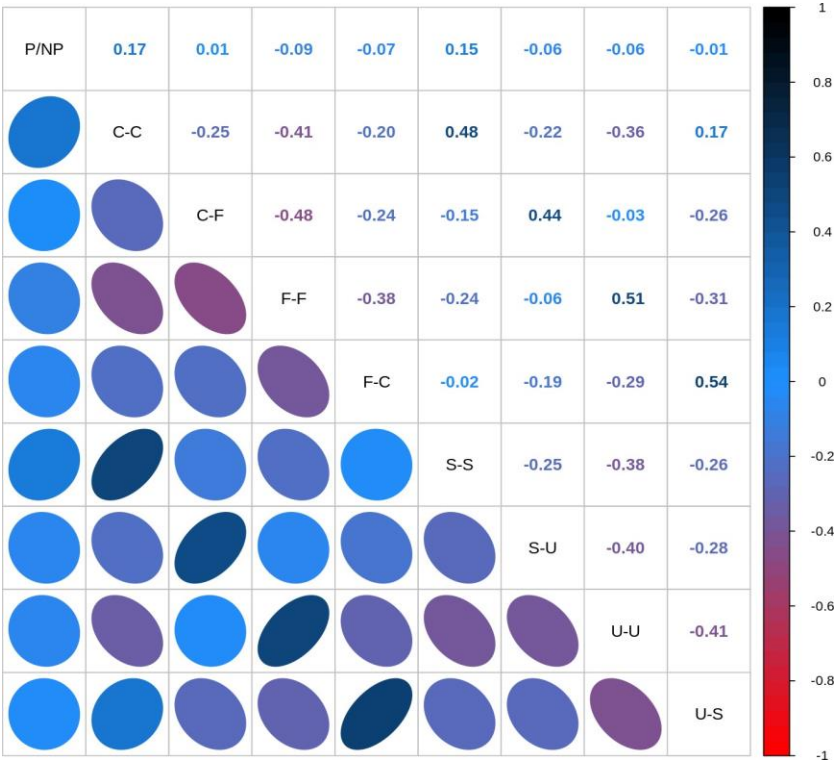


Figure 32. Correlation matrix for yod coalescence production and word-collocations and stress patterns combinations for yod coalescence in the AMS corpus.

The strongest degree of association between the production of yod coalescence and the type of word is found when both lexemes are content words ($r = 0.17$), while the weakest degree of association is found between the production of yod coalescence and the word combinations formed by two function words ($r = -0.09$). FC word combinations show a negative association concerning yod coalescence production ($r = -0.07$); while CF word combinations exhibit a positive, although weak association ($r = 0.01$). The association between the production of yod coalescence and stress patterns reveals that the strongest degree of association is found in the production of yod coalescence in word combinations containing stressed syllables in both the syllable likely to coalesce and the following syllable ($r = 0.15$). However, the degree of association is negative between the production

of the phenomenon of yod coalescence and the rest of the prosodic environments ($r = -0.06$; $r = -0.06$; $r = -0.01$; for SU, UU and US, respectively).

To find out whether the data obtained report statistically significant results on the relationship of word-class collocation and stress patterns regarding yod coalescence production, a Pearson's chi-squared test was applied to both factors. The test reveals a relationship between word-class collocation and stress patterns regarding yod coalescence production ($X^2(1) = 442.5$, $p < 0.05$, $V = 0.251$; and $X^2(1) = 103.26$, $p < 0.05$, $V = 0.164$; for word-class collocations and stress patterns respectively). This means that the relation found between the word combinations and stress patterns in the present study regarding yod coalescence production is statistically significant. The results reveal that the relationship between these variables may not be due to chance, but due to an influence of the word combinations and stress patterns affecting the production of yod coalescence in connected speech. Therefore, the hypotheses are confirmed, which indicates that there exists a relationship between the production of yod coalescence and word-class collocations and stress patterns.

4.1.4. General results

Once the results of each of the three studies were presented above, some general results emerged from the three studies. The general results indicate that in a corpus containing ca. 163,400 words, 2,157 cases (47.3%) were produced out of 4,560 potential cases of the three connected speech processes investigated (see Figure 33).

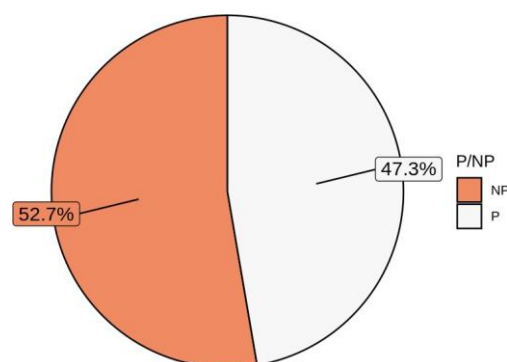


Figure 33. Overall cases produced and cases not produced.

Concerning the first hypothesis entertained in the first research question (RQ1i: each process will be used variably), the analysis of the data reveals that 1,320 (73.3%) and 526 cases (29.2%) of linking /r/ and t-tapping were produced, respectively. In the case of yod coalescence, 311 cases (32.4%) were produced (see Figure 34 below; see also 3.3.2.). Therefore, the results confirm that each process is variably used in spontaneous speech.

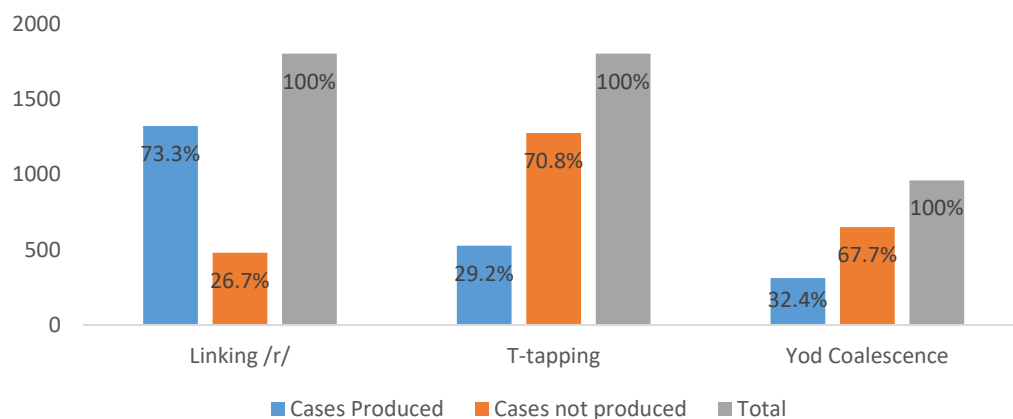


Figure 34. Data per connected speech phenomena.

Regarding the second hypothesis entertained in the first research question (RQ1ii: there will be significant differences in their rate of production), the analysis shows that the occurrence of the phenomena corresponds to 73.3% for linking /r/, 29.2% for t-tapping,

and 32.4% for yod coalescence (see Table 14). The results show that linking /r/ is the most frequent phenomenon because both t-tapping and yod coalescence scored lower ratios.

Table 14. Cases produced divided by phenomena.

Cases produced	n	%
Linking r	1320	73.3
T-Tapping	526	29.2
Yod coalescence	311	32.4

Several statistical tests were conducted to test the significance of the production of the three phenomena in spontaneous speech. A two-sample test for equality of proportions without continuity correction was applied to determine whether there was any difference between linking /r/ and t-tapping productions, which revealed statistically significant differences ($X^2(1) = 700.9, p < 0.05$). Another two-sample test for equality of proportions without continuity correction was applied to determine whether there exists any difference between t-tapping and yod coalescence productions. The test revealed no statistically significant differences ($X^2(1) = 3.0, p > 0.05$). Third, a two-sample test for equality of proportions without continuity correction was applied to determine whether there exists any difference between linking /r/ and yod coalescence production. The test revealed statistically significant differences ($X^2(1) = 434.1, p < 0.05$).

While the general results presented above consider six speakers as a group, it might also be interesting to see whether the data per speaker regarding the three phenomena reveal any specific idiosyncratic pattern. Table 15 shows that all speakers followed the general pattern of producing more potential cases of linking /r/ than cases of t-tapping or yod coalescence.

Table 15. Cases produced by speakers for linking /r/, t-tapping, and yod coalescence.

Speaker	Linking /r/ %	T-tapping %	Yod coalescence %
GO Male 1	66.67	36.00	26.25
TM Female 1	81.33	22.33	38.75
EB Male 2	59.00	22.33	41.25
YC Female 2	86.33	26.67	25.00
DC Male 3	65.33	38.33	27.50
HH Female 3	81.33	29.67	35.62

Moving now on to RQ2, which focuses on the factors that cause variation in the production of the three connected speech processes, the overall data are presented in connection with the different factors studied.

Concerning the differences across genders, a preliminary observation shows a contrast between the total number of cases recorded as a single group and the actual cases produced when divided by males and females. While the total number of occurrences in the AMS corpus (females and males) represents 47.30% of all potential cases for the three phenomena (see Figure 31 above), male informants produced 1,015 cases (47.06%), which indicates a slightly lower percentage of occurrences in comparison to female production with 1,142 cases (52.94%) (see Figure 35).

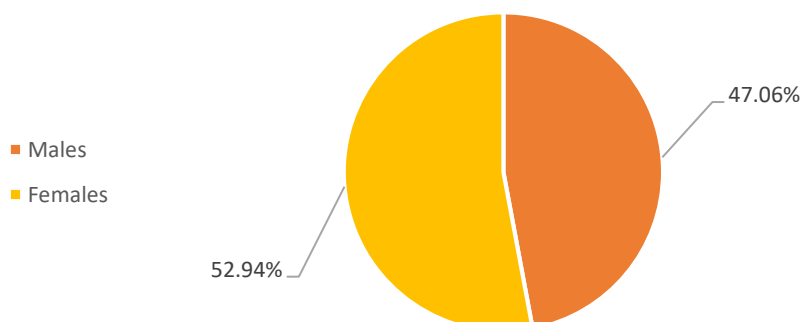


Figure 35. Males and Females overall production.

To determine whether the data report statistically significant results on the relationship between the production of the three connected speech phenomena and gender, the data were tested for statistical significance. A chi-square test revealed a relationship between gender and connected speech production ($X^2(1) = 14.2, p < 0.05$). This means that the difference found between the cases produced regarding gender is statistically significant. The results reveal that the relationship between these variables is because of variation in the use of connected speech phenomena in spontaneous speech. The Cramer's V test confirms that the association is extremely weak ($V = 0.1$).

Concerning the variation in the use of linking/r/, t-tapping, and yod coalescence in connected speech motivated by the grammatical status (content vs. function) of words, the coding of word-class patterns in the data analysis yielded two-word collocation patterns (see Table 16 below and see also 3.3.3.).

Table 16. Word-class combinations per phenomena: potential/actual cases per phenomenon and rate of actual cases

Word type	CC	%	CF	%	FF	%	FC	%
Linking /r/	210/141	67.1	718/480	66.9	521/435	83.5	351/264	75.2
T-tapping	120/15	12.5	554/95	17.2	733/288	39.3	393/128	32.6
Yod coalescence	173/86	49.7	219/73	33.3	416/114	27.4	152/38	25.0
Total	503/242	48.1	1,491/648	43.5	1,670/837	50.1	896/773	86.3

A chi-square test was also applied to find statistically significant differences between the three connected speech phenomena regarding the word class patterns. The test reveals a relationship between the production of the phenomena under investigation and the word-class patterns found in the AMS corpus ($X^2(1) = 216.1, p < 0.05$). This means that the difference between the cases produced and analysed in the present study and the word-class patterns is statistically significant. The results reveal that there exists a relationship between these variables in spontaneous speech. The Cramer's V test confirms that the association is extremely weak ($V = 0.2$).

As for lexical and collocational frequency, the data codification was twofold: lexical frequency and collocation frequency were quantified and tested separately. First, token frequency was coded as frequent and infrequent tokens using the median, in line with previous word-frequency studies (see 3.3.3.). The coding of the token frequency for the left-hand right-hand items represents 96.11% for linking /r/, 95.2%, for t-tapping, 88.54% for yod coalescence, and for the left and right items, 93.4% for linking /r/, 42.3% for t-tapping, and 48.3% for yod coalescence (see Table 17 below).

Table 17. Lexical frequency of the left-hand and right-hand tokens in the AMS corpus.

Position of token and frequency	Left-hand tokens				Right-hand tokens			
	Freq.	%	Infreq.	%	Freq.	%	Infreq.	%
Linking /r/	1,730	96.1	70	3.9	1,682	93.4	118	6.6
T-tapping	1,714	95.2	86	4.8	1,717	95.4	86	4.6
Yod coalescence	850	88.5	110	11.5	944	98.3	16	1.7

To consider the effect of lexical frequency on the production of linking /r/, t-tapping and yod coalescence, a chi-square test was conducted. The statistical test reveals a relationship between the production of the phenomena and token frequency (frequent and infrequent left-hand and right-hand tokens combined) in the AMS corpus ($X^2(1) = 8.8$, $p < 0.05$). This means that the difference found between the analysed in the present study and the lexemes quantified according to their frequency is statistically significant. The results, therefore, reveal that there exists evidence of the relationship between these variables in connected speech. The Cramer's V test confirms that the association is extremely weak ($V = 0.03$).

Second, the coding of collocation frequency was also quantified as frequent and infrequent word collocations using the median, in line with previous word frequency studies (see 3.3.3.). The coding of word-collocation frequency in the AMS corpus is 71.7% for linking /r/, 74.9% for t-tapping, and 74.6% for yod coalescence (see Table 18).

Table 18. Collocational frequency in the AMS corpus.

Frequency of word collocations	Frequent	%	Infrequent	%
Linking /r/	1,290	71.7	510	28.3
T-tapping	1,348	74.9	452	25.1
Yod Coalescence	716	74.6	244	25.4

To test the relationship between the production of the three connected speech phenomena and the word collocations computed as frequent and infrequent, a chi-square test was applied. The test reveals no relationship between the production of the phenomena and word collocations ($X^2(1) = 5.5, p > 0.05$). This means that the difference between the cases analysed, and collocation frequency is not statistically significant. The results reveal that the relationship between these variables is, therefore, negligible.

To interpret the influence of lexical and collocational frequencies, the data concerning tokens and word collocations in the present study were tested against a reference corpus (BNC Spoken) (see also 3.3.3.). The coding of the data for this part of the analysis was twofold, as was previously the case when testing lexical and word collocation frequency for the AMS corpus.

First, the token frequency (left-hand and right-hand items computed as frequent/infrequent) found in the AMS corpus was coded as frequent or infrequent against the reference corpus data (BNC Spoken). The token frequency represents 86.72% for linking /r/, 94.8% for t-tapping, 89.5% for yod coalescence for the left, and right items represents 87.3% for linking /r/; 91.2% for t-tapping and 93.2% for yod coalescence, in the reference corpus (see Table 19 below).

Table 19. Lexical frequency in the reference corpus (BNC Spoken).

Position of token and frequency	Left-hand tokens				Right-hand tokens			
	Freq.	%	Infreq.	%	Freq.	%	Infreq.	%
Linking /r/	1,561	86.7	239	13.3	1,571	87.3	229	12.7
T-tapping	1,706	94.8	94	5.2	1,641	91.2	159	8.8
Yod coalescence	859	89.5	101	10.5	895	93.2	65	6.8

To inspect the effect between the production of the three connected speech phenomena and lexical frequency in the reference corpus (BNC Spoken), a chi-square test was applied, which reveals a statistically significant relationship ($X^2(1) = 76.4, p < 0.05$). This means that the difference between the cases produced both corpora (AMS and BNC Spoken). The results reveal that there exists a relationship between the frequency of usage of a token and connected speech phenomena production. The Cramer's V test confirms that the association is extremely weak ($V = 0.1$).

Second, the collocability frequency (two-word combination patterns computed as frequent/infrequent) found in the present AMS corpus was also coded as frequent or infrequent using reference corpus data (BNC Spoken). The coding of frequency for word collocations in the reference corpus represents 67.8% for linking /r/, 81.1% for t-tapping, and 74.0% for yod coalescence, in the reference corpus (BNC Spoken) (see Table 20).

Table 20. Collocational frequency according to the BNC Spoken corpus.

Frequency of word collocations	Frequent	%	Infrequent	%
Linking /r/	1,220	67.8	580	32.2
T-tapping	1,429	79.4	341	20.6
Yod Coalescence	710	74.0	250	26.0

To determine the relationship between the production of the three connected speech phenomena under investigation and the word-collocations (frequent and infrequent) of the reference corpus (BNC Spoken), a chi-square test was applied. The test reveals a relationship between the production of the phenomena and word collocation frequency in the reference corpus ($X^2(1) = 83.2, p < 0.05$). This means that the relationship between these variables may be due to an effect of variation in the use of connected speech phenomena in spontaneous speech. The Cramer's V test confirms that the association is weak ($V = 0.1$).

These results on collocational frequency indicate that there exists a relationship between the word collocations catalogued as frequent/infrequent in the AMS corpus and the collocation frequency found for the same word collocations in the reference corpus (BNC Spoken).

Concerning the production of linking/r/, t-tapping, and yod coalescence in connected speech motivated by stress pattern combinations, the coding of stress patterns in the data analysis yielded four stress patterns (see Table 21 below; see also 3.3.4.).

Table 21. Stress pattern combinations per phenomena: potential/actual cases per phenomenon and rate of actual cases.

Stress Patterns	SS	%	SU	%	UU	%	US	%
Linking /r/	174/74	42.5	265/181	68.3	943/811	86.0	418/254	60.8
T-tapping	189/14	45.5	444/101	22.8	805/311	38.6	362/100	27.6
Yod coalescence	186/86	46.2	203/54	26.6	357/103	28.9	215/68	32.6
Total	549/174	31.7	912/336	36.8	2,105/1,225	58.2	995/422	42.4

A chi-square test was applied to test whether the data report statistically significant results on the relationship between the production of the three connected speech phenomena under investigation and the stress patterns. The test reveals a relationship between the production of the phenomena and stress patterns in the AMS corpus ($X^2(1) = 135.9$, $p < 0.05$). This means that the relationship between the cases produced regarding the prosodic environment is statistically significant. The Cramer's V test confirms that the association is weak ($V = 0.1$).

The results show that each process is used variably, and that there are significant differences in the rate of production in spontaneous speech due to phonetic, linguistic, sociolinguistic, and usage-based factors, which have contributed to the variable rate of production of the phenomena.

Summarising, the overall results show that 47.3% of the total number of cases analysed were produced, and that the phenomena are used variably in spontaneous discourse (see Table 14). In terms of sociolinguistic patterns, male speakers produced a slightly lower percentage than female speakers. As for linguistic factors, the analysis reveals that linking /r/ and t-tapping are favoured in FF word-collocations while yod coalescence is favoured in CC word-collocations. While the statistical results reveal no relationship between the connected speech phenomena and word-collocations, the tests

indicate a relationship between the lexical and collocational frequency and the cases produced in the AMS corpus. Concerning the production of linking /r/, t-tapping, and yod coalescence in relation to stress patterns, the statistical tests reveal statistically significant differences between the variables.

4.2. Discussion

Following the presentation of the results for each of the three studies, in the present section, the findings are discussed taking into account the theoretical approaches and previous empirical research in the field (see Chapter II). A discussion of each study is presented first. This is followed by a general discussion of the three studies.

4.2.1. Study 1: linking /r/

The results obtained for linking /r/ indicate that the production of this phenomenon in spontaneous speech is notably frequent (73.3%; 1,320 cases produced out of 1,800 potential cases –see Figure 18 in 4.1.), but not categorical. The high rate of linking /r/ usage in the RP accent of the six speakers analysed is not unexpected given that linking /r/ has traditionally been described as a common feature in several English accents (e.g., Foulkes & Docherty, 1999; Hay & Sudbury, 2005), including RP where it is claimed to be generally present (e.g., Cruttenden, 2014; Hannisdal, 2006; Mompeán, 2022). This finding is in line with some recent corpus-based studies looking at the speech of groups of RP speakers where scores for linking /r/ range from around 60% to 65% (e.g., Hannisdal, 2006; Mompeán & Gómez, 2011; Pavlík, 2016; see also 2.1.3.).

The high frequency of use of linking /r/ is consistent with its hiatus-breaking function, as pointed out by previous studies (e.g., Cox et al., 2014; McCarthy, 1993; Mompeán & Gómez, 2011). Hiatus occurs when adjacent syllables contain two separate vowels occurring alongside one another, both within and across word boundaries. Since this adjacency is phonologically undesirable in many languages, the latter often use a range of strategies to avoid hiatus. In English, linking /r/ (or /r/-sandhi more generally) is one of such strategies in a relatively complex system of resolving vowel-hiatus in vernacular varieties of English (Britain & Fox, 2009). Therefore, the high percentage of

/r/-sandhi resolution in the present study is not unexpected given that linking /r/ is thought to be the rule rather than the exception at potential /r/-sandhi sites in connected speech.

While the frequent use of linking /r/ is consistent with the description of the connected speech processes as a hiatus-breaking strategy, the use of laryngeal strategies was expected to occur in the data as well, even at very low rates. As a noticeable finding also reveals, the use of glottalisation instead of linking /r/ is also present as a strategy to break hiatus at potential /r/-sandhi sites. The well-established yet highly complex hiatus breaking system of British English seems, therefore, to be undergoing a radical reorganisation (see Britain & Fox, 2009).

The low percentage of laryngeal strategies found in the current study is remarkable, but perhaps unsurprising at the same time, since glottal stops have particularly been long stigmatised in RP (Mompeán, 2022), although this use of glottalisation as a hiatus-breaker has been described in previous studies (Cox et al., 2014; Mompeán, 2022; Mompeán & Gómez, 2011). One reason for this use of glottalisation as a hiatus-breaking strategy may be that the hiatus-breaking system is reorganising itself in some varieties of English, not only in cases of /r/-sandhi but also in contexts where other strategies are used with indefinite and definite articles (e.g., *a/an, the*) or function words (e.g., *to, of*) to block hiatus (Britain & Fox, 2009). Evidence of glottalisation is also found in lab-based studies (e.g., Davidson & Eker, 2014) as a more frequent hiatus-breaking strategy across word boundaries than glide insertion. In London adolescent speakers' speech, for example, the glide insertion system (i.e., [j] or [w] preceding a high vowel) to block hiatus seems to be regularising by using a glottal stop as hiatus-breaker instead of a glide (Britain & Fox, 2009).

Thus, glottal stops are not only replacing glides after high vowels but also replacing /r/ after non-high vowels, as it is a segment which is typically considered unmarked (e.g., Anttila & Cho, 1998). One potential justification for this is given by Lombardi (1977, 2002), who claims that glottal stops are less marked than coronal, labial, and dorsal consonants. This makes laryngeals (or glottal stops, more precisely) optimal epenthetic segments to override the /r/-sandhi linguistic function (i.e., linking /r/ as hiatus-breaker), as they are arguably “the next best choice” after pharyngeals (Lombardi, 2002, p. 223), at least, in areas with a significant language and sociocultural contact (Britain & Fox, 2009). However, the lack of a longitudinal scope in the present study does not allow to corroborate or refute the overriding of /r/-sandhi as a hiatus resolution strategy.

Additionally, /r/-sandhi and glottalisation can co-occur, although at not very high rates (Cox et al., 2014; Mompeán, 2022). Given that both /r/-sandhi and glottalisation block the occurrence of hiatus, the concomitant presence of both strategies could serve some prosodic function (Redi & Shattuck-Hufnagel, 2001).

While the hiatus-breaking system of RP still remains relatively conservative, it can now be confidently described as making ample use of glottalisation as an alternative hiatus-breaking strategy in potential cases of linking /r/ when the latter is not used. It should also be pointed out that the likelihood of glottalisation, particularly a full glottal stop, to be used instead of linking /r/ correlates with the prosodic pattern of the syllables at the potential linking /r/ site. In line with previous research (e.g., Hannisdal, 2006; Mompeán & Gómez, 2011), glottalisation is mostly used when the syllable after the potential linking /r/ site is stressed and less likely when it is unstressed.

One difference between the current study and the previous corpus-based studies is the fact that the former has analysed spontaneous speech while the corpus-based studies were based on scripted speech. Since the average rate of production of linking /r/ in the current study (73.3%) is higher than that of the studies in Hannisdal (2006), Mompeán & Gómez (2011), and Pavlík (2006), this may suggest that the type of speech delivery (i.e., scripted vs. spontaneous) may have some impact on the production of linking /r/. In fact, it has been pointed out that linking /r/ usage is a feature of fluent colloquial style (e.g., Hannisdal, 2006; Mompeán, 2022) and non-deliberate speech contexts (e.g., Lewis, 1975/2014), in contrast to speech with a slow delivery rate, which produces fewer [r]s (e.g., Heselwood, 2006; McMahon, 2000). Notwithstanding, the relative formality or informality and speech tempo of the type of speech delivery alone may not explain the slightly higher rate of linking /r/ in spontaneous speech than in scripted speech.

While the type of speech delivery may have some impact on the rate of production of linking /r/, the specialised literature suggests that several factors may motivate the production of linking /r/ and may also account for the variability found in this and previous studies. There is some suggestion that one reason for the lack of a categorical production in the use of linking /r/ could be that, in avoiding the use of intrusive /r/ given its traditionally stigmatised character, some speakers suppress linking /r/ as a type of hypercorrection (e.g., Brown, 1988a; Cruttenden, 2014; Gimson, 1962/1980; Mompeán, 2022; Wells, 1982; Wells & Colson, 1971). Nonetheless, the hypothesis concerning a hypercorrection effect could not be tested in the present study.

Several sociolinguistic, linguistic/phonetic, and usage-based factors that might condition variability in the production of linking /r/ were also analysed. Since some of these factors are discussed and analysed in previous studies, this provides a point of comparison between the current study and the previous literature.

In sociolinguistic terms, the findings show that linking /r/ is remarkably common in contemporary RP for both males and females, since half of the speakers have a ratio of production above 80% (females), and the other half have ratios in the 59%-70% (males) range. The high rates found in this dissertation might be due to individual factors and the fact that the speakers analysed are, on average, higher r-linkers than those in previous studies. The study by Mompeán (2022) on the speech of Queen Elizabeth II found that she produced linking /r/ at a rate of 76.3%, which underscores the individual variation that there may be among speakers in the use of one and the same connected speech process.

As mentioned previously, in the case of linking /r/, some recent corpus-based studies on TV newsreaders, radio correspondents, and newscasters RP speakers (e.g., Hannisdal, 2006; Mompeán & Gómez, 2011, Pavlík, 2016) show rates of linking /r/ that are lower (59.8%, 66.5%, 62%, respectively) than QEII's linking /r/ rates (76.3%) (Mompeán, 2022) - see also 2.1.3, and the current study (73.3%). This seems to point to the fact that, not only QEII -at least, at individual trend level- may be more of an /r/-linker, but also UK politicians, at group level, seem to be more-than-average /r/-linkers when compared to TV newsreaders, radio correspondents and newscasters RP speakers.

There are many explanations as to why individuals belonging to the same speech community (i.e., RP speakers) resemble group trends or norms in their language choices (e.g., Mompeán & Gómez, 2011, Pavlík, 2016) or differ, although slightly, from them (i.e., the current study vs. previous corpus-based studies). For instance, as mentioned in previous studies (e.g., Mompeán, 2022), in the case of /r/-sandhi usage, it can be explained in terms of language ideology, which is commonly said to play a critical role in shaping and influencing linguistic choices (Silverstein, 1979). As a case in point, when examining group level trends across the studies mentioned earlier and the current study, the RP speakers in this study belong to an upper-class social background, in contrast to other RP speakers (e.g., Hannisdal, 2006; Mompeán & Gómez, 2011, Pavlík, 2016) that possibly belong to a middle-to upper-class social background. This might probably

explain their higher rate of production that seems to justify a social stratification pattern of speech regarding the use of linking /r/.

Alongside individual variation, the differences in rates of production across gender reveal the existence of certain variation as well. The assumption that the use of linking /r/ is the prestigious variant and that women tend to use higher percentages of prestige or standard forms than male speakers (e.g., Coates, 2004) led to expect higher scores for the female speakers (e.g., Bauer, 1984, Hannisdal, 2006; Hudson, 1996; Labov, 1990; Mompeán & Mompeán-Guillamón, 2009). In the current study, female speakers stand out as the ones with scores over 80% whereas male speakers have the lowest scores of linking /r/. Female 2 (Yvette Cooper), for example, has the highest linking /r/ usage whereas Male 2 (Ed Balls) has the lowest linking /r/ usage (see Table 4 in 4.1.1.). The difference was found to be statistically significant.

The findings in this study are, therefore, in contrast with previous studies that found no correlation between the production of linking /r/ and gender and studies where men exhibited a higher production of linking /r/ (e.g., Bauer, 1984; Hannisdal, 2006; Hay & Sudbury, 2005; Mompeán & Mompeán-Guillamón, 2009; Pavlík, 2016). One reason for the discrepancy between the current study and previous studies regarding gender may also relate to the type of speech delivery. Previous studies have analysed the scripted speech of newscasters or TV presenters (e.g., Hannisdal, 2006; Mompeán & Gómez, 2011; Pavlík, 2016). In this type of data, both males and females may feel some pressure to use standard varieties in broadcasting when addressing the general public. In contrast, the relatively more relaxed setting of a conversation, despite also being broadcast on TV, may make some speakers produce linguistic forms more similar to their everyday use thereof. If this is the case, then the three female politicians may have used a higher rate of prestige forms (in this case, linking /r/) than males, whose conversations may mirror their everyday speech more accurately.

Related to the differences in rates across gender, one possible justification relies on the fact that female speakers exhibit a higher usage of prestige forms than male speakers as well as avoid stigmatised forms more regularly than men, mainly in formal speech (i.e. the relatively formal context of the interviews analysed in the present study) (Coates, 2004). The fact that women's production corresponds to 81% and above (Female 1, Theresa May, and Female 3, Harriet Harman, both have 81.3%, and Female 2, Yvette Cooper has 86.3%) while men's highest production corresponds to 66.7% (Male 1,

George Osborne has 66.7%, followed by Male 3, David Cameron, 65.3%, and Male 2, Ed Balls, 59%) might correlate with the use of /r/-sandhi (or linking /r/, more specifically) at gender level.

It is not surprising, therefore, that, as mentioned previously, women use a high proportion of a prestige form, such as linking /r/, which is supposed to be the norm, and not the exception at potential linking /r/ sites. The study by Mompeán (2022) provided high rates of linking /r/ as well (76.3% in QEII's speech, which is even higher compared to other group level trends of RP speakers -see 2.1.3.), which might point to the fact that women in positions of power/authority may be more of /r/-linkers than male RP speakers. The use of prestige forms in their speech can be a reflection of their authority and leadership skills in public life. Women in these studies (Mompeán, 2022, and the current study) represent public figures, as compared to women participants in previous studies, who, although being easily recognised because of their TV and/or radio contributions (e.g., Hannisdal, 2006), might not feel as pressured when speaking to an audience as QEII or the female participants in the current study.

Co-variation of language and gender in the present study seems to correlate, therefore, to social class, and maybe age, to a certain extent (i.e., all participants are middle-aged). Since it can be reasonably argued that the use of one specific linguistic structure or another might be directly related to women's position in society, and to women's linguistic usage, which probably explains women's linguistic choices. Following Coates (2004), this might explain why the language of women differs from the language of men. But as to why gender establishes a difference in terms of language use, it is a sociolinguistic pattern to be studied in depth in future research. Therefore, gender, or in other words, women as a social group, should not be ignored as a potential sociolinguistic factor to explain inter-speaker variation and to show the interrelationship of language and society.

As for linguistic factors, one interesting finding relates to the content (C) vs. function (F) grammatical status of the linking and linked words. Linking /r/ was found to be more frequent in combinations of FF (83.5%), and FC (75.2%), whereas the rate of production is slightly lower in CF (66.9%) and CC (67.1%). These findings are in line with previous studies where the r-link is particularly common between function words (e.g., Bauer, 1984; Foulkes, 1997; Viollain, 2014), but in contrast to Hay and Sudbury (2005), Pavlík (2016) and Mompeán (2022) who found little or no evidence concerning

word class. One possible explanation for this may be the prosodic structure since function words are usually unstressed and carry little information value (Hannisdal, 2006). Another possible factor contributing to the high frequency of FF, FC, and to a lesser extent, CF combinations may be the high corpus frequency of many set phrases or collocations of those word combinations including, among others, *there is* and *there are* (FF), *for example* and *our economy* (FC), or *number of* and *year ago* (CF), and in which linking /r/ is sometimes categorical or nearly categorical.

Finally, the finding that lexical and collocational frequency has an impact on the production of linking /r/ underscores the role of frequency of occurrence in the entrenchment or lexicalisation of a connected feature such as linking /r/ as part of the lexical collocation (see Bybee, 2001). This is most evident in the case of frequent set phrases or expressions (e.g., *for example/instance*) or highly frequent collocations (e.g., *the/a number of*), where linking /r/ was found to be categorical or nearly categorical.

One very relevant finding relates to the prosodic environment of the linking and linked words. The findings indicate that, in line with previous empirical research, UU word combinations represent the stress pattern where linking /r/ is most likely to be produced (86%). Besides, these results suggest that the presence of a stressed linking syllable followed by an unstressed linked syllable favours linking /r/ (68.3%), in line with previous research (e.g., Mompeán, 2022). The data regarding US word combinations (60.8%) and SS word combinations (42.5%) seem not to strongly favour linking /r/ (see Table 16 in 4.1.4.), which is also in line with previous research (e.g., Cox et al., 2014; Foulkes, 1997; Hannisdal, 2006; Hay & MacLagan, 2010; Mompeán & Gómez, 2011; Pavlík, 2016). The findings indicate that a prototypical environment for linking /r/ would be in a UU prosodic site and least likely of all is an SS site, which according to Giegerich (1999) coincides with a left-dominant foot boundary in English. As mentioned above, the presence of a stressed syllable after the potential linking /r/ also correlated positively with the use of glottalisation instead of linking /r/ as a hiatus-breaking strategy.

This fact that the production of linking /r/ is not categorical and the role of frequency of occurrence in its variability seems particularly relevant for the theoretical interpretation of the overall findings of the current study. From a usage-based framework, modelled exemplars of remembered sequences would contain information about the context in which they usually appear (e.g., Hay & MacLagan 2012; Hay & Sudbury, 2005; Mompeán, 2022; Mompeán & Mompeán-Guillamón, 2009; Pavlík, 2016). Following the

specialised literature, the findings may indicate that underlying lexico-phonological forms stored in the speaker's mental representations model the surface phonetic output. A usage-based approach would, therefore, consider high-frequency sequences to be more entrenched in terms of morphosyntactic structure (e.g., Bybee, 2001). Within this view, surface r-link forms would not be generated from underlying representations due to deletion or insertion rules (see 1.2.5. and 2.1.2.) but rather stored with higher or lower entrenchment given the frequency of occurrence, itself partly determined by other factors (e.g., phonetic environment).

These usage-based findings contribute to the debate regarding the phonological modelling of linking /r/ variation in non-rhotic accents of English. While most traditional analyses claim that there is no underlying coda /r/ present in word-final positions and that /r/ arises as a function of /r/-insertion (e.g., Vennemann, 1972), later approaches state that there is an underlying /r/ present word-finally but that an /r/-deletion rule operates if /r/ occurs before a consonant or pause (e.g., Mohanan, 1986). However, other phonological accounts assume that both rules (i.e., /r/-insertion and /r/-deletion) can actually co-occur in the synchronic phonology of varieties (e.g., Antilla & Cho, 1998). All these approaches, however, exclude considerations of functions and usage because of their aim is the synchronic model of competence. Therefore, they fail to predict inter-/intra-speaker variation (Mompeán, 2022).

The findings confirm the hypotheses that a number of factors may potentially condition /r/-sandhi usage. Mostly, phonetic, linguistic, sociolinguistic, and usage-based factors, although the weight of each factor is variable. Regarding phonetic factors, the current thesis has shown that the prosodic environment predicts the use of linking /r/. The results suggest that the presence of unstressed linking syllables and the absence of stress in the linked syllable favour the production of linking /r/. Besides, SS, SU, and US prosodic sites are potential, yet not categorical, linking /r/ sites, while UU prosodic sites are the least likely to trigger linking /r/. As to the impact of linguistic factors in the production of linking /r/, the results show that FF, FC, and CF exhibit higher rates of linking /r/ production that may correlate with the effect of prosodic stress and the use of schwa in function words, which are typically unstressed.

Turning now to the usage-based factor, the findings indicate that lexicalisation or entrenchment of /r/ in frequent set phrases or collocations underscores a clear effect of frequency on the use of linking /r/. From a usage-based viewpoint, frequent set phrases

or word-collocations in potential linking /r/ contexts have higher chances of lexicalisation or entrenchment, and therefore, of being produced with /r/. If particular word-collocations, or modelled exemplars, are recurrently used, and consequently stored containing /r/, speakers' mental representation will contain many memories of that word-collocation produced with /r/, a robust /r/ production will result (Cox et al, 2014; Hay & MacLagan, 2010, 2012; Mompeán, 2022). However, high lexical frequency leads to items being more resistant to change, and linking /r/ in a frequent set phrase or word-collocations can be interpreted as the strengthening of /r/ representations in memory of the "historical /r/ which continues to resist /r/-loss in British English" (Pavliik, 2012, p. 112).

4.2.2. Study 2: t-tapping

The results obtained for t-tapping indicate that the phenomenon is not very frequent in spontaneous speech in RP, as instantiated by the AMS corpus (29.2% in the whole AMS corpus – see Figure 23 in 4.1.2.), at least for the spontaneous speech of the six politicians analysed in the present dissertation. Although several interpretations of the t-tapping phenomenon suggest that it is spreading in British English (e.g., Hannisdal, 2006), t-tapping does not seem to be fully established in RP as a realisation of alveolar plosives /t, d/ in rapid or spontaneous speech across word boundaries. The relatively low rate of t-tapping suggests that, as a connected speech phenomenon, the process is a recent development in RP. This is also consistent with the fact that t-tapping has traditionally been ignored or referred to only marginally in classic descriptions of the RP accent.

The analysis has also revealed that, while the use of alveolar stops [t, d] is the default realisation of /t, d/ when a tap is not produced, as expected, there is also some use of glottal stops in cases in which /t/ is word-final at the boundaries of potential t-tapping contexts (e.g., *what if*). The results reveal that the glottal stop variant is used in 10.6% (135 cases) of potential t-tapping in which /t/ is the word-final consonant.

The use of glottal stops in intervocalic positions across word boundaries has been described previously for contemporary RP (e.g., Hannisdal, 2006) -as well as other accents of English such as NZE (e.g., Holmes, 1994) and contrasts with traditional descriptions of RP where such use was described as stigmatised or simply avoided (Wells,

1982). This use of glottal stop in intervocalic positions, as is the case of t-tapping, can also be considered to be a recent innovation in RP. According to Cruttenden (2014), some RP speakers (or GB speakers, in his terminology) may replace /t/ with a glottal stop “before words beginning with vowels, e.g. *not on* [nɒ? ʔɒn], *it opens* [ɪ? ʔəʊpmz]” (p. 184), which “was until recently stigmatised as non-GB” (ibid.) Use of the glottal stop in word-internal intervocalic position (e.g., *water* [ˈwɒ?ə], however, remains stigmatised in RP.

The non-marginal presence of t-tapping in RP, however, is sufficient to indicate that the phenomenon is spreading across varieties of English and it is not only a feature of varieties from the Southern Hemisphere (e.g., Holmes, 1994; Tollfree, 2001), or American English (e.g., Patterson and Connine, 2001; Zue and Laferriere, 1979), where it is generally recognised as a standard variant, but also of varieties of English in England (e.g., Amos, 2007; Badia Barrera, 2015).

Despite the likely status of t-tapping in RP as a recent innovation, the use of t-tapping in the AMS corpus might notwithstanding be less frequent than in the broader community of RP speakers in spontaneous speech. Several factors might point in this direction. One such factor may relate to the actual degree of formality of the interviews in the corpus. As in other varieties of English such as NZE (Holmes, 1994), alveolar plosive variants would be expected to be more likely to occur than taps in the speech of newsreaders. A conversational setting in which politicians (not newsreaders) chat with a journalist and where speech (at least the politicians’ speech) is supposed to be spontaneous might seem to favour the use of t-tapping. However, the low rate of t-tapping could relate to the interviews taking place in a relatively formal setting for the speakers, who are not only on camera but also talking about political topics to a general audience as professional politicians.

Given that t-tapping is often associated with American English and popular broadcasting (Alderton, 2022), the level of informality that the speakers may have wished to convey may therefore have been somewhat limited. Given that the use of t-tapping in RP has been described as indexing a combination of authority and informality by elite speakers to assert themselves from a position of privilege while maintaining an image of openness and approachability (Alderton, 2022). A trade-off between the desire to express both formality and informality as well as professionalism and approachability may thus

have resulted in a rate of t-tapping that could be higher than in formal, scripted speech but lower than fully informal, conversational speech.

Another factor suggesting that the rate of t-tapping found in the current study may be lower than expected could be related to the age of the speakers. The six speakers analysed in this dissertation range from 39 to 67 years of age, according to the different years of production (see Table 1 in 3.3.1.). They are, therefore, middle-aged. The use of t-tapping among middle-aged speakers has been suggested to point to some very specific social meaning. According to Alderton (2022), the speech of self-satisfied middle-aged male RP-speaking disc jockeys from the British TV comedies *I'm Alan Partridge* and *Harry Enfield's Television Programme* contain taps but these seem to be “highlighted for comic effect” (p. 290). While Alderton points out that not all these examples derive from naturally occurring speech, taken together, they seem to indicate that t-tapping “may index a kind of trans-Atlantic sophistication that can be used by (especially male) SSBE speakers to construct their desired personae” (ibid.).

It can be hypothesised that if younger speakers had been analysed, the rate of t-tapping production could have been higher given that there is some evidence that young speakers may use t-tapping more often than older speakers. Holmes (1994) found that t-tapping was considerably more favoured in young NZE speakers than in middle-aged NZE speakers. In the case of RP, Alderton (2022) analysed the speech of 45 RP speakers aged 16-19 and their pronunciation of the items *sort of*, *little*, and *pretty* using the tap variant was even lower (*sort of*: 8.4%; *little*: 10.5%; *pretty*: 7%). Badia Barrera (2015), who looked at data from sociolinguistic interviews of 20 teenagers, aged between 13 and 17 as well as data from 15 older speakers aged 27, found an average rate of 9.5%. These even lower rates suggest that a younger age may not only be correlated with a higher rate of t-tapping in RP. Interestingly, both studies point out the increasing use of t-glottaling in intervocalic position in RP as the main variant when alveolar plosives are not used (12.1% and 11.2% in Alderton (2022) and Badia Barrera (2015), respectively), despite the traditional stigmatisation of its use in RP (e.g., Badia Barrera, 2015; Hannisdal, 2006). This trend has been observed in young RP speakers for a long time now (e.g., Fabricius, 2000). Elsewhere in Britain, the participants in Amos (2007) were young and old speakers of Mersea English, and this study also shows that tap realisations are the second most frequent variant after glottal stops, since t-glottaling was preferred by the young

generation over t-tapping, whereas for the older generation taps are more prevalent in their speech.

Badia Barrera (2015) also points out an interesting fact. In her study, taps were mainly found in word-final pre-vocalic, word-medial intervocalic and word-medial syllabic /l/ contexts, three phonetic contexts traditionally regarded as stigmatised positions for a glottal stop in RP. Therefore, according to Badia Barrera, “the choice for taps by certain young RP speakers in these environments would indicate that they are trying to find an “in-between” variant halfway between the “stigma” of the glottal stop [ʔ] and the “stiffness” of the full alveolar [t]” (p. 201). In the current study, the use of glottal stops in intervocalic position is marginal or even unattested, which is the expected pattern among traditional RP speakers. The fact that the speakers analysed for this dissertation are middle-aged and relatively conservative in their pronunciations, suggests that they may have used t-tapping both to signal informality while at the same time avoiding glottal stops.

While the relative formality of the speech context or the speaker’s age alone provide some explanation for the use of t-tapping in the current study, further factors were explored that could potentially motivate their use, as is the case of the other two connected speech phenomena analysed in this dissertation. These include linguistic factors that involve the grammatical status of the word (i.e., content vs. function word) as well as word-class collocations; sociolinguistic factors, such as gender; usage-based factors that involve lexical and collocability frequency; and phonetic factors, such as the prosodic environment.

Regarding the grammatical status of words (content vs. function words), the findings indicate that monosyllabic function words favour t-tapping production (e.g., *it*, *that*, *but*, *what*, *etc.*), with a relatively high frequency in word collocations such as *it is*, *that it(’s)*, *but I*, *what I*, *but I*, *etc.*, and in frequent word collocations (e.g., *(a) lot of*,) as well. Unlike the rate of production of t-tapping involving monosyllabic function words, the production of t-tapping is disfavoured in monosyllabic content words, such as *not* and *get* (e.g., *not a/an*, *not about*, *not actually*, *get a*, *get any*, *get invited*, *etc.*). Word collocations containing *not* are categorically disfavoured, while word collocations that contain *get* slightly disfavour t-tapping.

One of the main reasons for monosyllabic words to be more likely to favour t-tapping production is because they are shorter and more frequent than polysyllabic or complex words, a pattern which is supported by other studies that evidenced that t-tapping was more frequent in morphologically simple words (e.g., Patterson & Connine, 2001). Furthermore, morphologically simple words are also normally unstressed, and therefore, more likely to undergo phonetic reduction in connected speech (Hannisdal, 2006).

Regarding the grammatical nature of the words flanking the t-tapping site (word-class collocations), taps are found to be disfavoured in CC word combinations (12.5%), particularly word combinations containing long content words (e.g., *right approach*, *put interest*). T-tapping is found to be more frequent in FF word combinations (e.g., *but I*, *it is*) and FC word combinations (e.g., *that actually*, *at all*) and somewhat infrequent in CF word combinations (e.g., *lot of*, *had a*) (39.3%, 32.6, 17.2%, respectively; see Table 16 in 4.1.4.). The results obtained indicate that the production of t-tapping is likely to be avoided when the left-hand word flanking the potential t-tapping site is a content word, whereas two function words flanking the potential t-tapping site is the most frequent environment for t-tapping to occur.

Given that function words normally carry little information value, one possible interpretation for this usage pattern (i.e., preference for FF word combinations for t-tapping) could again be related to a process of phonetic reduction. If t-tapping is considered a reduction process (e.g., Shockey, 2003), a combination for the greater predictability of function words from the context (Bell et al., 2009), the absence of prosodic prominence, the fact that unstressed syllables are generally shorter, and the fact that phonetic reduction is generally greater in contexts of durational shortening (e.g., Riehl, 2003) may all add up to t-tapping being more frequent in FF word combinations.

Furthermore, t-tapping is mainly produced in monosyllabic function words because longer content words are normally stressed and carefully pronounced, which favours the production of standard variants (/t, d/) over the tap variant (Hannisdal, 2006).

In terms of sociolinguistic factors, the combination of the status of t-tapping with existing knowledge of women's speech patterns who tend to use more standard variants leads to expect lower scores for female speakers (e.g., Hannisdal, 2006; Hudson, 1996; Labov, 1990). Indeed, previous studies on t-tapping found that men use more t-tapping productions than women (Hannisdal, 2006; Holmes, 1994). One possible explanation for

the two different realisations of t-tapping (i.e., the tap variants vs. the voiced/voiceless plosive) is that they may have different social status and prestige. For instance, the tap variant is commonly associated with young speakers in varieties of English other than RP, such as NZE (e.g., Holmes, 1994) and it is also associated with young speakers and young adults in RP (e.g., Badia Barrera, 2015).

These findings reveal that speakers group particular sounds into mental categories (i.e., function words as carrier tap words), a cognitive procedure that is sensitive to orthography. Mental categories of the lexicon match the access to lexical representations of previously experienced or remembered pronunciations, which have been automatised for particular grammar/morphological categories (Eddington, 2007).

The individual frequency obtained for the production of t-tapping in the present sample indicates that all the speakers have scores below 40%. The highest scores are above 35% and correspond to two of the three male speakers, while the other male participant (Male 2, Ed Balls) has a score below 25% (22.3%). Females represent a more consistent group with a score ranging from 22% to 30% (see Table 7 in 4.1.2.). The percentage of production for men is 9% higher than females. The results from the present study, therefore, indicate that females stand out as the ones with the lowest score -below 30%- whereas males have the highest scores -above 30%. Then, concerning the difference between males and females in the production of the tap variant, female speakers as a group show a slight preference for standard variants over tap pronunciation.

In terms of usage-based factors, lexical and collocability frequency were analysed to account for the degree of variation in t-tapping production, since word frequency has long been claimed to account for word shortening or reduction processes such as t-tapping (Gregory et al., 1999).

In line with relevant studies (e.g., Patterson & Connine, 2001), t-tapping occurs more frequently in high-frequency function words (e.g., *it, that, what, get, but, don't, etc.*), notably in word-class collocations such as *that it, that is, that I, I don't, what I, but also, but again, but I, it is, etc.* Taps are fairly uncommon, although not inexistent, in some high-frequency content words such as *need (need a), instead (instead of) rid (rid of) set (set out)*. The production of t-tapping is, however, less likely to occur in bisyllabic or polysyllabic and less frequent content words or word collocations, such as *deficit (deficit in/is), and budget (budget is)*.

One possible explanation for the occurrence of t-tapping in high-frequency function words might be related to the fact that reduction processes, such as t-tapping, affect the most common words first, through the process of lexical diffusion. Following Bybee (2002), sound change occurs in a gradual fashion and diffusion patterns affect normally high-frequency words in the lexicon first. In her view, sound change is mainly motivated by articulatory changes, which involve weakening in the case of t-tapping. In considering t-tapping as a reduction process, and therefore, a process that is sensitive to frequency, its linguistic production can be considered automatised and prone to minimise articulatory effort, which is reflected in the use of particular linguistic forms and their cognitive representation. Such cognitive representation results in high-frequency items exhibiting stronger mental representations and, therefore, easier access to reduced forms. Therefore, an exemplar model of phonological representation allows for sound change of high-frequency words to be both phonetically and lexically gradual.

The lexical analysis has identified that, in line with previous research, t-tapping is more commonly produced in high-frequency function words (e.g., Badia Barrera, 2015; Connine, 2004; Hannisdal, 2006; Patterson & Connine, 2001), and that t-tapping production seems to be primarily constrained by lexical diffusion occurring in high-frequency words and contexts, in contrast to previous studies where frequency did not have any impact (e.g., Gregory et al., 1999). T-tapping exhibits a similar distributional pattern in RP as in other accents of English where it is a well-established feature, such as American English since high-frequency words and morphologically simple words exhibited a higher ratio of t-tapping production than low-frequency words and morphologically complex words (e.g., Patterson and Connine, 2001).

In line with Holmes (1994) and her study of NZE speakers using t-tapping in intervocalic position, her analysis showed that the lexical item *sort of* is approaching a semi-categorical production (16 actual cases out of 21 potential cases in the present sample), which support the process of lexical diffusion that t-tapping is undergoing for high-frequency items, especially /t/ across boundaries. Other examples that support the statement that the production of t-tapping in the present study is affected by lexical and collocational frequency are the use of the third singular pronoun *it* and the conjunctions *but*, *that* and *because*, as well as the negative contraction *don't*. For instance, *but I* represents the word collocation with the highest number of actual cases of t-tapping (41 actual cases), followed by *that is* with 38 actual cases.

However, a word collocation such as (a) *lot of*, which could be considered to be frequent in spoken language use has a relatively low number of actual cases produced (8 actual cases out of 27 potential cases) (see Table 10 in 4.1.2.). The reason why t-tapping seems to be disfavoured in this word collocation might be that *lot* is functioning as a quantifier rather than a noun.

In terms of phonetic factors, the prosodic environment is a relevant aspect involving the production of t-tapping. In this respect, t-tapping tends to be avoided in stressed environments (e.g., *put interest, said all, get any*), since stressed syllables (S) need more articulatory force than unstressed (U) ones and are more likely to be preceded by a glottal onset, which disfavours the production of t-tapping (Hannisdal, 2006). In line with previous empirical research (e.g., Eddington & Elzinga, 2008; Hannisdal, 2006, Warner & Tucker, 2011), UU word combinations represent the stress pattern where t-tapping is most likely to be produced (38.6%). The results also suggest that the presence of SU and US word combinations slightly favour t-tapping (27.6% and 22.8%, respectively), in line with previous research (e.g. Eddington & Elzinga, 2008; Hannisdal, 2006, Zue & Laferriere, 1979). The data also suggest that SS word combinations (7.4%) seem not to favour t-tapping (see Table 21 in section 4.1.4.), in contrast to Holmes (1994) in NZE. The findings indicate that a prototypical environment for t-tapping would be in a UU prosodic site and least likely of all in a SS site. The findings indicate that tap productions are blocked in foot-initial position but are frequent in non-foot-initial contexts (Tollfree, 2001).

One possible interpretation for these findings could be that absolute word-initial vowels often resist reduction processes when stressed in English, as in the SS prosodic environment cases. These typically ‘strong’ vowels (as opposed to ‘weak’ vowels, such as schwa) (Cruttenden, 2014) are typically more stable and less subject to reduction compared to vowels in other positions within a word (e.g., unstressed syllables after stressed ones). This interpretation alone does not explain, though, why US prosodic environments, have a rate of production of 22.8%, which is well above that of the UU pattern (i.e., 7.4%). This may point to the importance of function words and their tendency towards phonetic reduction in the use of t-tapping in spontaneous speech which may override the effect of word-initial stressed vowels.

If this above interpretation is correct, an interaction between grammatical factors (word class) and prosodic information may potentially affect t-tapping production.

Interestingly, the SU and US prosodic environments have relatively similar rates of t-tapping (27.6% and 22.8%, respectively). Since function words tend to consistently co-occur with content words from specific word classes (e.g., determiners such as *a* or *the* co-occur with nouns, while pronouns such as *they* or *he* tend to co-occur with verbs (de Carvalho, He, Lidz & Christophe, 2019)), speakers could use statistical/distributional information in language use in learning. A relatively high frequency of SU and US word class combinations or lexical collocations help in the entrenchment of such patterns, and reduction processes such as lenition operate more frequently in SU and US prosodic environments than in SS environments, where the frequency of lexical collocations may be lower.

4.2.3. Study 3: yod coalescence

The results obtained for yod coalescence indicate that the phenomenon in spontaneous speech is not very frequent in RP, since the percentage of production corresponds to 32.4% in the AMS corpus (see Figure 28 in 4.1.3.). As a preliminary observation, it is worth noting that the number of tokens containing a potential yod coalescence is notably lower in comparison to linking /r/ and t-tapping. This, in turn, made the quantitative analysis more limited, although the data were still large enough to reveal some important findings in the use of yod coalescence in RP in spontaneous speech.

The data indicate that yod coalescence is not favoured over the uncoalesced variants (/t, d, s, z/ followed by /j/), as yod coalescence is produced in ca. 35% of the cases in the whole AMS corpus, so the phenomenon is moderately low (i.e., one-third of all potential cases). This indicates that the production of yod coalescence as a variant across word boundaries is still in progress in RP.

Various factors may be called upon to explain the low production of yod coalescence including the relative formality of the speech context, the stigmatisation of the phenomenon in terms of its morphosyntactic distribution, or the speaker's language awareness.

The results support the claim that yod coalescence across word boundaries is not exclusive of either casual speech or non-standard accents of English (e.g., Estuary English) (Wells, 1982). The current study analysed conversational, spontaneous speech,

so the use of yod coalescence would seem to be unsurprising. However, the relative formality of the speech situation along with the social status and language awareness of the speakers. A low rate of production of yod coalescence forms is therefore far from categorical, which could be a direct consequence of the tendency among some speakers to avoid stigmatised forms of language. However, Hannisdal (2006) found that the rate of yod coalescence amounted to 46.4% of the cases in broadcast scripted speech, so the speech-conscious context of broadcasting (news in the case of Hannisdal's study and political interviews in the current study) do not prevent the use of yod coalescence. Taken together, the results of the present and those of Hannisdal's indicate that yod coalescence is found in both spontaneous and scripted speech in RP and that the rate of production may be around 35%-45% in this accent. Incidentally, these rates are also similar to those found in Amos' study of conversational interviews with Mersea English speakers, where /t, d/ + /j/ coalescence was produced between 31%-51%.

The relatively low percentage of yod coalescence may be linked to its stigmatisation of being often regarded as a vulgarity when produced word-internally in stressed syllables (e.g., Wells, 1990), which may be carried over to the word-boundary contexts. Despite some claims that yod coalescence is a well-established process word-internally (Hannisdal, 2006; Wells, 1994a), speech-conscious speakers who deliberately avoid producing yod coalescence word-internally may avoid it across word boundaries as well, as a type of hypercorrection. This may explain the low rate of production even though, from a coarticulatory point of view, coarticulation would seem to facilitate coalescent assimilation.

The results also show that the production of yod coalescence was constrained by a number of linguistic factors, such as the grammatical status of the word (i.e., content vs. function word), and word-class collocations; sociolinguistic factors, such as gender; usage-based factors that involve lexical and collocability frequency; and phonetic factors, such as the prosodic environment.

Regarding linguistic factors, yod coalescence is found to be more frequent in CC (49.7%), while it is slightly disfavoured in CF (33.3%) and FF (27.4%) word collocations, and the lowest rate in FC (25%) word combinations (see Table 16 in 4.1.3.). Some examples of FF word combinations containing actual yod coalescence realisations are *and you, because you, as you, what you*, etc. Some set phrases where yod coalescence is frequent in CC and CF word collocations are *last year, next year, is you*, or in compounds,

such as *trade union* (see Table 13 in 4.1.3.). The pronoun *you* is by far the most frequent item in the AMS corpus, and it is one of the function words with potential yod coalescence pronunciation. Indeed, there are altogether 164 word combinations that involve the clitics *your* and *you* and 45 word combinations that involve the use of *you'll*, *you're*, *you've*, and *you'd*, which may have inflated to some extent the number of CF or FF cases.

A possible explanation why CC word combinations have the highest percentage of production for yod coalescence might be that some of these word collocations contain adjective-noun, verb-noun or noun-noun combinations with the noun *year*, which could be frequently found in a speaker's everyday lexicon (e.g., *good years*, *next year*, *recent years*, *last year*, *takes years*, *wait years*, *hundred years*, etc.).

The findings, which are in line with previous studies (e.g., Glain, 2012), may indicate that the content words found in the sample refer to either technical/non-everyday lexicon or proper nouns (e.g., *government*, *market*, *UKIP*, *interests*, *taxes*, *inspectorate*, *conservatives*, *allowances*, *Keynes*, *crisis*, *referendums*, *manifestos*, *Parliament*). These findings are in line with previous studies where yod coalescence was particularly uncommon regarding specialised or technical vocabulary (e.g., Glain, 2012), which could be linked to the principle of lexical diffusion in which high-frequency words are subject to change first (e.g., Bybee, 2001).

Concerning the linguistic constraints for yod coalescence, the grammatical pattern that has been identified as a relevant factor affecting variation is the grammatical status of the words involving the potential yod coalescence site. Yod coalescence is more commonly found when both segments are content words but it is quite frequent when the word-class collocation contains a high-frequency item such as the pronoun *you*, which is remarkably common to undergo coalescence. However, it is worth noting that even though most of the word collocations found in the sample contain the pronoun *you* (e.g., *that you*, *and you*, *as you*, *but you*, *what you*, etc.), none of the word combinations containing *you* exhibit a categorical production of yod coalescence.

Previous empirical studies also investigated correlations between the production of yod coalescence and several sociolinguistic factors, such as age, gender or class (see 2.4.). Gender is expected to have an impact on yod coalescence production as it conveys sociolinguistic salience. The individual frequency obtained for the production of yod coalescence indicates that there is a slight variation, although fairly consistent across genders (males vs. females), in the rate of production between the participants in the

present sample. Half of the speakers have scores above 35%, while the other half have scores between 25% and 28%, and the average score is 32.4% (see Table 11 in 4.1.3.).

In the current study, male and female speakers exhibit similar rates of production for yod coalescence between 30% and 35%, which indicates that all speakers in the sample use yod coalescence to a certain extent. The finding is in line with Hannisdal's (2006) study where both men and women exhibited a similar use of coalesced forms. Nevertheless, these findings are in contrast to Altendorf (2003) in which female teenage speakers used yod coalescence in less than 10% of the cases, as well as previous studies on other varieties of British English, such as Mersea Island English (Amos, 2007), where a great deal of variation across gender was found. While the data obtained from the female speakers in the present study represent a ratio of production that ranges from 25 to 38%, in Mersea Island English there was a shift in production across generations and genders where a categorical or semi-categorical production was found for both genders in younger speakers whereas male speakers were the only ones who produced yod coalescence in the older speakers group.

The results show that there are no differences across gender and that the stigmatisation of yod coalescence could have influenced the relatively low level of production for both, males and females. It is, however, significant that all individual scores correspond to 25% or above. This finding suggests that a similar rate of yod coalescence usage for men and women may indicate "a shift in the evaluative status" of yod coalescence (Hannisdal, 2006, p. 216). The fact that there is a slight difference in the production between males and females can be considered to be a linguistic indicator of becoming a standard variant and losing stigmatisation, at least, across word boundaries.

In terms of usage-based factors, lexical and collocability frequency were analysed to account for the degree of variation of yod coalescence production. The findings show that certain combinations of words which exhibit yod coalescence may be related to a high rate of frequency of the clitic pronoun *you* (see table 13 in section 4.1.3.), as previously mentioned, which results in adopting a categorical pronunciation containing yod coalescence due to lexical entrenchment. These results may also suggest that the lexicalisation of certain lexemes affects their phonological production. Following the specialised literature, the findings may indicate that underlying lexico-phonological forms stored in the speaker's mental representations would model the surface phonetic output (e.g., Bybee, 2001). A usage-based approach would, therefore, consider high-

frequency sequences to be more entrenched in terms of morphosyntactic structure, and consequently, modelled exemplars would contain phonetic information about the context in which they usually appear (e.g., Bybee, 2001; Bybee & Beckner, 2009; Cooper & Paccia-Cooper, 1980; Shi et al, 2005).

A usage-based account of yod coalescence may explain its occurrence as motivated by coarticulatory factors which are subsequently processed by language users given some threshold frequency of occurrence of yod coalescence in specific word combinations leading to the entrenchment of coalesced forms and their use as an articulatory routine in subsequent uses of the same linguistic material. This process may also be sensitive to the high lexical frequency of certain items in the word combinations (e.g., the clitic pronoun *you*). The usage-based account accommodates the fact that high-frequency words tend to be shortened as compared to low-frequency words, not only in terms of lexicon but also during phonetic production (see Bybee, 1999, 2001; Fidelholtz, 1975; Gregory, 1999). In this respect, yod coalescence could also be considered a step in the direction of phonetic reduction of the word combinations with potential yod coalescence. This usage-based account is not incompatible with subsequent lexical diffusion or the gradual spread of the entrenched patterns through a network of words or forms in speakers' mental lexicon and the lexicon of a language.

Additionally, according to Recasens (2018), the degree of articulatory accommodation between the coda and onset consonants seems also to be conditioned by the frequency of occurrence of the lexical item that contains the phonetic segment (Bybee, 2001). Thus, coalescent assimilation in English /C#j/ sequences is more probably to occur when /j/ belongs to a highly frequent lexical item than a less frequent one (e.g., *you* vs. *unit*, Shi et al., 2005).

While many scholars have explored the assimilation of function words in phonological descriptions of different languages (e.g., Kaisse, 1985; Kohler, 2000; McCarthy, 1993), it is however, not clear whether the occurrence or non-occurrence is because of the grammatical category or frequency of usage, since the most common function words also exhibit high usage-based frequencies (Shi, Gick, Kanwischer & Wilson, 2005). Given the fact that function words generally occur fairly frequently, it is unclear whether the assimilation observed for function words in the findings of the current study is because they belong to a distinct grammatical status from content words or because of their high frequency.

In terms of phonetic factors, the prosodic environment was also analysed to observe their impact on the production of yod coalescence. According to the relevant literature (e.g., Altendorf, 2003; Amos, 2007; Glain, 2012; Hannisdal, 2006), the findings reveal that yod coalescence is more likely to occur in a SS environment (46.2%, e.g., *good years*), followed by US (31.6%, e.g., *at university*), UU (28.9%), and SU (26.6%). The findings indicate that a prototypical environment for yod coalescence is an SS prosodic domain and least likely of all an SU prosodic environment for yod coalescence to occur across word boundaries.

4.2.4. General discussion

The studies discussed above addressed two research questions: (a) RQ1: how frequent is the use of linking /r/, t-tapping, and yod coalescence in conversational language use?; and (b) RQ2: what factors condition variation in the production of linking /r/, t-tapping, and yod coalescence?

Regarding the first research question, it can be concluded that each of the processes analysed is used variably in connected speech (see 3.2.). Even though linking /r/ is the connected speech phenomenon that exhibits the highest production, it is used non-categorically. T-tapping and yod coalescence are two connected speech phenomena that, although some covariation may be expected due to sociolinguistic factors, present a relatively low ratio of production that confirms how variable they are in connected speech. The differences in rates of production across the three connected speech phenomena reveal the existence of a wide range of inter-speaker and intra-speaker variation, although this type of variation was not the focus of the current study given the small number of speakers analysed.

As for the second research question, it can be concluded that variation across the three connected speech phenomena studied is motivated (or less so) by –or correlates with– multiple linguistic (grammatical status), sociolinguistic (gender), usage-based (frequency), and phonetic (prosodic environment) factors, although the contribution of each of the factors or type of factors depends on the specific phenomenon. In some cases, a given factor has an impact on the rate of production of the phenomena, such as the grammatical status of the words (content vs. function) involved at the different production

sites. In other cases, there is no sociolinguistic correlation between the variable and usage patterns (e.g., gender). Moreover, some groups of factors may have a greater impact on some phenomena. Regarding t-tapping variation, for example, phonetic factors have a greater effect than linguistic or sociolinguistic factors. Finally, oftentimes, one variable has an impact on rates of production but the specific impact varies across phenomena so the weight of the factor is different across phenomena. The prosodic environment seems to have an impact on the three phenomena. However, while linking /r/ and t-tapping productions are most likely to occur in a UU environment, yod coalescence is most favoured in an SS environment, an environment where t-tapping is least likely to be produced.

The studies carried out in this dissertation have several implications for phonological theory, and the field of linguistics more generally. These implications encompass the study of lesser-studied connected speech phenomena, the accommodation of variation within linguistic theory, or the need for an update of linguistic descriptions of linguistic varieties.

The general aim of the current dissertation was to provide empirical evidence on the use of connected speech phenomena in conversational speech in real contexts. The analysis of language data in such contexts, therefore, has not targeted linguistic material produced *in vacuo* and decontextualized from its original setting but *in situ*, anchored to the context it was produced in. From the perspective of the study of connected speech phenomena, such an approach has allowed to gather information about phenomena in a specific language variety, RP, that have drawn the attention of linguists for some time (i.e., linking /r/) but that have generally gone ‘under the radar’ for this variety (i.e., t-tapping and yod coalescence). One contribution from this thesis, therefore, is an increase in our understanding of connected speech processes such as the above. In doing so, the thesis contributes towards building a stronger case for the study of connected speech processes in conversational speech in English.

The interpretation of the results within the general background of existing linguistic and phonological theories needs to take into account, therefore, the inherent variability in the use of connected speech phenomena such as linking /r/, t-tapping, and yod coalescence. This variability can be dealt with in various ways since the field of phonology is not highly unified with a shared body of principles, aims, and methods and seems more like a ‘loose federation’ than a homogeneous ‘community’ (Bermúdez-Otero,

2002). In the theoretical landscape of current phonological theory, it can be identified different approaches: rule-based, constraint-based, representation-based, and usage-based approaches (Balogné Bérces & Honeybone, 2020). Classical generative phonology, the most prominent representative of rule-based approaches, made little emphasis, if any, on variation, generally considered to be the results of phonetic implementation at the late stages of phonological derivation rather than of variable phonological rule application (e.g., Johansson, 1973 -/r/-insertion rule-; Vennemann, 1972 -/r/-inversion rule-; Gick, 1999; Mohanan, 1986 -/r/-deletion rule).

OT (e.g., Cutillas, 2003; Prince & Smolensky, 1993), the most popular approach among constraint-based ones, has increasingly paid attention to variation. Anttila (2004) and Anttila and Cho (1998), for example, underscore the importance of variation to the theoretical phonologist and how it can be treated within OT. Apart from rule-based and constraint-based approaches, variation has been dealt within representation-based approaches, such as Government Theory (e.g., Scheer & Nancy, 2017). One interesting approach regarding the interpretation of connected speech phenomena is that of Strict CV phonology, a subbranch of Government Phonology. Balogné Bérces (2004, 2008), for example, analysed English /t/-allophony and the distribution of /t/-allophones. In this analysis, a prosodic structure is made up of strictly alternating C and V positions together with an empty CV-span as the boundary-marker of words.

Finally, usage-based approaches in phonology (e.g., Bybee, 2001) may also be used to explain variation such as the one found in this dissertation. Within this more functional approach, words in the lexicon consist of a cloud of phonetic exemplars or phonetic variants that contain detailed information about the recurrent contexts in which they appear (e.g., Bybee, 1999, 2001). As storage implies a network of connections between the related items, following this usage-based framework, mental phonetic images are stored in a speaker's knowledge of the language, which operates by the influence of factors such as they type of token and collocation, lexical and collocability frequency (e.g., Bybee, 1999, 2001). These factors may result in lexicalisation or lexical entrenchment, which refers to a cognitive routinisation process of the linguistic units and structures based on repetitive events in language use (e.g., Bybee, 2001; Langacker, 1999; Mompeán, 2022).

Within this more functional, usage-based view, tokens containing linking /r/, tap productions and coalescence assimilation would be the result of how entrenched or

lexicalised these processes are in specific utterances. The more frequent a word-class collocation including a potential connected speech process is, the more likely it is that the process becomes lexicalised in it, and thus be produced.

This phonological framework helps to understand how speakers store and access their phonetic inventory of connected speech processes by using those exemplars that appear in the same or similar phonetic contexts more recurrently. This is supported by the word combination *for example* which is produced categorically in the sample, by the use of certain function words, such as the clitic pronoun *you* in yod coalescence pronunciations, or certain recurrent function words in word combinations found in the t-tapping study, such as *that, it, don't*. These findings explain that high-frequency lexemes and word combinations lead to higher rates of linking /r/, t-tapping and yod coalescence, which points to a lexicalisation of /r/ in high-frequency contexts and the automatisisation of tap productions and yod coalescence in high-frequency contexts as well (e.g., Bybee, 2001, 2006). However, the relatively low frequency of production of any type of word combination for yod coalescence, in contrast to linking /r/ or t-tapping, might be linked to the type of content words found in the AMS corpus (i.e., technical vocabulary related to politics). From this perspective, some of the findings from the present dissertation can be viewed as a substantiation of some of the observations of a usage-based phonological account.

Thus, the present study not only measured the production of different connected speech phenomena in a non-rhotic variety of English (RP) in conversational spontaneous discourse but also correlated the variation in terms of four different parameters or factors. This was made by using a corpus-oriented method which enabled to encompass the findings to the existing phonological literature since the examination of traditional and current approaches to phonology (see 1.1.5.) seem to point to the fact that traditional phonological theory often lacks -usually overlooks-, or sometimes generates rules whereby connected speech processes need a specific (re)interpretation depending on the stand of the approach. This occasionally leads to highlight that the mechanisms that operate in connected speech are usually unclear since neither all existing phonological theories nor the ones that describe connected speech are concerned with connected speech processes (i.e., there seems to be a lack of explicit predictors to explain the presence/absence of yod coalescence in the same way as there are many rules to account for linking /r/ production). Furthermore, predicting the insertion/deletion of a particular

rule for the production/absence of connected speech processes (e.g., rule insertion vs. rule deletion for linking /r/), traditional approaches seem to be somewhat uncertain to explain categorical applications of such rules.

Thus, the present dissertation tried to evaluate, using a more current approach, the weight of variation (and what motivations are implied) in spontaneous language use. Regarding the presence or direction of a factor and without ruling out any consideration (i.e., traditional analysis tended to overlook inter-/intra-speaker variation), a usage-based approach enables to understand in what way predictors have an impact on production (e.g., grammatical status, lexical and collocability frequency, prosodic environment) with a focus on language use. Therefore, one of the main implications of the present doctoral thesis aims to build knowledge on the multiple factors that are inherent to variation in connected speech regarding the presence of a meaningful association as well as the possibility to develop an updated phonological theory in the future which can predict variation more accurately.

Another theoretical implication emerging from the three studies is the need for the descriptive linguist to update their descriptions of linguistic varieties. More specifically, the results also underscore the necessity of updating descriptions of RP, a variety traditionally described impressionistically by linguists (e.g., Cruttenden, 2014; Wells, 1982) based on the empirical data from corpus-based studies and other types of empirical studies (e.g., Alderton, 2022; Hannisdal, 2006; etc.) Despite the fact that linguists have rarely paid much empirical attention to elite accents such as RP (Britain, 2017), these recent studies have contributed significantly to studying RP in a more similar way to the manner in which other linguistic varieties are analysed in fields such as sociolinguistics. Alongside other previous corpus-studies (e.g., Hay & Sudbury, 2005; Mompeán, 2022), the current study has shown, for example, that linking /r/ is less frequent than previously thought or traditional descriptions of RP seem to imply. Moreover, the study has also shown that t-tapping is now found in RP, particularly among young speakers (e.g., Alderton, 2022). The impact of different factors or variables in the use of connected speech phenomena is yet another important theoretical implication of this study for the description of linguistic varieties. Oftentimes, descriptions of varieties mention the use or presence of a given phenomenon but not the variants that may condition its variation (e.g., Bauer, 1984; Lewis, 1975/2014).

CHAPTER V: Conclusion, limitations of the study, directions for future research, and implications of the findings

This last chapter provides the conclusion and a general overview of the main findings obtained in the current doctoral thesis, the limitations of the present research and the directions for future work, as well as some possible applications of the results. The first section presents a conclusion of the key findings of the study that were previously correlated to the theoretical framework and empirical research. The second section presents the limitations of the current study and some suggestions for future research. Finally, some possible applications of the research findings are discussed in connection with areas such as foreign language teaching, speech technology and automatic speech recognition (ASR), or forensic linguistics, among others.

5.1. Main findings

This dissertation is an empirical investigation of the use of linking /r/, t-tapping and yod coalescence in the spontaneous speech of UK political figures while interviewed on TV. The present study has provided valuable insights into the three connected speech phenomena in contemporary RP regarding the factors that motivate variation in their use.

The present dissertation, then, concludes with the following main findings, which have both theoretical and empirical significance. The first main finding is that the use of the three connected speech phenomena is not categorical. Restating the first research question addressed in this study (see Chapter III), the results obtained have shown that they exhibit significant variability, which points to the need for any attempt to account for these processes in phonological theory to take variation into account. The existence of this variability is perhaps unsurprising since language use is full of variation of different sorts and connected speech phenomena, by their very nature, are likely to exhibit variation in their use. In fact, although a categorical decision is often made regarding the presence or absence of a given feature (e.g., linking /r/ vs. no linking /r/), a thorough phonetic analysis of actual productions often shows more phonetic variability. In the case of linking /r/, for example, hiatus may occur instead of a rhotic consonant. However, laryngeal strategies (glottalisation, creaky voice) are also used to resolve hiatus when linking /r/ does not apply.

While variation is inherent in the use of linking /r/, t-tapping and yod coalescence, the current study has also shown that the three processes do not have the same degree of variability in terms of production rates. One of the connected speech phenomena, linking /r/, is notably more frequent than the other two (t-tapping and yod coalescence), which may indicate its well-established use in RP (as well as other non-rhotic accents) while yod coalescence across word boundaries, and especially t-tapping, may be considered as more recent trends. Finally, a detailed analysis of the variation in the phenomena reveals interesting findings that point to linguistic tendencies in RP.

The analysis of the three connected speech phenomena reveals a surprisingly non-categorical use of linking /r/ while t-tapping and yod coalescence exhibit low overall scores. For instance, the findings show that, according to the relevant literature, linking /r/ seems to be the rule rather than the exception at potential /r/-sandhi sites. In a similar vein, the empirical analysis reveals that the low rate of t-tapping production suffices to point to the use of certain (more modern, ongoing) linguistic choices (taps) instead of traditional ones (/t, d/). As for the use of yod coalescence, which displays a low percentage of production, the absence of more yod coalescence forms might be linked to its traditional stigmatisation in RP.

The second main finding is that the production of the connected speech processes or the variants is conditioned –or at least correlates- with a variety of factors. In terms of factors that motivate the production of connected speech processes, for the speaker sample under study (middle-aged UK politics), a combination of linguistic (e.g., lexical status of words), usage-based (e.g., frequency) and phonetic factors (e.g., stressed/unstressed nature of the words flanking the potential connected speech phenomenon site) are found to make an impact on how variable the production of the different processes is, while the sociolinguistic variable of gender was not found to be as relevant. Thus, restating the second research question addressed in this study (see Chapter III), the results obtained have shown that the variation in the production of linking /r/, t-tapping, and yod coalescence is motivated by multiple factors. The specific weight of the different factors, however, is dependent on the specific connected speech process studied.

The study has also revealed that there is a high use of pure hiatus and of laryngeal strategies (glottalisation, creaky voice) –the latter are slightly higher- to resolve hiatus when linking /r/ does not apply, and that the use of the latter is lower in spontaneous speech than in scripted speech. Similarly, the analysis reveals that there is also use of t-

glottaling in /t/ contexts, although negligible, when t-tapping does not apply, and that there exists a difference in /t/ and /d/ contexts at potential t-tapping sites. The production of yod coalescence, in similar trend to that of t-tapping, reveals an unsurprising low percentage of production.

While the production of t-tapping and yod coalescence differs in some respects from the trend that is evidenced in the analysis of linking /r/, the production of t-tapping and yod coalescence reveals that they are not connected speech phenomena that can be considered exclusive from non-standard accentual varieties of English or colloquial speech contexts. Even though the lack of a higher production for these two phenomena (t-tapping and yod coalescence) indicates that their standardisation is still in progress, the relatively low percentage of production is unsurprising given that the unconscious level of formality adopted by the speakers in the sample.

One interesting finding regarding the different factors studied is the fact that the prosodic environment is often very relevant for the production of a connected speech process or its absence. The same can be said of word class collocations as well as lexical collocability frequency. This may reveal complex interactions of prosodic, linguistic, and usage-based factors leading to the speakers being exposed, during linguistic interactions, to recurrent contexts of high-frequency prosodic sequences of lexical and grammatical chunks. This might, in turn, lead speakers to spend less time processing the contexts and the phonetic output, leading to higher levels of connected speech phenomena given the entrenchment of the variants exemplifying connected speech phenomena (Bybee, 2001). A usage-based account of the impact of the factors studied is, however, not exclusive of other sociolinguistic or phonetic motivations, which may vary depending on the phenomenon or the context.

For instance, one possible reason for the use of laryngeal strategies (glottal stops, creaky voice) when linking /r/ does not apply seems to point to a reorganisation of the hiatus-breaking strategy in some varieties of English. In the case of t-tapping productions, while the use of glottal stops at word boundaries in /t/ contexts (t-tapping study) is marginal, the production of taps in spontaneous speech in RP might be explained in terms of sociolinguistic indexation (i.e., a combination of authority and informality by elite speakers to assert themselves from a position of privilege while maintaining an image of openness and approachability). As for the low percentage of production of yod coalescence, it might be argued that, as yod coalescence is a type of assimilation and a

phonetically motivated process, the low production seems to reflect that it represents a change from below in the sense that it originated in non-standard varieties. As all speakers in the sample displayed yod coalescence forms to some extent, the overall score also indicates that yod coalescence is, although not frequent, at least, recurrent in RP speakers' speech, and that it is not limited to other accentual varieties.

5.2. Limitations and directions for future research

The present study has shed some light on the factors that motivate connected speech phenomena in spontaneous speech in contemporary RP. The analysis has provided valuable insights into how the processes are realised by speakers in discourse, thus contributing to the growing empirical body of data on connected speech processes and the analysis of spoken language in actual situations. Despite the contributions to our knowledge of connected speech processes that the current dissertation has made, the studies conducted have some limitations that should be acknowledged, which can motivate directions for future research.

First and foremost, it is worth mentioning that one of the limitations of the study is the number of speakers who were analysed since the speech of only six speakers was analysed and a sample consisting of only six informants is not representative of the sociolinguistic group that the speakers belong to. However, the sample used was considered to be suitable for the present investigation as the most practical approach for the present study was to delimit a sizeable number of tokens. It is nevertheless acknowledged here that while using a narrow choice of speakers offered many practical advantages at the time of examining the data, a wider range of speakers would undoubtedly provide more robust findings.

In addition to this, another limitation of the present study concerns the speakers' age, as the study only covers one age group (middle-aged speakers), so neither the potentially more conservative language use of older speakers nor the innovative trends of younger speakers could be ascertained. Future research that covers more age groups (i.e., young, middle-aged, and older speakers) would produce a sample with different linguistic data.

Limiting the study to a given age group at a particular point in time (at least a few consecutive years) leads to another limitation of the study that concerns the lack of a

longitudinal perspective, which would have allowed to compare speech community trends. This is because languages are not static entities, but systems with a huge amount of synchronic as well as diachronic variation. A longitudinal analysis involving the connected speech phenomena under investigation over several decades would provide valuable insights into identifying speech community trends diachronically. Future studies, therefore, could look at the use of connected speech phenomena such as the ones studied in this dissertation not only from an apparent-time perspective (i.e., comparing the speech of individuals of different ages at one point in time) but also from a real-time perspective (i.e., comparing data from two or more points in time to observe change directly) (Cuckor-Avila & Guy, 2013).

As to how a longitudinal perspective would provide important insights into connected speech research, when conducting research on language variation and change it is important to reflect on the different methodologies of this type of studies: trend and panel. Trend studies normally involves surveying independent groups of language users who belong to the same community over time while panel studies specifically analyse the language of particular individuals or groups of subjects over time (e.g. Beaman & Buchstaller, 2021; Wagner & Buchstaller, 2018). Then, the ways to research on connected speech are diverse and give the opportunity to reflect on how research is conducted. For example, by analysing different speakers from those in this dissertation sample belonging to a different age group, but in the same time span (e.g., 2010s), this would shed light on distinctive connected speech parameters by using an apparent time and panel study. Given this, future studies on connected speech phenomena should include more potential cases per speaker, more years, more speakers, or a combination of these.

In terms of the types of connected speech processes under study, while two of the connected speech phenomena that have been analysed for the present investigation have been extensively investigated (linking /r/ and t-tapping), yod coalescence remains scarcely investigated. However, the fact that the sample did not include the same number of potential cases for yod coalescence as for linking /r/ and t-tapping can be understood as another limitation of the study since the number of tokens containing a potential yod coalescence is inferior in comparison to linking /r/ and t-tapping. This was due to potential yod coalescence cases being substantially less frequent in the sample compared with potential cases of linking /r/ and t-tapping. Thus, future studies could include more tokens of yod coalescence.

Additionally, future studies should also examine linking /r/, t-tapping, and yod coalescence in other accentual varieties of English, such as non-standard accents. Moreover, connected speech processes somewhat related to the ones studied in this dissertation could also be included. These include, among others, intrusive /r/, t-glottaling, or other types of assimilation processes. Given that the current study has focused on the phenomena across word boundaries, future studies could also look at word-internal morpheme-boundary levels.

Another way in which future studies could improve on the current one is by analysing more variables potentially conditioning variation in the use of connected speech processes. The analysis in this dissertation has focused on several factors (i.e., lexical class, word class combination, prosodic environment, usage-based frequency); nevertheless, a study with a richer set of factors, such as the presence of tone units, vowel type, vowel quality collocation frequency, among others, would reveal a clearer picture of what type of factors prevail.

Despite this particular type of linguistic research being directly associated with the field of phonetics and phonology, it is also true that the findings obtained in it could serve as a valuable resource tool for investigation concerning other linguistic disciplines, such as sociolinguistics or conversation analysis (CA)-based study of talk-in-interaction. Drawing a line between phonetics, phonology, and other linguistic fields of enquiry that have an interest in spoken language is difficult and spoken language is such a multifaceted phenomenon that considering all its components is a substantial challenge in any linguistic study. Nevertheless, future research on connected speech processes could attempt to integrate theoretical and methodological aspects of various research traditions to gain a better understanding of the use of those processes in conversational language.

Finally, future studies could also be conducted by combining different methodologies and data sources. On the one hand, research could use techniques from corpus phonology and more experimental approaches. By way of example, future research involving a parallel study where conversational speech is recorded and later reproduced (i.e. interviews as role plays), by the same or different speakers, might provide interesting results on the differences between conversational and scripted speech. In the exploration of conversational speech from TV, as was the case for the current study, the inclusion of the different TV interviewers' speech and their comparison with the interviewees' speech might reveal some interesting results in future research as well. On

the other hand, data could be collected from different resources other than the traditional elicitation of speech with carrier sentences or spoken/multimodal corpora. These additional sources could include social media platforms, naturalistic settings outside media, or human-machine interactions, among others.

5.3. Implications for applied work

To conclude, some possible implications for applied work based on the findings of the present dissertation are discussed. For instance, future applications of the findings obtained are in connection with areas such as foreign language teaching/learning, speech technology and automatic speech recognition (ASR), or forensic linguistics, among others.

As mentioned above, some of the key implications of the study are related to the teaching and learning of English pronunciation (or second language acquisition, more generally). An updated synchronic description of an accent, such as RP, has practical implications for future research in areas such as teaching and learning English as a foreign language. RP has traditionally been the pronunciation model for English taught in Europe, but the choice of a pronunciation model depends on numerous factors (Mompeán, 2008). For RP, or any other accent, chosen as a target model (e.g., General American), information about connected speech processes can help learners have a more realistic view of what a model actually represents: not an ideal, invariable way of pronouncing specific sounds and lexical items, but a model that encompasses the pronunciation of lexical items in isolation as well as pronunciation in connected speech.

In addition to this, empirically-informed pronunciation models may also help avoid possible stereotypes and/or misconceptions about the L2 target variety or speakers. It is a common misperception, for example, that RP might not be the most suitable English pronunciation model if communicative criteria are taken into consideration (see Mompeán, 2008). According to some of these criteria, the choice of a pronunciation model relies on the degree of intelligibility of speakers when using a given model. Then, the question of how intelligible speakers could be is also linked to the degree of exposure to a particular model that students may have outside the learning environment. As a case in point, a given model of pronunciation such as General American might have a higher

communicative and learning value than RP, given the large number of speakers who use it. This might lead to not chose RP as a pronunciation model for learning/teaching purposes in the classroom setting, since it is commonly, yet wrongly, assumed that RP is spoken by only 3% of the British population (however, the actual percentage is unknown). Therefore, RP would not be a suitable pronunciation model given that it does not follow this criterion.

Considering misconceptions such as the one described above, an updated model of an accent -in this case, RP- can help avoid stereotypes about such pronunciation models. As English as a Foreign Language (EFL) or English as a Second Language (ESL) learners using RP as a pronunciation model would be easily understood by most British citizens, if not all, even if learners find (very) few or no RP speakers when they are in the UK.

Other stereotypes concerning RP are related to some other popular misconceptions, such as the claim that RP is not used in spontaneous speech but exclusively in formal speech contexts. Despite what has been stated, the current thesis proves that RP is not the accent used by public figures only and in extremely formal circumstances (e.g., the Budget, or Financial Statement delivered by the Chancellor of the Exchequer, The State Opening of Parliament, BBC newscasting, etc.). This helps to avoid to dismiss a given accentual variety because of the script type/situational context in which it is wrongly believed the accent is used.

For example, another common myth about RP is that it is an accent that does not change, but as Wells (1982) argued, ‘no accent is a homogeneous invariant monolith - certainly not RP’ (p. 279). Hence, its phonological/phonetic variability must be considered (or reconsidered, at least, in broad terms). Even though RP is one of the most thoroughly described and studied accents within the British Isles, no description includes new (i.e., ongoing, not fully established at the moment) features in its inventory, such as t-tapping, which other empirical studies tend to attribute almost exclusively to other varieties of English (i.e., American English, Canadian English).

Thus, one area where the findings from the current dissertation, as well as from related studies on connected speech, could be used in an applied way is the teaching of oral skills (speaking and listening comprehension) in EFL or ESL, which can be conveniently referred to as L2. Adult native speakers have little trouble understanding

each other and dealing with the outcomes of phonetic speech processes in spoken language, given the large amount of linguistic experience they have gathered over the years and decades of linguistic interaction in their native language. However, it is well known that learners of English often have difficulty with listening comprehension, even at advanced levels. This is a real problem that EFL learners may face given their potential expectation that speech may be a sequence of words with clear-cut boundaries, which is reinforced by the standard graphic representation of written language as a sequence of words with blank spaces in-between.

The frequent ‘disconnect’ between written language and spoken language at word boundaries given the complexities of coarticulation and connected speech would seem to require more attention from L2 pronunciation practitioners, many of whom seem well aware of this. According to Dalton and Seidlhofer (1994), for example, “[w]e do not articulate sounds in isolation but connect them in strings” (p. 22). Mastering this would presumably help learners not only understand their interlocutors, but also make their own speech sound more natural, fluent, and perhaps even more native (or native-like). Since the use of connected speech processes “is completely acceptable, natural and an essential part of speech” (Alameen & Levis, 2015, p. 3), teaching it should be an important part and parcel of the teaching of oral skills and pronunciation of the L2.

In addition, future research could also be oriented into the development of areas such as speech technology and automatic speech recognition (ASR), or forensic linguistics. For instance, Foulkes and Docherty (1999) already identified speech technology as maybe the most noticeable area in which the comprehension of variation in speech has been pursued, where the involvement of linguistic research can be of paramount importance.

The research findings can also have possible applications in areas such as ASR systems (i.e., automatic speech recognition) to explain sociolinguistic variation. Even though much is known about how speech is produced, the knowledge of speech production can be of paramount importance in current mainstream approaches to use the findings from the present study to improve ASR systems. By considering how different social groups (e.g., RP vs. working-class speakers) may pronounce words or phrases differently, ASR systems can become more suitable and effective in both understanding and transcribing speech.

By accounting for sociolinguistic variation, for example, this dissertation provides knowledge of the different parameters that motivate speech production (as far as connected speech is concerned) which cannot be straightforwardly obtained from the acoustic signal or a phonetic transcription alone (i.e., linguistic or usage-based factors). Therefore, the modelling of ASR systems can benefit from this knowledge, so as to speech is not only modelled as a “surface” type of phenomenon. Thus, incorporating this type of speech production knowledge into ASR systems will enable improved recognition of spontaneous speech, as explicit modelling at the production level will specify where, when, and how connected speech processes occur to construct production-based representations (King, Frankel, Livescu, McDermott, Richmond & Wester, 2007).

Some other implications for applied work using the research findings can be used in the field of forensic linguistics to help identify speakers, examine voice recordings, and provide linguistic evidence in legal cases. The fact that linguistic aspects are often used in legal cases draws attention to how important linguistic evidence is. As Coulthard, Johnson & Wright (2016) claimed, there is something unique and valuable in voice in such a way that a tape recording of a suspected murderer more than twenty-five years ago could be effectively matched with a recently arrested suspect’s vocal output “in spite of the passage of time and its effect on the voice” (p. 2). Hence, the findings obtained in the present study can be useful in this particular area as well as in the others mentioned earlier, to improve the knowledge of casual speech in speaker identification. Linguistic evidence, spoken one, as in the case mentioned earlier, can sometimes be decisive for phoneticians or dialectologists who are employed to analyse recordings in legal cases. Findings from empirical work on connected speech, such as the ones obtained in the present doctoral thesis, could be essential for the update of accentual varieties and how particular features describe certain patterns of sociolinguistic variation. Thus, findings on connected speech phenomena enable to have more accurate and updated descriptions of linguistic profiles for forensic linguists. As it is described in Coulthard et al. (2016, p. 4), “corpora and real data, though are often limited in size and availability, are central to our endeavour”.

References

- Alameen, G., & Levis, J. M. (2015). Connected Speech. In M. Reed & J. M. Levis (eds). *The Handbook of English Pronunciation* (pp. 159-174). Wiley: Blackwell.
- Alderton, R. (2022). T-tapping in Standard Southern British English: An ‘elite’ Sociolinguistic Variant? *Journal of Sociolinguistics* 26, 287-298. <https://doi.org/10.1111/josl.12541>
- Allerton, D. J. (2000). Articulatory Inertia Vs ‘Systemzwang’: Changes in Liaison Phenomena in Recent British English. *English Studies* 81(6), 574-581. <https://doi.org/10.1076/enst.81.6.574.9183>
- Almela, M. & Marqués, T. (2016). *Introduction to English Linguistics*. Diego Marin.
- Altendorf, U. (2003). *Estuary English: Levelling at the interface of RP and south-eastern British English*. Tübingen: Gunter Narr Verlag.
- Amos, J. (2007). *Wadda boo’iful place: An Analysis of the variable (ju) and (t) in Mersea Island English* [Unpublished Master’s Dissertation]. University of Essex.
- Andersen, H. (1986). *Sandhi Phenomena in the Languages of Europe*. Berlin, Boston: de Gruyter Mouton.
- Anderson, J. R. (1993). *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum.
- Anderson, S. R. (2001). Phonology. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences* (pp. 11386-11392). Pergamon. <https://doi.org/10.1016/B0-08-043076-7/02982-X>
- Anderson-Hsieh, J., Rinev, T., & Koehler, K. (1994). Connected speech modifications in the English of Japanese ESL learners, *IDEAL*, 7, 31-52.
- Anttila, A. (2004). Variation and phonological theory. In J. K. Chambers, Peter Trudgill, Natalie Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 206–243). Oxford: Blackwell Publishing Ltd.
- Anttila, A., & Cho, Y. Y. (1998). Variation and Change in Optimality Theory. *Lingua* 104(1–2), 31–56.
- Ash, S., & Grossman, M. (2015). Why study connected speech production? In R. Willems (Ed.), *Cognitive Neuroscience of Natural Language Use* (pp. 29-58). Cambridge: Cambridge University Press.
- Ashby, M. G., & Maidment, J. (2005). *Introducing Phonetic Science*. Cambridge: Cambridge University Press.

- Ashby, M. G., & Przedlacka, J. (2010). The stops that aren't. *Journal of the English Phonetic Society of Japan*, 1, 14-15.
- Badia Barrera, B. (2015). A Sociolinguistic Study of T-glottalling in Young RP: Accent, Class and Education. [Unpublished Doctoral Dissertation]. University of Essex.
- Balogné Bérces, K. (2004). Connected speech phenomena in Strict CV phonology. *The Even Yearbook* 6: 1–10.
- Balogné Bérces, K. (2008). *Strict CV Phonology and the English Cross-word Puzzle: Empty Categories in Connected Speech*. Saarbrücken: VDM Verlag Dr. Müller.
- Balogné Bérces, K. & Honeybone, P. (2020). Representation-based models in the current landscape of phonological theory. *Acta Linguistica Academica* 67(1), 3-27. <https://doi.org/10.1556/2062.2020.00002>
- Bauer, L. (1984). Linking /r/ in RP: some facts. *Journal of the International Phonetic Alphabet* 14, 74-79. <https://doi.org/10.1017/S0025100300002802>
- Bauer, L. (1994). *Watching English Change: An Introduction to the Study of Linguistic Change in the Twentieth Century*. London: Longman.
- BBC. (n.d.). *The Andrew Marr Show*. Retrieved from <http://www.bbc.co.uk/programmes/b0080bbs> (Date of access 15 August 2017).
- BBC News. (n.d.). *UK Politics. Cooper: police poll 'troubling'* <http://www.bbc.com/news/uk-politics-20196772> (Date of access 8 September 2017)
- BBC News. (n.d.). *UK Politics. Shadow home secretary: police elections are 'shambles'* <http://www.bbc.com/news/uk-politics-20197498> (Date of access 8 September 2017)
- Beal, J. C., Sen, R., Yáñez-Bouza, N., & Wallis, C. (2020). En[dj]uring [ʃ]unes or ma[tj]ure [ɔ̃]ukes? Yod-coalescence and Yod-dropping in the Eighteenth-Century English Phonology Database. *English Language and Linguistics*, 24(3), pp. 493-526.
- Beaman, K.V., & Buchstaller, I. (Eds.). (2021). *Language Variation and Language Change Across the Lifespan: Theoretical and Empirical Perspectives from Panel Studies* (1st ed.). Routledge. <https://doi.org/10.4324/9780429030314>
- Bell, A. G. (1977). The language of radio news in Auckland: a sociolinguistic study of style, audience and subediting variation. Ph.D. thesis. University of Auckland.
- Bell, A. G. (1984). Language style as audience design. *Language in Society* 13, pp. 145-204.
- Bell, A. (1991). *The Language of News Media*. Oxford: Blackwell.
- Bell, A. & Hooper, J. (1978). *Syllables and Segments*. Amsterdam: North Holland.

- Beňuš, Š. (2021). *Investigating spoken English: A practical guide to phonetics and phonology using Praat*. Palgrave Macmillan.
- Bermúdez-Otero, R. (2002). Review of Noel Burton-Roberts, Philip Carr, and Gerard Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*. Oxford: Oxford University Press, 2000. Pp. x+352. *Journal of Linguistics* 38(2): 403–410.
- Blevins, J. & Garret, A. (2004). The Evolution of Metathesis. In Bruce H., Robert K., & Donca S. *Phonetically Based Phonology*, pp. 117-156. Cambridge: Cambridge University Press.
- Bloomfield, L. (1933). *Language*. New York: Henry Holt.
- Boberg, C. (2015). North American English. In, M. Reed; J. M. Levis. *The Handbook of English Pronunciation* (pp- 229-250). Wiley: Blackwell.
- Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32(1): 45–86.
- Borowsky, T. (1991). *Topics in the lexical phonology of English*. New York: Garland Press.
- Boyland, J. T. (1996). *Morphosyntactic Change in Progress: A Psycholinguistic Approach*. Dissertation: University of California at Berkeley.
- Britain, D. (2017). Beyond the ‘gentry aesthetic’: elites, Received Pronunciation and the dialectological gaze in England. *Social Semiotics* 27 (3): 288-298.
- Britain, D. & Fox, S. (2009). The regularisation of the hiatus resolution system in British English: A contact-induced ‘vernacular universal’? In M. Filppula, J. Klemola & H. Paulasto (eds.), *Vernacular universals and language contacts: Evidence from varieties of English and beyond*, 177–205. London: Routledge.
- Broadbent, J. M. (1991). Linking and intrusive /r/ in English. *UCL Working Papers in Linguistics* 3, 281-302.
- Brown, A. (1988a). Linking, Intrusive and Rhotic /r/ in Pronunciation Models. *Journal of the International Phonetic Association* 18, 144-151.
- Brown, A. (1988b). The Staccato Effect in the Pronunciation of English in Malaysia and Singapore. In Joseph F. (ed.), *New Englishes: The Case of Singapore*. pp 115-147. Singapore: Singapore University Press.
- Brown, A. (2014). *Pronunciation and Phonetics: A Practical Guide for English Language Teachers*. New York: Routledge.
- Burridge, K. & Bergs, A. (2016). *Understanding Language Change*. London: Routledge.

- Butler, C. (1985). *Statistics in linguistics*. Oxford: Blackwell.
- Bybee, J. L. (1985). *Morphology: A Study of the Relation Between Meaning and Form*. Philadelphia: Benjamins.
- Bybee, J. L. (1999). Usage-based Phonology. In M. Darnell, E. Moravcsik, F. Newmeyer, M. Noonan, & K. Wheatly (Eds). *Functionalism and Formalism in Linguistics VI* (pp. 211-242). Amsterdam: John Benjamins.
- Bybee, J. L. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Bybee, J. L. (2002). Word Frequency and Context of Use in the Lexical Diffusion of Phonetically Conditioned Sound Change. *Language Variation and Change* 14, 261-290.
- Bybee, J. L. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4), 711-733.
- Bybee, J. L. (2013). Usage-based Theory and Exemplar Representation of Constructions. In T. Hoffman & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 49-69). Oxford: Oxford University Press.
- Bye, P. (2011). Dissimilation. In M van Oostendorp, C. Ewen, E. Hume, & K. Rice (eds.) *The Blackwell Companion to Phonology, 5 Volume Set* (Vol. 3: Segmental Processes), pp. 1408-1433. Oxford: Wiley-Blackwell.
- Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication* 15, 39-54.
- Campbell, L. (2013). *Historical linguistics*. Edinburgh: Edinburgh University Press.
- Cantos-Gómez, P. (2013). *Statistical Methods in Language and Linguistic Research*. Sheffield/Bristol: Equinox.
- Carterette, E., & Hubbard Jones, M. (1974). *Informal Speech: Alphabetic and Phonemic Text with Statistical Analyses and Tables*. Berkeley: University of California Press.
- Chambers, J. K. (1998). Social Embedding of Changes in Progress. *Journal of English Linguistics* 26 (1), 5-36.
- Chambers, J.K. (2002). *Sociolinguistic Theory*. 2nd edn. Oxford: Blackwell.
- Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects* (Vol. 92). Cambridge: Cambridge University Press.
- Chomsky, N. (1964). *Current Issues in Linguistic Theory*. The Hague: Mouton.
- Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.

- Clark, J., Yallop, C., & Fletcher, J. (2007). *An Introduction to Phonetics and Phonology*, 3rd ed. Blackwell Pub.
- Clarke, S. (1993). The Americanization of Canadian Pronunciation: A Survey of Palatal Glide Usage. In S., Clarke. (Eds.), *Focus on Canada*, (pp. 85-108). Amsterdam: Benjamins.
- Carr, P. (1991). Lexical Properties of Post-Lexical Rules: Postlexical Derived Environment and the Elsewhere Condition. *Lingua* 85, 255-268.
- Clements, G. N. & Keyser, S. J. (1983). *CV Phonology: A Generative Theory of the syllable*. MIT Press. Cambridge, Mass.
- Coates, J. 2004. *Women, men and language: A sociolinguistic account of gender differences in language*, 3rd edn. London & New York: Longman.
- Collins, B., & Mees, I. M. (2003). *Practical Phonetics and Phonology. A resource book for students*. London: Routledge.
- Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychomic Bulletin & Review* 11 (6), 1084-1089.
- Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and Speech*. Cambridge, MA: Harvard University Press.
- Côté, M. H. (2011). *French Liaison*. In the Blackwell Companion to Phonology (eds. M. Oostendorp, C. J. Ewen, E. Hume and K. Rice).
- Côté, M. H. (2013). Understand cohesion in French liaison. *Language Sciences*, 39, 156-166.
- Cox, F., Palethorpe, S., Buckley, L., & Bentink, S. (2014). Hiatus resolution and linking 'r' in Australian English. *Journal of the International Phonetic Association* 44, 155-178. DOI: 10.1017/S0025100314000036
- Crowley, T., & Bowern, C. (2010). *An Introduction to Historical Linguistics*. Oxford: Oxford University Press.
- Cruttenden, A. (2008). *Gimson's Pronunciation of English*. 7th edn. London: Arnold Education.
- Cruttenden, A. (2014). *Gimson's Pronunciation of English*. 8th edn. Routledge.
- Crystal, D. (1984). *Who Cares about English Usage?* Harmondsworth: Penguin Books.
- Crystal, D. (2003). *A Dictionary of Linguistics and Phonetics*. 5th edn. London: Blackwell.

- Cuckor-Avila, Patricia, & Bailey, Guy. (2013). Real and apparent time. In J. K. Chambers & N. Schilling (eds.), *The handbook of language variation and change*. 2nd ed. Malden, MA: Wiley-Blackwell. 239–262.
- Cutillas, J. A. (2003). *Teoría Lingüística de la Optimidad: fonología, morfología y aprendizaje*. Murcia: Universidad de Murcia, servicio de publicaciones.
- Dalton, C., & Seidlhofer, B. (1994). *Pronunciation*. Oxford, UK: Oxford University.
- De Jong, K. J. (1998). Stress-related Variation in the Articulation of Coda Alveolar Stops: Flapping Revisited. *Journal of Phonetics* 26 (3), 283-310.
- De Jong, K. J. (2011). Flapping in American English. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (p. 2711-2729). Malden, Mass.: Wiley-Blackwell.
- De Carvalho, A., He, A. X., Lidz, J., & Christophe, A. (2019). Prosody and Function Words Cue the Acquisition of Word Meanings in 18-Month-Old Infants. *Psychological Science*, 30(3), 319–332.
- Deterding, D. (2006). The North Wind Versus a Wolf: Short Texts for the Description and Measurement of English Pronunciation. *Journal of the International Phonetic Association* 36, 187-196. <https://doi.org/10.1017/S0025100306002544>
- Denes, P. B., & Pinson, E. N. (1993). *The Speech Chain: The Physics and Biology of Spoken Language* (2nd ed.). W. H. Freeman and Company.
- Denham, K. & Lobeck, A. (2012). *Linguistics for Everyone: An Introduction*. Boston: Cengage Learning.
- Donegan, P. (1993). On the phonetic basis of phonological change. In C. Jones (Ed.), *Historical Linguistics: Problems and Perspectives*, (pp. 98-130). London: Longman.
- Donegan, P. (2002). Phonological processes and phonetic rules. *Future challenges for natural linguistics*, 57-81.
- Donegan, P. & Stampe, D. (1979). The study of Natural Phonology. In D. A. Dinnsen (Eds.), *Current Approaches to Phonological Theory* (Conference on the Differentiation of Current Phonological Theories, Bloomington, Sept. 30-Oct. 1, 1977), (pp. 126-173). Bloomington: Indiana University Press.
- Dressler, W., & Wodak, R. (1982). Sociophonological Methods in the Study of Sociolinguistic Variation in Viennese German. *Language in Society* 11(3), 339-370.

- Eddington, D. (2007). Flaps and other variants of /t/ in American English: Allophonic distribution without constraints, rules, or abstractions. *Cognitive Linguistics* 18-1, 23-46.
- Eddington, D., & Elzinga, K. (2008). The Phonetic Context of American English Flapping: Quantitative Evidence. *Language and Speech* 51(3), 245-266.
- Eddington, D., Treiman, R., & Elzinga, K. (2013). Syllabification of American English: Evidence from a Large-Scale Experiment. Part 1. *Journal of Quantitative Linguistics* 20(1), 45-67.
- Egido, C. & Cooper, W. E. (1980). Blocking of alveolar flapping in speech production: The role of syntactic boundaries and deletion sites. *Journal of Phonetics* 8, 175-184.
- Encyclopedia Britannica. (n.d.). Theresa May [https://www.britannica.com/ Theresa May | Biography, Facts, & Policies | Britannica](https://www.britannica.com/Theresa-May/Biography,Facts,&Policies|Britannica) (Date of access 20 May 2018)
- Fabricius, A. H. (2017). Twentieth-century RP: Prevocalic /r/. In Raymond Hickey (ed.), *Listening to the past*, 39–65. Cambridge: Cambridge University Press.
- Fabricius, A. H. (2000). *T-glottalling Between Stigma and Prestige: A Sociolinguistic Study of Modern RP*. PhD Thesis. Copenhagen Business School.
- Farnetani, E., & Recasens, D. (2010). Coarticulation and Connected Speech Processes. In Hardcastle, W. J., Laver, J. & Gibbon, F. D. (Eds.), *The Handbook of Phonetic Sciences* (pp. 316-352). 2nd edn. Chichester, U.K: Wiley-Blackwell.
- Fasold, R. W. & Connor-Linton, J. (2014). *An Introduction to Language and Linguistics*. Cambridge: Cambridge University Press.
- Fidelholtz, J. L. (1975). Word Frequency and Vowel Reduction in English. *CLS* 11, 200-213.
- Fosler-Lussier, E., & Morgan, N. (1998). Effects of Speaking Rate and Word Frequency on Conversational Pronunciations. *ESCA Tutorial and Research Workshop on Modeling Pronunciation for Variation for Automatic Speech Recognition* (pp. 1-6). Kerkrade, Netherlands.
- Foulkes, P. (1997a). Rule Inversion in a British English Dialect: A Sociolinguistic Investigation of [r]-sandhi in Newcastle upon Tyne. *University of Pennsylvania Working Papers in Linguistics* 4(1), 259-270.
- Foulkes, P. (1997b). English [r]-sandhi: A sociolinguistic perspective. *Histoire Épistémologie Langage* 19, 73-96.
- Foulkes, P. & Docherty, G. (1999). *Urban Voices: Accent Studies in the British Isles*. London: Arnold.

- Fox, A. (1978). To 'r' is human? Intrusive remarks on a recent controversy? *Journal of the International Phonetic Association* 8(1-2), 72-74.
- Fry, D. B. (1979). *The Physics of Speech*. UK: Cambridge University Press.
- Fukaya, T., & Byrd, D. (2005). An Articulatory Examination of Word-Final Flapping at Phrase Edges and Interiors. *Journal of the International Phonetic Association*, 35(1), 45-58.
<https://doi.org/10.1017/S0025100305001891>
- Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallet, D. S. & Dahlgren, N. L. (1993). *DARPA TIMIT: Acoustic-Phonetic Continuous Speech Corpus*. Philadelphia: Linguistic Data Consortium.
- Gick, B. (1999). A gesture-based account of intrusive consonants in English. *Phonology* 16, 29-54.
- Gick, B. (2002). The American Intrusive l. *American Speech*, 77(2), 167-183.
- Giegerich, H. J. (1992). *English Phonology: An introduction*. Cambridge: Cambridge University Press.
- Giegerich, H. J. (1999). *Lexical Strata in English: Morphological Causes, Phonological Effects*. *Cambridge Studies in Linguistics*. Cambridge: Cambridge University Press.
- Gimson, A. C. (1980). *An Introduction to the Pronunciation of English*, 3rd edn. London: Edward Arnold. (Original work published 1962).
- Glain, O. (2012). The yod /j/: palatise it or drop it! How Traditional Yod Forms Are Disappearing from Contemporary English. *Cercles* 22, 4-24.
- Glain, O. (2014). Introducing Contemporary Palatalisation. *York Papers in Linguistics: PARLAY Proceedings Series*, 1(1), 16-29.
- Godfrey, J., McDaniel, J., & Holliman, J. (1992). SWITCHBOARD: A Telephone Corpus for Research and Developments. In *ICASSP Proceedings* (pp. 517-520). San Francisco, United States.
- Goldsmith, J. A. (1976). *Autosegmental Phonology*. Bloomington: Indiana University Linguistics Club.
- Gordon, E., Maclagan, M., & Hay, J. (2007). The ONZE Corpus. In J. C. Beal, K. P. Corrigan, & H. L. Moisl (Eds.) *Creating and Digitizing Language Corpora, vol. : Diachronic Databases* (pp. 82-104). Basingtoke: Palgrave Macmillan.
- Grammont, M. (1936). *Traité de phonétique*. Paris: Delagrave.

- Greenberg, S. (1999). Speaking in shorthand- A Syllable-centric Perspective for Understanding Pronunciation Variation. *Speech Commun* 29, 159-176.
- Greenberg, S., & Fosler-Lussier, E. (2000). The Uninvited Guest: Information's Role in Guiding the Production of Spontaneous Speech. *Proceedings of the CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling* (pp. 129-132). Kloster Seeon, Germany.
- Greenberg, S., Hollenback, J., & Ellis, D. (1996). Insights into spoken language gleaned from phonetic transcription of the Switchboard corpus. *Proceedings of the International Conference on Spoken Language Processing* 24-27. Philadelphia.
- Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society* 35, 151-166.
- Gussenhoven, C. (1986). English Plosive Allophone and Ambisyllabicity. *Gramma* 10, 119-141.
- Gussenhoven, C., & Jacobs, H. (2011). *Understanding Phonology: Understanding Language* (3rd ed.). London: Hodder Education.
- Hall, N. (2007). R-Dissimilation in English. Ms., California State University, Long Beach.
- Halle, M. & Idsardi, W. J. (1997). r, Hypercorrection, and the elsewhere condition. In: I. Roca (ed.), *Derivations and Constraints in Phonology*, 331-348. Oxford: Clarendon Press.
- Halliday, M. A. K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.
- Hammond, M. (1997). Vowel Quality and Syllabification in English. *Language* 73(1), 1-17.
- Hannahs, S. J., & Bosch, A. (2018). *The Routledge Handbook of Phonological Theory*. London & New York: Routledge.
- Hannisdal, B. R. (2006). *Variability and change in Received Pronunciation. A study of six phonological variables in the speech of television newsreaders*. Doctoral Dissertation. Department of English, University of Bergen.
- Harriet Harman MP (n.d.). About Harriet – Harriet Harman MP (Retrieved from [Harman, Harriet \(1950–\) | Encyclopedia.com](#)) (Date of access 12 June 2018..
- Harris, J. (1994). *English Sound Structure*. Oxford: Blackwell.
- Harris, Z. S. (1951). *Methods in Structural Linguistics*. Chicago: University of Chicago Press.

- Hartmann, D., & Sabine, Z. (2009). Rhoticity in Black South African English: A Sociolinguistic Study. *African Linguistics and Applied Languages Studies* 27(2), 135-148.
- Haspelmath, M., & Tadmor, U. (2009). *World Loanword Database*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Last accessed 11 Sept 2022 <https://wold.clld.org/>
- Haugen, E. (1938). Notes on ‘voiced t’ in American English. *Dialect Notes*, 6, 627-634.
- Hay, J., & Foulkes, P. (2016). The Evolution of Medial /t/ Over Real and Remembered Time. *Language* 92, 298-330. <https://doi.org/10.1353/lan.2016.0036>
- Hay, J., Jannedy, S., & Mendoza-Denton, N. (1999). Oprah and /ay/: Lexical Frequency, Referee Design and Style. *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, 1389-92.
- Hay, J., & MacLagan, M. (2010). Social and phonetic conditioners on the frequency and degree of ‘intrusive /r/’ in New Zealand English. In Preston, D., Niedzielski, N. (eds). *A reader in sociophonetics* 249, 41-69.
- Hay, J., & MacLagan, M. (2012). /r/-sandhi in early 20th century New Zealand English. *Linguistics* 50, 745-763.
- Hay, J., & Sudbury, A. (2005). How rhoticity became /r/-sandhi. *Language* 81, 799-823.
- Hay, J., & Warren, P. (2002). Experiments on /r/-intrusion. *Wellington Working Papers in Linguistics* 14, 47-58.
- Hayes, B. (1981). *A metrical theory of stress rules*. PhD Thesis, MIT, Distributed by Indiana University Linguistics Club.
- Heselwood, B. (2006). Final schwa and r-sandhi in RP English. *Leeds Working Papers in Linguistics and Phonetics* 11, 78-95.
- Hock, H. H. (1991). *Principles of historical linguistics*. Berlin, Boston: de Gruyter Mouton.
- Holm, J. (1988). *Pidgins and Creoles. Vol. 1: Theory and Structure*. Cambridge: Cambridge University Press.
- Holmes, J. (1994). New Zealand Flappers: An Analysis of T-Voicing in New Zealand English. *English World-Wide* 15(2), 195-224.
- Holmes, J. (1995). Glottal stops in New Zealand English: An Analysis of Variants of Word-final /t/. *Linguistics* 33, 433-63.
- Hooper, J. (1976). *An Introduction to Natural Generative Phonology*. New York: Academic Press.

- Hovarth, B. (2008). Australian English: phonology. *Varieties of English* 3, 89-109.
- Huang, Y., Athanasopoulou, A., & Vogel, I. (2018). The effect of focus on creaky phonation in Mandarin Chinese tones. *University of Pennsylvania Working Papers in Linguistics*, 24(1), pp. 96-104.
- Hudson, R. A. (1996). *Sociolinguistics*. 2nd edn. Cambridge: Cambridge University Press.
- Ingram, J. C. L. (1989). Connected Speech Processes in Australian English. *Australian Journal of Linguistics*, 9(1), 21-49.
- Janda, R. D. & Joseph, B. D. (2003). Reconsidering the Canons of Sound-Change. In B. Blake & K. Burridge (Eds.) *Historical Linguistics 2001*. Selected Papers from the 15th International Conference on Historical Linguistics, Melbourne 13-17 August 2001, (pp.205-219). Amsterdam: John Benjamins Publishing Co.
- Jespersen, O. (1928). *A Modern English Grammar on Historical Principles*. Vol. 1, *Sounds and Spellings*. London: Allen and Unwin.
- Jensen, J. T. (1993). *English Phonology*. Amsterdam: John Benjamins.
- Johansson, S. (1973). Linking and intrusive /r/ in English: A case for a more concrete phonology. *Studia Linguistica* 27, 53-68.
- Johnson, L. (1973). Dissimilation as a natural process in phonology. *Stanford Occasional Papers in Linguistics* 3, 45-56.
- Jones, D. (1956). *The Pronunciation of English*, 4th edn. Cambridge: Cambridge University Press. (Original work published 1909).
- Jones, D. (2011). *English Pronouncing Dictionary*. Cambridge: Cambridge University Press. (Original word published 1917).
- Jones, D. (1972). *An Outline of English Phonetics*. 9th edn. Cambridge: Cambridge University Press. (Original word published 1918).
- Joseph, B. D. (1999). Utterance Finality –Framing the Issues. In *Proceedings of LP'98*, vol. 1, O. Fujimura, B. D. Joseph, B. Palek (Eds), 3-14. Prague: The Karolinum Press.
- Kahn, D. (1976). *Syllable-based generalizations in English phonology*. Doctoral dissertation, MIT, (published by Garland Publishing Company: New York. 1980).
- Kamińska, T. E. (1995). *Problems in Scottish English Phonology*. Tübingen: Niemeyer.

- Kaze, J. W. (1989). *Metaphony in Italian and Spanish dialects revisited*. Doctoral dissertation. University of Illinois at Urbana-Champaign.
- Kenstowicz, M (1994). *Phonology in Generative Grammar*. Cambridge, Massachusetts: Blackwell.
- Kerswill, P. (1985). A Sociophonetic Study of Connected Speech Processes in Cambridge English: An Outline and Some Results. *Cambridge Papers in Phonetics and Experimental Linguistics* 4, 25-49.
- Kijak, A. (2009). How stressful is L2 stress? A cross-linguistic study of L2 perception and production of metrical systems. Doctoral Dissertation, Universiteit Utrecht.
- King, S., Frankel, J., Livescu, K., McDermott, E., Richmond, K., & Wester, M. (2007). Speech production knowledge in automatic speech recognition. *The Journal of the Acoustical Society of America*, 121(2), 723-742.
- Kiparsky, P. (1982). Lexical Morphonology and Phonology. In I. S. Yang (Ed.), *Linguistics in the Morning Calm* (pp. 3-92). Seoul: Hanshin
- Knowles, G. (1987). *Patterns of spoken English: An introduction to English Phonetics*. London & New York: Longman.
- Kolachina, S., Sharma, D. M., Gadde, P., Vijay, M., Sangal, R., & Bharati, A. (2011). External Sandhi and its Relevance to Syntactic Treebanking. *Polibits* 43, 67-74.
- Krämer, M. (2005). English schwa insertion before liquids and phonological opacity. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 41(1), 267-281. Chicago Linguistic Society.
- Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2(2), 205-254.
- Labov, W. (1994). *Principles of Linguistic Change. Vol. 1: Internal Factors*. Oxford: Blackwell.
- Labov, W. (2006). *The Social Stratification of English in New York City* (2nd ed). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511618208
- Ladefoged, P. (1965). The Nature of General Phonetic Theories. *Georgetown University Monograph Series on Language and Linguistics* 18, 27-42.
- Ladefoged, P. (1971). *Preliminaries to Linguistics Phonetics*. Chicago: University of Chicago Press.

- Ladefoged, P. (2001) *Vowels and Consonants: An introduction to the sounds of languages*. Oxford: Blackwell Publishers.
- Ladefoged, P., & Johnson, K. (2015). *A Course in Phonetics* (7th Ed.). Cengage Learning.
- Ladefoged, P., & Maddieson, I. (1996). *The Sounds of World's Languages*. Oxford: Blackwell.
- Laferriere, M. & Zue, V. (1977). Flapping Rule in American English: An Acoustic Study. *Journal of the Acoustical Society of America* 61, S31.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar*, vol. 1: *Theoretical Prerequisites*. Stanford: Stanford University Press.
- Langacker, R. W. (1999). A Dynamic Usage-Based Model. In R. W. Langacker, *Grammar and Conceptualization*, 91-145. Berlin & New York: Mouton de Gruyter.
- Lass, R. (1984). *Phonology*. Cambridge: Cambridge University Press.
- Lass, R. (1987). *The Shape of English*. London: Dent & Sons.
- Lavoie, L. M. (2001). *Consonant strength: Phonological patterns and phonetic manifestations*. Routledge.
- Levshina, N. (2015). *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Netherlands: John Benjamins Publishing Company.
- Levon, Erez. (2010). *Organising and Processing Your Data: The Nuts and Bolts of Quantitative Analyses*. In Lia Litosseliti (Ed.), 68–92.
- Lewis, J. W. (2014). Linking /r/ in the General British Pronunciation of English. *Journal of the International Phonetic Association* 5(1), 37-42. (Original word published 1975).
- Lewis, J. W. (1977). The r-link Business - A Reply. *Journal of the International Phonetic Association* 7(1), 28-31.
- Lindsey, G. A. (2019). *English after RP: Standard British Pronunciation Today*. Springer International Publishing.
- Lutz, A. (1991). *Phonotakitsch Gesteuerte Konsonantenveränderungen in der Geschichte des Englischen*. Tübingen: Niemeyer.
- Malmkjaer, K. (2002). *The Linguistics Encyclopaedia*. London: Routledge
- Manchaiah, V. (2007). *The Effect of Stimulation Rate on Speech Perception in 4th Generation Cochlear Implant Users*. Unpublished MSc Dissertation, University of Southampton.

- Masutti, V. (2014, February, 13-15). *The Liaison in French IP and DP: A Syntactic Analysis* [Talk]. Incontro di Grammatica Generativa IGG40, Italy, Trento.
- Mathisen, A. G. (1999). Sandwell, West Midlands: Ambiguous Perspectives on Gender Patterns and Models of Change. In P. Foulkes & G. Docherty (Eds.), *Urban Voices: Accent Studies in the British Isles*, (pp. 107-123). London: Arnold.
- McCarthy, J. J. (1991). Synchronic Rule Inversion. *Berkeley Linguistics Society* 17, 192-207.
- McCarthy, J. J. (1993). A Casa of Surface Constraint Violation. *Canadian Journal of Linguistics* 38, 169-95.
- McCarthy, J. J. (2001). Nonlinear phonology. In N. J. Smelser & P. B. Baltes (Eds), *International Encyclopedia of the Social and Behavioral Sciences*. Oxford: Permanon, pp. 11392-11395.
- McCarthy, J.J. & Prince, A. (1993). Generalized Alignment. In G. Booji & J. van Marle (Eds.), *Yearbook of Morphology* 6 (pp. 79-153). Dordrecht: Kluwer.
- McMahon, A. (1994). *Understanding Language Change*. Cambridge: Cambridge University Press.
- McMahon, A. (2000). *Lexical phonology and the History of English*. Cambridge: Cambridge University Press.
- McMahon, A. (2002). *An Introduction to English Phonology*. Edinburgh: Edinburgh University Press.
- Mendoza-Denton, N., Hay, J., & Jannedy, S. (2003). Probabilistic Sociolinguistics: Beyond Variable Rules. In R. Bod, J. Hay, & S. Jannedy (eds.) *Probabilistic Linguistic*, pp. 97-138. Cambridge, MA: MIT Press.
- Milroy, L., & Gordon, M. (2003). *Sociolinguistics: Method and Interpretation*. Oxford: Blackwell.
- Minkova, D. (2014). *A Historical Phonology of English*. Edinburgh: Edinburgh University Press.
- Mohanan, K. P. (1982). *Lexical Phonology*. Doctoral Dissertation, Massachusetts Institute of Technology.
- Mohanan, K. P. (1986). *The theory of Lexical phonology*. Dordrecht: Reidel Publishing Co.
- Monroy-Casas, R. (2001). Profiling the phonological processes shaping the fossilised IL of adult Spanish learners of English. Some theoretical implications. *International Journal of English Studies*, 1, 157–217.

- Mompeán, J. A. (2022). /r/ Sandhi in the Speech of Queen Elizabeth II. *Journal of the International Phonetic Association* 52(2), 246-277. doi:10.1017/S0025100320000213
- Mompeán, J. A. (2008). Consumers' preferences and the choice of English pronunciation models. In R. Monroy & A. Sanchez (Eds.), *25 years of applied linguistics in Spain: Milestones and challenges* (pp. 959–964). Murcia: Editum.
- Mompeán, J. A., & Gómez, A. (2011). Hiatus-resolution Strategies in Non- Rhotic English: The Case of /r/-liaison. In W. Lee & E. Zee (Ed), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1414- 1417). Hong Kong, China.
- Mompeán, J. A., & Mompeán-Guillamón, P. (2009). /r/-liaison in English: An empirical study. *Cognitive Linguistics*, 20(4), 733–776.
- Mompeán, P., & Mompeán, J. A. (2007). Phonetic factors in /r/-liaison usage: A First Report. J. Trouvain & W. J. Barry (Dir.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1397-1400). Sarrebruck, Germany: U. Saarlandes.
- Nagy, N., & Irwin, P. (2010). Boston (r): Neighbo(r)s nea(r) and far(r). *Language Variation and Change* 22(2), 241-278.
- Nathan, G. S. (2008). *Phonology: A Cognitive Grammar Introduction* (Vol. 3). Amsterdam & Philadelphia: John Benjamins.
- Navarro, S. (2013). *Rhoticité et 'r' de sandhi en anglais: Du Lancashire Á Boston*. Ph.D. dissertation, Université de Toulouse Jean Jaurès.
- Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.
- Nespor, M., & Vogel, I. (2007). *Prosodic Phonology: With a New Foreword*. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110977790>
- O'Connor, J. D. (1967). *Better English Pronunciation*. Cambridge: Cambridge University Press.
- Ohala, J. J. (1981). The listener as a source of sound change. In *Papers from the Parasession on Language and Behaviour*. Chicago: Chicago Linguistics Society, 251-274.
- Oswald, V. A. (1943). Voiced t- a misnomer. *American Speech* 18, 18-25.
- Parker, F., & Walsh, T. (1982). Blocking Alveolar Flapping: A Linguistic Analysis. *Journal of Phonetics* 10(3), 301-314.
- Patterson, D. & Connine, C. M. (2001). Variant Frequency in Flap Production. A corpus analysis of variant frequency in American English flap production. *Phonetica*, 58(4), 254-275. <https://doi.org/10.1159/000046178>

- Pavlík, R. (2009). A Typology of Assimilations. *SKASE Journal of Theoretical Linguistics*, 6(1), 2-26.
- Pavlík, R. (2016). A usage-based account of /r/-liaison in Standard British English. *Journal of Phonetics* 54, 109-122.
- Phillips, B. S. (2006). *Word Frequency and Lexical Diffusion*. New York, NY: Palgrave Macmillan.
- Picard, M. (1979). The Syllable Boundary in Generative Phonology. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, 24(2), 119-131.
- Pierrehumbert, J. B. (2006). Phonetic Diversity, Statistical Learning, and Acquisition of Phonology. *Language and Speech* 46, 115-154.
<https://doi.org/10.1177/00238309030460020501>
- Pisoni, D. B., Nusbaum, H. C., Luce, P. A., & Slowiczek, L. M. (1985). Speech Perception, Word Recognition, and the Structure of the Lexicon, *Speech Communication* 4, 75-95.
- Plag, I. & Uffmann, C. (2000). Phonological Restructuring in Creole: The Development of Paragoge in Sranan. *CREOLE LANGUAGE LIBRARY* 22, 309-336.
- Posner, R. (1961). Consonantal Dissimilation in the Romance Languages. Oxford: Blackwell.
- Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constrain Interaction in Generative Grammar*. Ms. Rutgers University and the University of Colorado at Boulder.
- Prince, A., and P. Smolensky. (2004). *Optimality Theory*. Oxford: Blackwell.
- Pring, J. T. (1976). More thoughts on the r-link business. *JIPA*, 6, 92-5.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ramsaran, S. (1978). *Phonetic and Phonological Correlates of Style in English: A Preliminary Investigation*. PhD thesis, University of London.
- Ramsaran, S. (1990). RP: fact and fiction. In S. Ramsaran (Eds.) *Studies in the pronunciation of English: a commemorative volume in honour of A C Gimson*, (pp. 178-190). New York: Routledge.
- Recasens, D. (2018). Coarticulation. *Oxford Research Encyclopedia of Linguistics*.
- Riehl, A. K. (2003). American English flapping: Perceptual and acoustic evidence against paradigm uniformity with phonetic features. *Working Papers of the Cornell Phonetics Laboratory* 15, 271-337.

- Roach, P. (1992). *Introducing Phonetics*. Penguin English.
- Roach, P. (2004). British English: Received Pronunciation. *JIPA* 34, 239-245.
- Roach, P. (2009). *English Phonetics and Phonology: A Practical Course* 4th ed. Cambridge: Cambridge University Press.
- Roach, P. Setter, J., & Esling, J. (2011). *Daniel Jones' Cambridge English Pronouncing Dictionary* 18th ed. Cambridge: Cambridge University Press.
- Roca, I., & Johnson, W. (1999). *A course in phonology*. Blackwell Publishing Cambridge: Cambridge University Press.
- Rosewarne, D. (1984). Estuary English. *Times Educational Supplement* 42.
- Rosewarne, D. (1994). Estuary English: Tomorrow's RP? *English Today* 10(1), 3-8.
- Ross, S. (1994). The ins and outs of paragoge and apocope in Japanese-English interphonology. *Second Language Research* 10(1), pp. 1-24.
- Sampson, G. (1980). *Schools of linguistics. Competition and evolution*. London: Hutchinson.
- Sapir, E. (1925). Sounds Patterns in Language. *Language* 1(2), 37-51.
- Saussure, F. de. (1916). *Cours de Linguistique Générale*. Lausanne & Paris: Payot.x
- Selkirk, E. O. (1972). *The phrase phonology of English and French*. PhD dissertation, Massachusetts Institute of Technology.
- Selkirk, E. O. (1982). The Syllable. In H. van der Hulst & N. Smith ((Eds.), *The Structure of Phonological Representations II* (pp. 337-383). Dordrecht: Foris.
- Shapiro, M. (1995). A Case of Distant Assimilation: /str/ → /ʃtr/. *American Speech*, 70(1), 101-107.
- Sheskin, D. J. (2020). *Handbook of Parametric and Nonparametric Statistical Procedures*. 5th edn. CRC Press.
- Shi, R., Gick, N., Kanwischer, D., & Wilson, I. (2005). Frequency and category factors in the reduction and assimilation of function words: EPG and acoustic measures. *Journal of Psycholinguistic Research*, 34, 315–364.
- Shockey, L. (2003). *Sound Patterns of Spoken English*. Malden, MA: Blackwell Pub.
- Sihler, Andrew L. 2000. *Language History: An Introduction*. Amsterdam: John Benjamins.
- Silverman, D. (2004). On the phonetic and cognitive nature of alveolar stop allophony in American English. *Cognitive Linguistics* 15, 69-93.

- Skandera, P. & Burleigh, P. (2016). *A manual of English phonetics and phonology: twelve lessons with an integrated course in phonetics transcription*. Frankfurt: Gunter Narr Verlag.
- Sketch Engine. (n.d.). *Sketch Engine*. (Retrieved from <https://www.sketchengine.eu/>) (Date of access 26 June 2020)
- Skousen, R. (1989). *Analogical Modelling of Language*. Dordrecht: Kluwer.
- Skousen, R. (1992). *Analogy and Structure*. Dordrecht: Kluwer.
- Spurling, J. (2004). *Yod Dropping in Ipswich English: Dialect Attention to the East Anglian County of Suffolk*. Unpublished BA Dissertation. Colchester, UK: University of Essex.
- Sóskuthy, M. (2013). Analogy in the emergence of intrusive-r in English. *English Language and Linguistics* 17, 55-84.
- Stevens, K. (1998). *Acoustics Phonetics*. Cambridge, Massachusetts: MIT Press.
- Stone, M., & Hamlet, S. (1982). Variation in Jaw and Tongue Gestures Observed During Production of Unstressed /ɪ/s and Flaps. *Journal of Phonetics* 10, 401-415.
- Strassel, S. (1998). Variation in American English flapping. In Paradis, C., Vincent, D., Deshaies, D., & Laforest, M. (Eds.), *Papers in sociolinguistics: NWAV-26 à l'Université Laval* (pp. 125-135). Quebec: Nota Bene.
- Tarone, E. (1980). Some Influences on the Syllable Structure of Interlanguage Phonology. *IRAL, International Review of Applied Linguistics in Language Teaching* 18, 139-152.
- Ternes, E. (1986). A Grammatical Hierarchy of Joining. In H. Andersen (Ed.), *Sandhi Phenomena in the Languages of Europe*, (pp. 11-21). Berlin, Boston: de Gruyter Mouton.
- Tollfree, L. (1999). South East London English: Discrete Versus Continuous Modelling of Consonantal Reduction. In P. Foulkes & G. Docherty (Eds.), *Urban Voices: Accent Studies in the British Isles*, (pp. 163-184). London: Arnold.
- Tollfree, L. (2001). Variation and Change in Australian English Consonants: Reduction of /t/. In D. Blair & P. Collins (Eds.), *English in Australia* (pp. 45-67). John Benjamins.
- Trager, G. L. & Smith, H. L. (1951). *An Outline of English Structure*. Battenburg Press: Norman, Okla.
- Trask, R. L. (1996). *A Dictionary of Phonetics and Phonology*. London: Routledge.
- Trubetzkoy, N. S. (1939). *Grundzüge der Phonologie*. Göttingen: Vandenhoeck & Ruprecht. Also [Travaux du Cercle Linguistique de Prague 7]. Prague. Trans. Christiane A. M. Baltaxe, *Principles of Phonology*. Berkeley: University of California Press.

- Trudgill, P. (1974). *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.
- Trudgill, P. (1986). *Dialects in Contact*. Oxford: Blackwell.
- Trudgill, P. (2000). *Sociolinguistics: An Introduction to Language and Society*. London: Penguin.
- Trudgill, P. (2001). Received Pronunciation: sociolinguistic aspects. *Studia Anglica Posnaniensia* 36, pp. 3–13.
- Trudgill, P. (2008). The Dialect of East Anglia: Phonology. In B. Kortmann & C. Upton (Eds.) *The British Isles*, 178-193. Berlin, New York: De Gruyter Mouton.
- Trudgill, P. & Hannah, J. (2008). *International English: A Guide to Varieties of Standard English*. London: Arnold.
- Tummers, J., Heylen, C., & Geeraerts, D. (2005). Usage-based Approaches in Cognitive Linguistics: A Technical State of the Art. *Corpus Linguistics and Linguistics Theory* 1, 225-261.
- Turk, A. (1992). The American English flapping rule and the effect of stress on stop consonant durations. *Working Papers of the Cornell Phonetics Laboratory* 7, 103-133.
- Uffmann, C. (2007a). Intrusive /r/ and optimal epenthetic consonants. *Language Sciences* 29, 451-476.
- Uffmann, C. (2007b). *Vowel Epenthesis in Loanword Adaptation*. Tübingen: Niemeyer.
- UK Government. (n.d.). *The Rt Hon Ed Balls*. [The Rt Hon Ed Balls - GOV.UK \(www.gov.uk\)](http://www.gov.uk) (Date of access 20 May 2018).
- UK Government. (n.d.). *The Rt Hon George Osborne*. <https://www.gov.uk/> ([hace 2 años](#))
[The Rt Hon George Osborne - GOV.UK \(www.gov.uk\)](http://www.gov.uk) (Date of access 20 May 2018).
- UK Government. (n.d.). *The Rt Hon David Cameron*, [The Rt Hon David Cameron - GOV.UK \(www.gov.uk\)](http://www.gov.uk) (Date of access 20 May 2018).
- UK Politics. (n.d.). *Cooper, Yvette*. [Yvette Cooper - The Wife of Britain's Second Most Popular Politician \(politics.co.uk\)](http://www.politics.co.uk) <https://www.politics.co.uk/> (Date of access 20 May 2018).
- University of Amsterdam. (n.d.). *Praat*. (Retrieved from <https://www.fon.hum.uva.nl/praat/>) (Date of access 30 June 2018).
- Vennemann, T. (1971). Natural generative phonology. Paper read at the annual meeting of the Linguistic Society of America, St. Louis, MO.

- Vennemann, T. (1972). Rule inversion. *Lingua* 29, 209-242.
- Viollain, Céline. (2014). *Sociophonologie de l'anglais contemporain en Nouvelle-Zélande: Corpus et dynamique des systèmes*. Ph.D. dissertation, Université de Toulouse Jean Jaurès.
- Wagner, S. E., & Buchstaller, I. (2018). *Panel studies of variation and change*. Routledge.
- Wang, J., & Smith, N. (2013). *Studies in Chinese Phonology*. Berlin, Boston: De Gruyter Mouton.
<https://doi.org/10.1515/9783110822014>
- Warner, N. & Tucker, B. V. (2011). Phonetic Variability of Stops and Flaps in Spontaneous and Careful Speech. *The Journal of the Acoustical Society of America* 130, 1606-1617.
<https://doi.org/10.1121/1.3621306>
- Watt, D., & Milroy, L. (1999). Patterns of Variation and Change in Three Newcastle Vowels: Is This Dialect Levelling? In P. Foulkes & G. Docherty (eds.), *Urban Voices: Accent Studies in the British Isles* (pp. 25-46). London: Longman.
- Wilner, J. (2003). *Wortubuku ini Sranan Tongo (Sranan Tongo – English dictionary)*. 4th edn. Paramaribo: Summer Institute of Linguistics.
- Wells, J. C. (1962). A specimen of British English. *Maître Phonétique*. 2-5.
- Wells, J. C. (1982). *Accents of English*, 3 vols. Cambridge: Cambridge University Press.
<https://doi-org.ezproxy.lib.gla.ac.uk/10.1017/CBO9780511611759>
- Wells, J. C. (1990). Syllabification and Allophony. In S. Ramsaran (Ed), *Studies in the Pronunciation of English: A Commemorative Volume in Honour of A. C. Gimson* (pp. 76-86). London and New York: Routledge.
- Wells, J. C. (1994a). The Cockneyfication of RP? In M. Gunnel, & J. Nils-Lennart (eds.), *Nonstandard varieties of language*. Papers from the Stockholm Symposium, 11-13 April 1991. Stockholm: Almqvist & Wiksell International.
- Wells, J. C. (1994b). Transcribing Estuary English: a discussion document. In *Speech Hearing and Language*, 8, 259-26.
- Wells, J. C. (1997). Whatever happened to Received Pronunciation? In C. Medina-Casado & C. Soto-Palomo (eds.), *II Jornadas de Estudios Ingleses*. Jaén: Universidad de Jaén, Servicio de Publicaciones. 19-28.
- Wells, J. C. (2000). *Longman Pronunciation Dictionary*. 2nd Edition. London: Longman.
- Wells, J. C., & Colson, G. (1971). *Practical Phonetics*. London: Pitman Publishing.

- Williams, A. & Kerswill, P. (1999). Dialect Levelling: Change and Continuity in Milton Keynes, Reading and Hull. In P. Foulkes & G. Docherty (Eds.), *Urban Voices: Accents Studies in the British Isles* (pp. 141-162). London: Longman.
- Woods, H. B. (1991). Social Differentiation in Ottawa English. In J. Cheshire (Ed.), *English Around the World: Sociolinguistics Perspectives* (pp. 134-149). Cambridge: Cambridge University Press.
- Yip, M. (2002). *Tones*. Cambridge: Cambridge University Press.
- YouTube. (n.d.). *Liar Politicians*. Retrieved from <https://www.youtube.com/user/liarpoliticians/featured> (Date of access 8 September 2017)
- Yuan, J. H., & Chen, Y. (2014). 3rd Tone Sandhi in Standard Chinese: A Corpus Approach. *Journal of Chinese Linguistics*, 42(1), 218-237.
- Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, 27(1), 153-201.
- Zipf, G. K. (1929). Relative Frequency as a Determinant of Phonetic Change. *Harvard Studies in Classical Philology* 15, 1-95.
- Zue, V. W., & Laferriere, M. (1979). Acoustic study of medial /t, d/ in American English. *Journal of the Acoustic Society of America* 66, 1038-1050. <https://doi-org.ezproxy.lib.gla.ac.uk/10.1121/1.383323>
- Zue, V. W., & Seneff, S. (1996). Transcription and Alignment of the TIMIT Database. In H. Fujisaki (ed.) *Recent Research Toward Advanced Man-Machine Interface Through Spoken Language* (pp. 464-447). Amsterdam: Elsevier.