UNIVERSIDAD DE MURCIA

Escuela de Doctorado

TESIS DOCTORAL

Polisemia Multimodal en los Verbos de Percepción en Inglés: Un Enfoque Experimental y Basado en Corpus Aplicado a los Sentidos de Touch, Taste y Smell

Multimodal Polysemy in English Perception Verbs: An Experimental and Corpus-Based Approach to the Senses of Touch, Taste and Smell

AUTORA Irene Bolumar Martínez

DIRECTORES Javier Valenzuela Manzanares
Daniel Alcaraz Carrión

2025

UNIVERSIDAD DE MURCIA Escuela de Doctorado

TESIS DOCTORAL

Polisemia Multimodal en los Verbos de Percepción en Inglés: Un Enfoque Experimental y Basado en Corpus Aplicado a los Sentidos de Touch, Taste y Smell

Multimodal Polysemy in English Perception Verbs: An Experimental and Corpus-Based Approach to the Senses of Touch, Taste and Smell

AUTORA Irene Bolumar Martínez

DIRECTORES Javier Valenzuela Manzanares
Daniel Alcaraz Carrión

2025

UNIVERSIDAD DE MURCIA

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA EN MODALIDAD DE COMPENDIO O ARTÍCULOS PARA OBTENER EL TITULO DE DOCTOR/A

Aprobado por la Comisión General de Doctorado el 19 de octubre de 2022.

Yo, Dª Irene Bolumar Martínez, habiendo cursado el Programa de Doctorado en Artes y Humanidades: Bellas Artes, Literatura, Teología, Traducción e Interpretación y Lingüística General e Inglesa de la Escuela Internacional de Doctorado de la Universidad de Murcia (EIDUM), como autor/a de la tesis presentada para la obtención del título de Doctor/a titulada:

Multimodal Polysemy in English Perception Verbs: An Experimental and Corpus-Based Approach to the Senses of Touch, Taste and Smell/Polisemia Multimodal en los Verbos de Percepción en Inglés: Un Enfoque Experimental y Basado en Corpus Aplicado a los Sentidos de Touch, Taste y Smell

y dirigida por:

D.: Javier Valenzuela Manzanares
D.: Daniel Alcaraz Carrión

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.

Además, al haber sido autorizada como prevé el artículo 29.8 del reglamento, 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.

Murcia, a 3 de diciembre de 2025

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

Esta DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD debe ser insertada en la quinta hoja, después de la portada de la tesis presentada para la obtención del título de Doctor/a.

"Gestures are an integral part of language as much as are words, phrases, and sentences—gesture and language are one system"

Hand and Mind: What Gestures Reveal about Thought – David McNeill (1992)

"Metaphor is as much a part of our functioning as our sense of touch, and as precious"

Metaphors We Live By – George Lakoff and Mark Johnson (1980)

"Everything we do, everything we feel, everything we know, is mediated through the senses"

Sensory Linguistics: Language, Perception, and Metaphor – Bodo Winter (2019)

Acknowledgements

The journey of a thesis is filled with enthusiasm and learning, but also with sacrifice and challenging moments. Fortunately, I have received invaluable support and collaboration from many individuals and institutions throughout this process. For this reason, I would like to express my sincere gratitude to all those who have supported me in various ways during the development of this project.

First of all, this research would not have been possible without the funding provided by the Fundación Séneca and the Fulbright Program. I would also like to extend my appreciation to the staff at both institutions, who consistently demonstrated their willingness to help. The support I received from these institutions not only made this thesis possible but also enriched my academic journey, both professionally and personally.

This thesis would not have been possible without the support of my supervisors, Javier Valenzuela and Daniel Alcaraz-Carrión. I first met Javier, who instilled in me his passion for studying the relationship between language and cognition. In particular, I am grateful for the way he demonstrates that researching what you love and sharing what you know is "the best job in the world." Shortly afterwards, Daniel joined, bringing a fresh perspective and always sharing his immense motivation and intellectual curiosity. I would also like to acknowledge his role in encouraging my critical thinking at every stage of the project. To both of them, I am deeply grateful for their patience, dedication and trust in this work; I truly value every minute shared. My professional admiration for them has undoubtedly extended to a personal level.

Further, I would like to express my heartfelt gratitude to Susan Goldin-Meadow, who welcomed me into her laboratory as a Fulbright visiting researcher. Her guidance not only brought clarity and rigor to my ideas but also deepened my understanding of gestures. Her

support greatly enriched the results of this thesis. Beyond her invaluable mentorship, I remain deeply appreciative of the kindness with which she welcomed me to Chicago.

Beyond the support of my supervisors, I would like to acknowledge the important role of my home research group, the Daedalus Lab. Every member of the group has contributed in some way throughout the course of this thesis. Cristóbal has always been generous with his support in various academic matters, trusting in the value of my contributions to the group. My colleagues Rosa and Águeda have shared every achievement with me and have been a constant source of support during the most challenging and uncertain moments of my predoctoral journey. I would also like to thank Paula for her willingness to help in any way, and I look forward to continuing to foster the collaborative spirit of the group alongside her. Finally, I am grateful to Raúl and Brian, who have always been ready to share their technical expertise—especially Raúl, whose help has been indispensable to this thesis. In short, I am thankful to all of them for the moments we have shared during my predoctoral stage, and I hope that such experiences will continue in future stages.

Moreover, I have been fortunate to find incredible colleagues in the academic world outside my own institution. I would like to acknowledge the warm welcome and support I received from the members of the Goldin-Meadow Laboratory. Jess, Zena, Uli, Casey, Michelle, Marine, Hannah, Nina and Anjana—you were not only an incredible source of professional support but also made me feel at home. In Chicago, I also had the pleasure of sharing moments with other colleagues in the same predoctoral stage. Taryn, thank you for welcoming me with open arms from the very beginning. Leire, Lucía and Ana, sharing the Fulbright experience with you was truly a pleasure. Without a doubt, many of you became true friends.

Mi mayor agradecimiento es, sin duda, para mi familia. Gracias, mamá y papá, por darlo todo por mí. A mi madre, gracias por enseñarme el valor de ser una mujer independiente desde que

tengo uso de razón y por ser el claro ejemplo de que todo sacrificio tiene recompensa. A mi padre, gracias por enseñarme el valor que tienen la calma y el humor en la vida y por inculcarme el valor del trabajo bien hecho. A mi hermano Samuel, gracias por cuidarme siempre y por hacerme saber el valor de mi esfuerzo. A los tres, por estar siempre a mi lado.

No puedo olvidarme de mis amigas, quienes son capaces de hacerme desconectar de lo profesional y me ayudan a olvidar mis mayores preocupaciones cuando estamos juntas. A mis amigas de siempre—Pilar, Irene, Clara, Majo, Isabel, Carmen, Lorena, Paula y Paloma—gracias por acompañarme durante todos estos años. Tampoco olvido a las amistades que me llevo de los primeros años de universidad—Almudena, Ester, Paula, Elena, Anabel y Laura—espero que podamos seguir reuniéndonos y ver que nada ha cambiado.

Before concluding, I would like to include in my acknowledgments the journals that have trusted my work and the reviewers for their comments, which undoubtedly improved the final version of this thesis. I would also like to thank the Spanish Association of Cognitive Linguistics (AELCO) for awarding me the prize for the best predoctoral presentation, which eventually became a publication in *Review of Cognitive Linguistics*.

Finally, I would like to thank in advance the members of the evaluation committee of this thesis for their time and dedication in reviewing my research. It is a privilege to benefit from their extensive expertise and insightful feedback.

Table of Contents

List of Figures

List of Tables

Abstracts

Abstract

The study of polysemy—phenomenon whereby a word has multiple related meanings—offers a unique opportunity to understand the cognitive mechanisms that underlie the structuring and generation of meaning. The essential aspect of polysemy does not lie in the multiplicity of meanings, but in the mechanisms responsible for creating and connecting them. These related, non-prototypical meanings are conceived as natural and systematic semantic extensions of the prototypical sense of a polysemous word. Polysemy is thus understood not only as a lexical phenomenon but also as a conceptual one, dependent on human cognitive structure.

Interest in the study of polysemy has generated a wide range of work in both linguistics and psychology. Among the central research questions is how interlocutors, apparently effortlessly, are able to select the appropriate sense of polysemous words. This question is relevant because, despite the potential ambiguity that polysemy may introduce into everyday communication, it remains unclear how humans manage to disambiguate the meaning of polysemous expressions so effectively.

With respect to meaning disambiguation, research has traditionally focused on the lexical items and syntactic structures that help interlocutors select the appropriate sense. Despite abundant research on both polysemy and language ambiguity, problematic situations arise, such as when the ambiguous expression is 'unpredictable' (i.e., uninterpretable). That is, when the verbal message is insufficient and access to contextual information is limited.

A potentially contributing factor lies in the multimodal nature of human communication, given that language involves much more than words. For this reason, this thesis adopts the perspective that language is inherently multimodal: all communicative modalities contribute to meaning construction and shape how we communicate, offering cues beyond words that help speakers convey their messages and listeners interpret the meaning expressed.

Given that gestures naturally co-occur with speech and can be used both to emphasize information conveyed verbally and to express ideas that go beyond the speaker's words, some researchers have begun to approach the study of ambiguous expressions from a multimodal perspective (Holler & Beattie, 2003; Holle & Gunter, 2007; Hinnell & Parrill, 2020). However, no work has explored whether the meaning of a polysemous word can be disambiguated through hand gestures. For this reason, this thesis focuses on gesture as a potential element contributing to meaning disambiguation in polysemous expressions.

This thesis posits that the relationship between speech and gesture may be crucial for resolving ambiguity in polysemous expressions. Nevertheless, the role of gesture in the expression and disambiguation of polysemous words remains largely unexplored. This thesis addresses this gap by analyzing how speakers use both speech and gesture to distinguish among related meanings of polysemous verbs. The main objective of this investigation, within the framework of Cognitive Linguistics, is to contribute to research on polysemy by exploring the role of hand gestures and linguistic markers in speech. In particular, the thesis examines three English perception verbs: *touch*, *taste* and *smell*. The research focuses on these verbs because the domain of perception is considered highly polysemous. In other words, these verbs are especially interesting because they are used not only to describe how we physically interact with the world through the senses, but also figuratively. Although the study of perception verbs has generated abundant research, most work has focused exclusively on written forms.

With this in mind, each of the three articles included in this thesis examines the role of speech and gesture in distinguishing the meanings of the perception verbs *touch*, *taste* and *smell*. The first and third articles identify, through large-scale corpus analyses, the strategies speakers use in speech and gesture to distinguish the different meanings of these verbs. The second article, presenting three experimental studies, analyzes how observers disambiguate the meaning of

touch using gestural information. By collecting spontaneous audiovisual data and incorporating laboratory data, this thesis brings a multimodal perspective to the phenomenon of polysemy.

In particular, the multimodal approach of this thesis combines the analysis of linguistic markers that accompany these polysemous verbs with that of *co-speech* gestures (i.e., gestures coordinated with the speaker's oral production). The linguistic markers examined include intensifiers and negation. This decision is grounded in previous studies (Argaman, 2009; Carrillo-de-Albornoz & Plaza, 2013) showing that such markers are frequently associated with emotional attributes of language. Co-speech hand gestures are included because they reveal dimensions of the speaker's conceptualization that are not accessible through the spoken message (Goldin-Meadow, 2003; Kendon, 2004).

Within the general objectives of this thesis, two specific research questions are addressed:

- Do speakers distinguish between the different meanings of polysemous expressions in speech and gesture?
- Can observers disambiguate the meaning of polysemous expressions through gesture?

The three research papers presented in this thesis share a multimodal approach to meaning disambiguation and are methodologically complementary. The first article investigates whether speakers use multimodal information to differentiate the physical and emotional meanings of the polysemous verb *touch*. The verb was searched in the multimodal corpus NewsScape and 302 co-speech gestures were identified. In each case the following were annotated: (1) the meaning of the verb, (2) what the speaker physically touched during the gesture (an external element, *other-touch*, or themselves, *self-touch*), (3) the personal pronoun that followed the verb and (4) the use of intensifiers and negation. The second article examines whether observers use gestural information to interpret the meaning of *touch* in ambiguous contexts. To address this question, three studies were conducted: the first tests whether participants

distinguish the meaning of the verb by viewing hand gestures alone; the second explores which combinations of gesture location and handshape are associated with the physical and emotional meanings of *touch*; the third investigates whether participants decide the meaning more quickly when a gesture is presented and whether reaction time varies as a function of the gesture observed. Finally, the third article, building on the corpus methodology of the first article, explores whether gestures and linguistic markers help to distinguish between the literal and figurative meanings of *taste* and *smell*. To determine whether the patterns found for *touch* generalize to other verbs, this study (1) compares the gestures associated with *taste* and *smell*, (2) examines the motivation of the most frequent gestures linked to these verbs and (3) describes the relationship between markers in speech and the meanings of both verbs.

Taken together, the three papers show differences in the relationship between gesture and speech for the perception verbs *touch*, *taste* and *smell*. Three main findings emerge from the first article. First, when speakers express the physical meaning of *touch*, they tend to physically touch an external element (other-touch), whereas when expressing the emotional meaning they tend to touch their own body (self-touch). Second, the most frequent co-speech gesture—touching one's chest (*chest-touching* gesture)—is associated with the emotional meaning, revealing the metaphor THE HEART IS A CONTAINER FOR EMOTIONS (Lakoff & Johnson, 1980; Kövecses, 1986). Third, this corpus study shows that, while the physical meaning of *touch* typically co-occurs with a wide variety of personal pronouns and negative words, the emotional meaning appears mainly with the pronoun *me* and is modified by intensifiers. In the second article, the results demonstrate that gesture helps observers to disambiguate the meaning of a polysemous verb such as *touch*. Gesture location and handshape bias the interpretation of the verb when the verbal message is ambiguous. In the third article, the gestural analysis reveals that, although gestures help to distinguish the meanings of *touch*, they do not display distinctive patterns between the literal and figurative meanings of *taste* and *smell*. The analysis of markers

in speech shows that object quantifiers are the most frequent markers when expressing literal meanings of both verbs.

Across the three studies included in this thesis, two conclusions stand out: (1) gestures help to distinguish the meanings of *touch* but not those of *taste* and *smell* and (2) overall—with differences depending on the verb—linguistic markers can contribute to differentiating the meanings of the three perception verbs.

The gestural analysis in this thesis confirms that gesture plays a key role in differentiating the meanings of *touch* (other-touch vs. self-touch), while not exhibiting clearly distinguishing patterns between the literal and figurative uses of *taste* and *smell*. In the case of *touch*, the relationship between a specific type of self-touch—chest-touching gesture—and the emotional meaning reflects the Western conceptualization of the heart as the seat of emotions. Besides, gesture location and handshape are confirmed as essential components that influence meaning interpretation when the verbal message is ambiguous. As for gestures co-occurring with *taste* and *smell*, they are not systematically associated with specific meanings.

The analysis of linguistic markers confirms that polysemy is not located solely in the word under observation, but also manifests in other elements accompanying the polysemous expression (Ibarretxe-Antuñano, 1999). In the case of *touch*, negation is associated with the physical meaning and intensifiers with the emotional one. By contrast, for *taste* and *smell*, only object quantifiers are linked to literal meanings, suggesting that markers in speech provide more discriminating information in the case of *touch*.

While the results of this thesis are relevant, several limitations exist. Despite obtaining a comparable number of occurrences per perception verb, the gestural datasets for *taste* and *smell*—especially the former—are considerably smaller than that for *touch*. Accordingly, the conclusions concerning *taste* and *smell* should be interpreted with greater caution. Among other

limitations, it should be borne in mind that the nature of the television recordings used for the analyses prevents all manual gestures from being coded. In short, any gesture occurring outside the camera shot cannot be annotated. The most significant limitation at the time of writing this thesis (November 2025) is that the NewsScape search engine is no longer functioning, which partially affects the replicability of the searches reported here. Nevertheless, this situation may be resolved at a later date.

Looking ahead, it would be interesting to extend this line of research to other ambiguous and polysemous expressions in order to strengthen the empirical basis of the findings of this thesis. For instance, particularly in the case of *touch*, it would be useful to incorporate new levels of analysis—such as prosodic patterns—to provide a more precise multimodal description. It would also be pertinent to explore, from a cross-cultural perspective, the use of gestures such as the chest-touching gesture, to determine whether they exhibit similar usage across different languages and communities.

In short, this thesis explores the role of co-speech gestures and linguistic markers in differentiating the meanings of the polysemous perception verbs *touch*, *taste* and *smell*. By adopting a multimodal approach, this thesis combines lexical and gestural analyses to study polysemous expressions, addressing a domain hitherto unexplored.

Resumen

El estudio de la polisemia, fenómeno por el que una palabra tiene varios significados relacionados entre sí, ofrece una oportunidad única para comprender los mecanismos cognitivos que subyacen a la estructuración y generación del significado. El aspecto esencial de la polisemia no reside en la multiplicidad de significados, sino en los mecanismos responsables de crearlos y conectarlos. Estos significados relacionados no prototípicos se conciben como extensiones semánticas naturales y sistemáticas del sentido prototípico de una palabra polisémica. Por lo tanto, la polisemia se concibe no solo como un fenómeno léxico, sino también conceptual, dependiente de la estructura cognitiva humana.

El interés en el estudio de la polisemia ha dado lugar a una gran variedad de trabajos tanto en el campo de la lingüística como de la psicología. Entre las preguntas principales de investigación de estos trabajos se encuentra cómo los interlocutores, sin esfuerzo aparente, son capaces de seleccionar el sentido adecuado de las palabras polisémicas. Esta pregunta es relevante ya que, a pesar de la posible ambigüedad que la polisemia introduce en conversaciones cotidianas, sigue sin estar demasiado claro cómo los seres humanos logran desambiguar el significado de expresiones polisémicas con tanta eficacia.

Respecto a la desambiguación del significado, la investigación se centra tradicionalmente en los elementos léxicos y las estructuras sintácticas que ayudan a los interlocutores a seleccionar el sentido adecuado. A pesar de la abundante investigación tanto sobre polisemia como sobre ambigüedad del lenguaje, se dan situaciones problemáticas, como cuando la expresión ambigua es ‘impredecible’ (i.e., no se puede interpretar). Es decir, cuando el mensaje verbal resulta insuficiente y el acceso a la información contextual es limitado.

Un factor potencialmente contribuyente radica en la naturaleza multimodal de la comunicación humana, dado que el lenguaje implica mucho más que palabras. Por ello, esta tesis adopta la

perspectiva de que el lenguaje es inherentemente multimodal: todas las modalidades comunicativas contribuyen a la construcción del significado y configuran la forma en que nos comunicamos, ofreciendo pistas adicionales, más allá de las palabras, que ayudan tanto a los hablantes a transmitir sus mensajes como a los oyentes a interpretar el significado expresado.

Teniendo en cuenta que los gestos coocurren naturalmente con el habla y pueden emplearse tanto para enfatizar la información transmitida verbalmente como para expresar ideas que van más allá de las palabras del hablante, algunos investigadores han comenzado a abordar el estudio de expresiones ambiguas desde una perspectiva multimodal (Holler & Beattie, 2003; Holle & Gunter, 2007; Hinnell & Parrill, 2020). Sin embargo, no existen trabajos que exploren si el significado de una palabra polisémica puede desambiguarse mediante el gesto. Por esta razón, esta tesis se centra en el potencial del gesto como elemento capaz de contribuir a la desambiguación del significado en expresiones polisémicas.

Esta tesis considera que la relación entre habla y gesto puede ser fundamental para resolver la ambigüedad en expresiones polisémicas. No obstante, el papel del gesto en la expresión y desambiguación de palabras polisémicas sigue estando en gran medida inexplorado. Esta tesis aborda este vacío analizando cómo los hablantes utilizan tanto el habla como el gesto para distinguir entre significados relacionados de verbos polisémicos. El objetivo principal de este trabajo, dentro del marco de la Lingüística Cognitiva, es contribuir a la investigación sobre polisemia mediante la exploración del papel de los gestos manuales y de los marcadores lingüísticos en el habla. En particular, esta tesis examina tres verbos de percepción del inglés: *touch*, *taste* y *smell* (en español, tocar, saborear y oler, respectivamente). La investigación se centra en estos verbos porque el dominio de la percepción se considera altamente polisémico. En otras palabras, estos verbos resultan especialmente interesantes porque se utilizan no solo para describir cómo interactuamos con el mundo a través de los sentidos físicos, sino también

de forma figurada. Aunque el estudio de los verbos de percepción ha generado abundante investigación, la mayoría de los trabajos se ha centrado exclusivamente en la modalidad escrita.

Con esto en mente, cada uno de los tres artículos incluidos en esta tesis examina el papel del habla y el gesto en la diferenciación de los significados de los verbos de percepción *touch*, *taste* y *smell*. El primer y el tercer artículo identifican, a través de un análisis de corpus a gran escala, las estrategias que los hablantes emplean en el habla y el gesto para distinguir entre los distintos significados de estos verbos. El segundo artículo, compuesto por tres estudios experimentales, analiza cómo los observadores desambiguan el significado de *touch* utilizando la información gestual. Mediante la recopilación de datos audiovisuales de naturaleza espontánea y la incorporación de datos de laboratorio, esta tesis aporta una perspectiva multimodal al fenómeno de la polisemia.

En particular, el enfoque multimodal de esta tesis combina el análisis de marcadores lingüísticos que acompañan a estos verbos polisémicos y el de los gestos manuales *co-habla* (i.e., coordinados con la producción oral del hablante). Los marcadores lingüísticos examinados incluyen intensificadores y negación. Esta decisión se basa en estudios previos (Argaman, 2009; Carrillo-de-Albornoz & Plaza, 2013), que muestran que dichos marcadores se asocian con frecuencia a atributos emocionales del lenguaje. Los gestos co-habla se incluyen porque revelan dimensiones de la conceptualización del hablante que no resultan accesibles mediante el estudio exclusivo del mensaje oral (Goldin-Meadow, 2003; Kendon, 2004).

Dentro de los objetivos generales de esta tesis, se abordan dos preguntas de investigación específicas:

- ¿Distinguen los hablantes entre los diferentes significados de las expresiones polisémicas en el habla y el gesto?

- ¿Pueden los observadores desambiguar el significado de las expresiones polisémicas a través del gesto?

Los tres artículos de investigación presentados en esta tesis comparten un enfoque multimodal de la desambiguación del significado y son metodológicamente complementarios. El primer artículo investiga si los hablantes utilizan información multimodal para diferenciar los significados físico y emocional del verbo polisémico *touch*. El verbo se buscó en el corpus multimodal NewsScape y se identificaron 302 gestos co-habla. En cada caso se anotó: (1) el significado del verbo, (2) lo que el hablante toca físicamente durante el gesto (contacto con un elemento externo, *other-touch*, o consigo mismo, *self-touch*), (3) el pronombre personal que sigue al verbo y (4) el uso de intensificadores y negaciones. El segundo artículo examina si los observadores utilizan la información gestual para interpretar el significado del verbo *touch* en contextos ambiguos. Para abordar esta cuestión se realizaron tres estudios: el primero comprueba si los participantes distinguen el significado del verbo únicamente viendo gestos manuales; el segundo explora qué combinaciones de localización del gesto y forma de la mano se asocian con los significados físico y emocional de *touch*; el tercero investiga si los participantes deciden el significado del verbo más rápido cuando se presenta un gesto y si el tiempo de reacción varía en función del gesto observado. Por último, el tercer artículo, basándose en la metodología de corpus del primer artículo, explora si los gestos y los marcadores lingüísticos ayudan a distinguir entre los significados literales y figurados de *taste* y *smell*. Para determinar si los patrones hallados para *touch* se generalizan a otros verbos, este estudio (1) compara los gestos asociados con *taste* y *smell*, (2) examina la motivación de los gestos más frecuentes vinculados a estos verbos y (3) describe la relación entre marcadores en el mensaje oral y los significados de ambos verbos.

En conjunto, los tres artículos muestran diferencias en la relación entre gesto y habla en los verbos de percepción *touch*, *taste* y *smell*. En el primer artículo surgen tres hallazgos

principales. En primer lugar, cuando los hablantes expresan el significado físico de *touch*, tienden a tocar físicamente un elemento externo (other-touch), mientras que, al expresar el significado emocional, tienden a tocar su propio cuerpo (self-touch). En segundo lugar, el gesto co-habla más frecuente—tocarse el pecho (*chest-touching*)—se asocia con el significado emocional, revelando la metáfora EL CORAZÓN ES UN CONTENEDOR DE EMOCIONES (Lakoff & Johnson, 1980; Kövecses, 1986). En tercer lugar, este estudio de corpus muestra que, mientras el significado físico de *touch* suele coexistir con una amplia variedad de pronombres personales y palabras de negación, el significado emocional aparece principalmente con el pronombre *me* (correspondiente al pronombre personal me en español) y se modifica con intensificadores. En el segundo artículo, los resultados demuestran que el gesto ayuda a los observadores a desambiguar el significado de un verbo polisémico como *touch*. La localización del gesto y la forma de la mano influyen en la interpretación del verbo cuando el mensaje verbal es ambiguo. En el tercer artículo, el análisis gestual revela que, aunque los gestos ayudan a diferenciar los significados de *touch*, no muestran patrones distintivos entre los significados literales y figurados de *taste* y *smell*. El análisis de marcadores en el habla muestra que los cuantificadores de objeto son los más frecuentes al expresar significados literales de ambos verbos.

El conjunto de los tres estudios incluidos en esta tesis destaca dos conclusiones: (1) los gestos ayudan a distinguir los significados de *touch*, pero no los de *taste* y *smell*; y (2) de forma global, aunque de manera diferente según el verbo, los marcadores lingüísticos pueden contribuir a diferenciar los significados de los tres verbos de percepción.

El análisis gestual de esta tesis confirma que el gesto desempeña un papel clave en la diferenciación de los significados de *touch* (other-touch vs. self-touch), sin mostrar patrones claramente diferenciadores entre los usos literales y figurados de *taste* y *smell*. En el caso de *touch*, la relación entre un tipo concreto de self-touch—tocarse el pecho (chest-touching)—y el significado emocional refleja la conceptualización occidental del corazón como lugar donde

se alojan las emociones. Asimismo, la localización del gesto y la forma de la mano se confirman como componentes esenciales que influyen en la interpretación cuando el mensaje verbal es ambiguo. En cuanto a los gestos que coocurren con *taste* y *smell*, no se asocian sistemáticamente con significados específicos.

El análisis de los marcadores lingüísticos confirma que la polisemia no se localiza únicamente en la palabra observada, sino que también se manifiesta en otros elementos que acompañan a la expresión polisémica (Ibarretxe-Antuñano, 1999). En el caso de *touch*, la negación se asocia con el significado físico y los intensificadores con el emocional. En cambio, para *taste* y *smell*, solo los cuantificadores de objeto se vinculan con los significados literales, lo que sugiere que los marcadores proporcionan información más discriminante en el caso de *touch*.

Aunque los resultados de esta tesis son relevantes, existen varias limitaciones. A pesar de obtener un número comparable de ocurrencias por verbo de percepción, los conjuntos de datos gestuales de *taste* y *smell*—especialmente del primer verbo—tienen un tamaño considerablemente menor en comparación con *touch*. Por tanto, las conclusiones relativas a *taste* y *smell* se deben interpretar con mayor cautela. Entre otras limitaciones, se debería tener en cuenta que la naturaleza de las grabaciones de televisión usadas para los análisis de la tesis impide codificar todos los gestos manuales. En pocas palabras, cualquier gesto que ocurra fuera de un plano de cámara no puede anotarse. Sin duda, la mayor limitación en el momento de la redacción de esta tesis (noviembre de 2025), es que el motor de búsqueda de NewsScape dejó de estar operativo, lo que afecta parcialmente la replicabilidad de las búsquedas de esta tesis. No obstante, esta situación podría resolverse en fechas posteriores.

De cara al futuro, sería interesante ampliar esta línea de estudio a otras expresiones ambiguas y polisémicas para reforzar la base empírica de los resultados de esta tesis. Por ejemplo, especialmente en el caso de *touch*, resultaría útil incorporar nuevos niveles de análisis—como

los patrones prosódicos—para ofrecer una descripción multimodal más precisa. También sería pertinente explorar, desde una perspectiva intercultural, el uso de gestos como tocarse el pecho (chest-touching), a fin de determinar si mantienen un uso similar en diferentes lenguas y comunidades.

En resumen, esta tesis explora el papel de los gestos coocurrentes con el habla y de los marcadores lingüísticos en la diferenciación de los significados de los verbos de percepción polisémicos *touch*, *taste* y *smell*. Al adoptar un enfoque multimodal, constituye una investigación que integra el análisis léxico y el análisis gestual en el estudio de expresiones polisémicas, abordando un ámbito hasta ahora inexplorado.

1. Introduction

1.1. General Introduction

Language plays a central role in human communication, since the complex system that underlies its use enables us to express meaning in diverse and efficient ways. One of the most remarkable features of language is its versatility, clearly illustrated by the fact that a single word can convey multiple meanings (e.g., the noun *spring* can refer to a coil, a natural source of water or the season before summer). The study of language becomes particularly challenging when the relationship between form and meaning does not follow a one-to-one correspondence. This capacity to associate a single form with multiple meanings may lead to ambiguity. The study of language ambiguity is essential for gaining insight into language as a cognitive system, as it allows us to explore how speakers produce and interpret meaning.

Ambiguity in language can arise from a variety of sources, such as polysemy, homonymy or an ineffective use of prosody. Among these sources, the phenomenon of polysemy—multiple related meanings coexisting under the same form—is particularly interesting. Polysemy offers a unique window into the cognitive mechanisms underlying the structuring and generation of meaning. An illustration of polysemy is the question 'can I close this window?,' which may be used to refer either to a framed area on a computer screen or to a physical opening in a building. Despite the potential ambiguity that polysemy may introduce in daily interactions, speakers and listeners usually infer the intended meaning of ambiguous expressions with apparent ease. Although contextual cues are crucial in this process, it remains unclear how humans manage to disambiguate meaning so efficiently.

A contributing factor may lie in the multimodal nature of human communication. Language involves much more than words. This thesis adopts the view that language is inherently multimodal, as it simultaneously unfolds across multiple channels—not only spoken or written—but also through elements such as facial expressions and gestures. All communication

modalities contribute to the construction of meaning and shape how humans communicate, offering additional cues beyond spoken words that help speakers convey messages and listeners interpret intended meaning.

In particular, the interaction between speech and gesture can be crucial for resolving ambiguity. Gestures naturally co-occur with speech, functioning as a unified system in communication, and they can be used not only to emphasize the information conveyed in speech but also to express ideas that extend beyond or diverge from the speaker's words. Thus, even when the spoken content is potentially ambiguous, speakers can transmit their intended message effectively by providing complementary information through gesture.

While prior research on polysemy has largely focused on lexical analyses, the role of gesture in the expression and disambiguation of polysemous items remains largely underexplored. This thesis addresses this research gap by investigating how speakers use both speech and gesture to distinguish between related meanings of polysemous verbs. The main aim of this pioneer investigation is to contribute to our understanding of polysemy by exploring the role of hand gestures and linguistic markers in speech, particularly in the use of the perception verbs *touch*, *taste* and *smell*. Overall, this thesis argues that multimodality—especially hand gestures—plays a key role in disambiguating polysemous expressions, offering new insights into how the meanings of such verbs are distributed and interpreted across modalities.

This research focuses on the perception verbs *touch*, *taste* and *smell* because the domain of perception is highly polysemous. Perception verbs are particularly interesting because they are used not only to describe how we interact with the world via our physical senses, but also in figurative ways (e.g., 'I see the end of the line' vs. 'I see what you mean'). The versatility of these polysemous verbs reveals how our perceptual experiences shape the way we construct meaning using metaphorical and metonymical mechanisms. Although the study of perception

verbs has generated extensive research, most of it has been centered on the written modality. Therefore, it seems especially relevant to focus on perception verbs when exploring how gesture contributes to the disambiguation of polysemous expressions.

Each of the papers included in this thesis explores the role of speech and gesture in distinguishing the meanings of perception verbs, with a focus on either *touch* or *taste* and *smell*. The first and third papers identify through large-scale corpus analysis the strategies speakers use in speech and gesture to differentiate between meanings of these perception verbs. The second paper, composed of three experimental studies, sheds light on how observers can disambiguate the meaning of *touch* by using gesture information. By collecting large-scale naturalistic data, this thesis examines the interplay between speech and gesture when speakers use polysemous verbs in spontaneous contexts. It also incorporates laboratory data to focus more precisely on the detailed aspects of how observers interpret the gesture patterns identified. Together, these two approaches provide new insights into the phenomenon of polysemy and the cognitive processes guiding the conceptualization of the perception verbs *touch*, *taste* and *smell*.

Since this thesis is framed within the Cognitive Linguistics approach to the study of polysemy, the remaining sections of the introduction cover the fundamental literature and ideas that guide this investigation. First, the relationship between language and the study of meaning is presented. Next, the phenomenon of polysemy is explained, followed by a detailed account of the relevance of perception verbs. The last section before the thesis structure overview is dedicated to describing the most important notions of multimodality and gesture. Each section builds on the previous one, working toward the overarching goal of understanding how a multimodal approach can contribute to the study of polysemy.

1.2. Language and Meaning

1.2.1. Cognitive Linguistics and the Study of Meaning

The field of Cognitive Linguistics emerged in the 1970s as a consequence of the reaction of multiple researchers—for example, Fillmore, Lakoff, Langacker and Talmy—against previous generativist accounts of linguistic phenomena. There was a general dissatisfaction with respect to the prominence of "notational formalism at the expense of descriptive adequacy and psychological realism" (Barcelona, 1997, p. 8). Cognitive Linguistics emerged as an alternative framework, aiming to explain linguistic phenomena by means of realistic descriptions that could also be used to establish their formal representations. Thus, Cognitive Linguistics can be defined as a research approach focused on the relationship between language and mind.

The term Cognitive Linguistics not only stands for the theoretical field but also for a set of fundamental principles concerning what language is and how humans use it. As Lakoff and Johnson (1999) stated, Cognitive Linguistics seeks to explain as much of language as possible by using cognitive science methods. This approach aims to identify comprehensive generalizations, collect the broadest range of converging evidence and adapt linguistic theory considering diverse empirical findings about the human mind. Ibarretxe-Antuñano and Valenzuela (2021) summarized the general assumptions of Cognitive Linguistics in four basic statements:

- Language is cognition. Language is regarded as both a linguistic and a psychological phenomenon, since it appears to be integrated in our cognitive capacities as another fundamental human ability.

- Language is symbolic. Linguistic forms, which do not have inherent meanings, function as clues that activate the conceptual structures in our minds (Barcelona, 1997). Thus, language is viewed as a symbolic system composed of the associations between a particular form and a specific meaning—this extends to non-verbal modalities.
- Language is motivated. The association between form and meaning is sometimes arbitrary, but it also draws in non-trivial ways on motivation-based links between form and meaning. For example, language is motivated when form-meaning associations are based on our experience and understanding of the world around us. This perspective connects with the idea that linguistic concepts are embodied (see Johnson, 1987) as they reflect our physical and sociocultural experiences.
- Language is usage-based. Language is a system that is shaped through its repeated use. This principle highlights the role of generalization, the passage from usage of specific, concrete cases to more abstract units, in establishing linguistic structure.

Given these fundamental principles, the study of meaning—of particular importance for this work—stands out as a central issue in Cognitive Linguistics. This is a transcendental topic since the boundaries of meaning permeate our embodied, culturally formed individual and shared understanding (Johnson, 1987).

1.2.2. Encyclopedic Meaning

The global term *meaning* covers an extensive range of phenomena which certainly surpasses the limits of linguistic research. This is why previous works have frequently made a distinction between lexical meaning, which could be perhaps thought as objective and explicit, and non-linguistic meaning, which refers to experience-based knowledge such as cultural background. Although this division can be useful in order to simplify the study of word meaning, it has also

been criticized for treating non-linguistic information as an additional element in the process of meaning construction.

Cognitive linguists claim that no clear distinction can be made between lexical meaning and world knowledge, since the boundaries between Semantics and Pragmatics are not clearly defined. In other words, the non-linguistic knowledge associated with a word is part of the meaning of that lexical item at every level (Alcaraz-Carrión, 2018). This holistic perspective, which considers all information—both linguistic and non-linguistic—to be pertinent to the study of meaning, is known as *encyclopedic meaning*. Cognitive Linguistics supports the encyclopedic view of meaning because there is no clear-cut distinction between experienced-based knowledge and linguistic meaning.

The notion of encyclopedic meaning is truly valuable since it involves not only the conventional meaning of a lexical item, but also other aspects that can affect its interpretation. Some of these 'environmental' factors are the context in which the word is expressed and the general knowledge shared by the speaker and the interlocutor. As an illustration of this, consider the word *fasting*; this word may be associated with three different contexts, each influencing its meaning in a unique way: 1) in a health-related context, it is seen as a healing mechanism; 2) in a social context, it can be understood as a protest strategy; and 3) in a religious context, it is considered a source of spiritual strength. One of the most influential approaches to account for this phenomenon is Frame Semantics (Fillmore, 1982), which proposes that words represent categorizations of experience that are shaped by a motivating situation grounded in prior knowledge and experience. The term *frame* refers to a "system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits" (Fillmore, 1982, p. 111). All in all, the influence of non-linguistic (cultural, pragmatic and/or contextual) factors on word meaning demonstrates the inadequacy of a straightforward one-to-one pairing of lexical meaning and linguistic form.

1.2.3. Categorization

Over the years, researchers have proposed various frameworks to explain how individuals make sense of the world through language and mental representation. The ways in which individuals structure meaning have been examined from different perspectives, particularly within linguistics, psychology and cognitive science. In linguistics, one of the most extensively studied cognitive abilities that helps humans organize meaning is *categorization*. This ability allows us to group conceptually similar objects or ideas, thereby facilitating communication and comprehension processes.

The approaches that examine the mechanisms underlying categorization vary considerably, ranging from discrete and universal models, such as *componential analysis* (Katz & Fodor, 1963; Katz, 1972), to more flexible and experience-based models, such as *prototype categorization* (Rosch, 1973). Componential analysis, rooted in structural semantics, assumes that word meanings can be broken down into features that are universally applicable across languages. This model has however been challenged by several factors; for example, meaning is context-sensitive and, accordingly, not always composed of invariant features. In contrast, the prototype categorization model provides a more suitable framework, as it emphasizes that category membership is graded rather than absolute, thus provides a better account of the variability observed in real-world language use and conceptual thought.

The classical understanding of the categorization process, known as componential analysis, considers that each category consists of a group of elements which share certain necessary conditions, referred to as *semantic markers* or *binary features*. For example, the term *bachelor* is used to identify an unmarried man (i.e., bachelor = +MALE, +ADULT, +UNMARRIED). From this viewpoint, the boundaries of each category are clearly established thanks to the general definition generated through the combination of these required conditions. Therefore,

the different elements of each category are linked to the same (core) meaning. However, this classical definition of meaning is problematic as it might incorrectly include unlikely candidates or exclude likely ones (Sweetser, 1986). For instance, a *priest* could be categorized as a *bachelor*, since he meets the binary features associated to the term (male, adult and unmarried), yet this categorization would be misleading. While such problematic cases—such as the term *priest*—might be unconsciously treated as 'exceptions,' a major limitation of this model lies in its inefficiency regarding context-dependent or new entities. For example, the category *game* lacks a fixed set of binary features: while some games may involve elements such as cards or dice (e.g., poker), others do not require any (e.g., hide-and-seek). This makes it difficult to define the precise features that determine the members of this category.

In contrast to the idea of a well-defined, core meaning, the cognitive approach adopted a different categorization model. The first author to provide an alternative for the problems arising from the previous model was Eleanor Rosch (1973). Her approach, known as the prototype categorization model, suggests that human categories consist of two types of elements: a *prototype* and *less-central members*. The former is believed to be the most typical member of a category, while the latter are considered relatively similar to the prototype. In this model, the non-prototypical members of a category are organized radially around the prototype, located at the center, and their distance from it depends on their degree of similarity (see Figure 1). Rather than boundaries determined by binary features, the conceptualizer places each member in the category based on its resemblance to the prototype: the more similar a member is to the prototype, the more central it will be within the category. In the visual representation of the category *dog* shown in Figure 1, the prototype is a German Shepherd; the less-central members are a Beagle and a Dalmatian; the peripheral member is a Dachshund; and the non-member is a wolf.

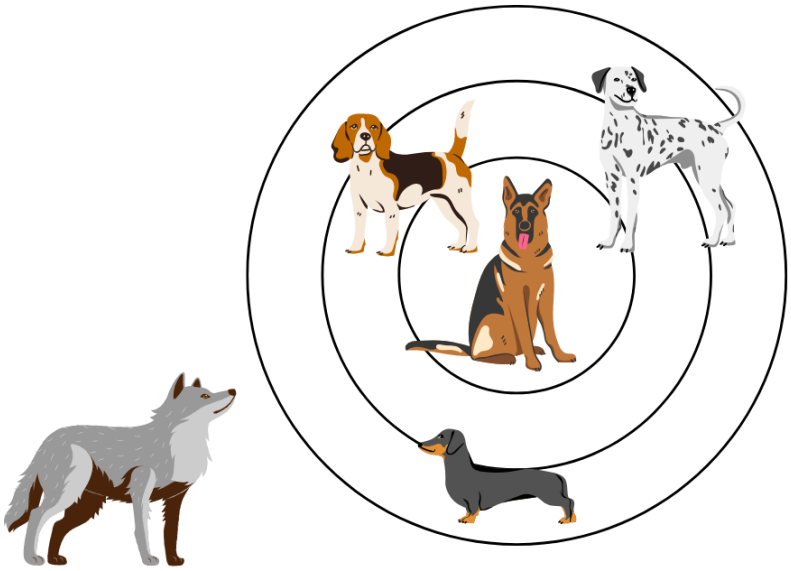

Figure 1. Category *dog* following the prototype categorization model

When analyzing the meaning of lexical items, the less-central members correspond to the "motivated, non-prototypical senses, related to the prototypical sense [i.e., prototype] in a systematic way" (Ibarretxe-Antuñano, 1999, p. 20). Therefore, prototype theory can be applied to the study of word meaning and, in particular, to the complex phenomenon of polysemy (i.e., multiplicity of related meanings) which can be understood as "a special case of prototype-based categorization, where the senses of the word are the members of a category" (Lakoff, 1987, p. 378). Expressed differently, the most frequent senses of a lexical item are conceptualized in the central area of the category (e.g., *slow* meaning 'without much speed'), whereas rarer semantic extensions gradually disperse toward the periphery (e.g., *slow* meaning 'not very clever,' *slow* meaning 'not very exciting'). The prototype model of categorization thus provides a valuable framework for studying meaning, as it not only accounts for psychological variation across contexts of use but also considers sociocultural variation among individuals—acknowledging that not every category is universally shared by all the speakers of a language (Geeraerts, 2009). In short, this categorization model, based on perceived similarity, reflects our capacity to draw connections, make generalizations and apply prior knowledge to novel situations—functions that are fundamental to both linguistic and non-linguistic processing.

1.2.4. Embodiment

Another key characteristic of the cognitive approach to the study of meaning is the notion of *embodiment*, which posits that human cognition is based on our bodily and perceptual interactions with the world. Differing with the traditional view of reason—objectivism—as being abstract and disembodied, the cognitive approach—experientialism—assumes that thought has a bodily basis.

Embodiment is considered intrinsic to our conceptual system and constitutes one of the fundamental tenets of Cognitive Linguistics (see Johnson 1987; Lakoff 1987; Lakoff & Johnson 1980, 1999). In the embodied cognition approach, mind and body form an integrated system, in which reason grows out of our bodily capacities. Perceptual and motor systems are crucial in shaping color concepts, basic-level concepts, spatial-relations concepts and aspectual (event-structuring) concepts. A clear example is the fact that the sky is blue "because the atmosphere transmits only a certain range of wavelengths of incoming light from the sun, and of the wavelengths it does transmit, it scatters some more than others" (Lakoff & Johnson, 1999); thus, color concepts arise from the interactions of the reflective properties of objects, electromagnetic radiation and our bodies and brains.

Cognitive Linguistics holds that our embodied experiences are extended from concrete cases to construct abstract ideas through a process known as *conceptual metaphor* (this idea is further developed in Section 1.2.5). Furthermore, the notion of embodiment is crucial to the study of meaning, since sensorimotor, social and cultural experiences can be seen to underlie the conceptualization of even the most abstract concepts. For example, in the sentence 'the political campaign kicks off today,' the beginning of the event is conceptualized as a sensorimotor action. Considering that embodiment produces non-arbitrary associations between cognition

and experience (Lakoff, 1987), studying meaning issues from an embodied perspective can be particularly relevant to examine the motivation behind semantic extensions.

Some of our recurring embodied experiences are more directly mapped in our minds as preconceptual abstract structures, which are known as *image schemata* (in singular, *image schema*). Johnson (1987, p. 29) defines an image schema as a recurrent pattern, shape or regularity in our actions which emerges as a meaningful structure for us mainly "at the level of our bodily movements through space, our manipulation of objects, and our perceptual interactions." Among the most common image schemata is the containment schema (see Figure 2), which reflects the many embodied experiences in which there is a space with an 'inside' separated from an 'outside.' For instance, people go in and out of containers every day—such as rooms, houses or vehicles (e.g., 'my cousins swim in the pool at their vacation home')—and even the human body can be construed as a container (e.g., 'I am full of energy once I drink my morning coffee').

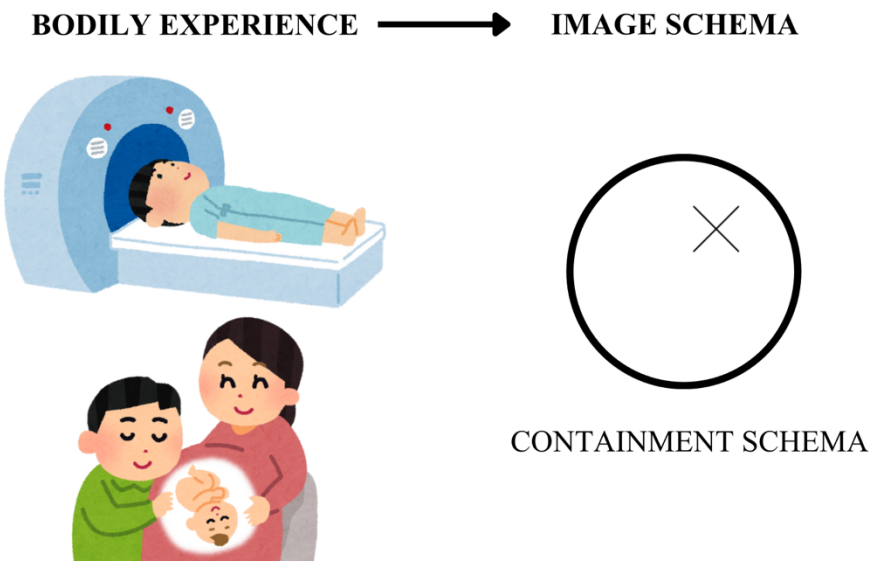

Figure 2. Representation of the bodily basis of the containment schema

Johnson (1987) further states that image schemata are dynamic structures that organize our experience and comprehension, such as in the case of polysemous expressions. In fact, he argues that polysemy provides some of the strongest evidence for the existence of image

schemata and their metaphorical or metonymical elaborations, since the related meanings of a polysemous term often share either an "underlying image schema, some extension of that schema, or some metaphorical projection of that schema" (Johnson, 1987, p. 107).

1.2.5. Conceptual Metonymy and Metaphor

The human ability to capture basic patterns common across diverse experiences also underlies our capacity to conceptualize one entity in terms of another. This can be achieved through two basic types of cognitive mappings: *metonymy* and *metaphor*. While metonymy involves entities that share conceptual contiguity within the same domain (UN peacekeepers are referred to as 'blue helmets' because of their uniform), metaphor establishes a cross-domain mapping between the entities (the phrase 'cold reception' conceives AFFECTION as WARMTH).

Metonymy uses one 'cognitively salient' entity—*vehicle* or *reference point*—to provide mental access to another entity—*target*—belonging to the same domain (Radden & Kövecses, 1999). In simple terms, metonymy occurs when a well-understood or easy-to-perceive aspect of an entity is used to stand either for the whole category or part of it (Lakoff, 1987). There are different types of metonymical mappings, depending on the relationship between the vehicle and the target: PART-FOR-WHOLE (e.g., 'Lisa gave me a hand'), WHOLE-FOR-PART (e.g., 'the Government pardoned the prisoner') and PART-FOR-PART (e.g., 'Martha did not pick up the phone'). Metonymy can also be categorized based on the specific relationship between the vehicle and the target. For example, the CONTAINER FOR CONTENT alternation underlies sentences such as 'John downed the glass in one gulp' where *glass* stands in for the beverage.

Metaphor is another cornerstone of the cognitive approach to the study of meaning. According to Conceptual Metaphor Theory (henceforth CMT), devised by George Lakoff and Mark Johnson (1980), metaphors help us understand and talk about one experiential domain (i.e., target domain) in terms of another (i.e., source domain). As a general rule, target domains are

abstract or less clearly delineated than source domains, which encompass concrete concepts. These mappings between different domains are central to our cognitive capacity of producing, transferring and processing meaning (Fauconnier, 1997). This is evident in the way speakers often conceptualize 'non-physical' ideas using 'physical' terms; for example, 'it makes my blood boil' expresses the emotion of anger. In short, CMT claims that human understanding arises not from individual words or expressions, but from broader experiential domains.

An important argument in CMT is that the mapping between source and target domains is selective, allowing multiple mappings to coexist within the same domain. This means that only specific aspects of a target domain are highlighted in a conceptual metaphor, while the remaining aspects are hidden. Although different metaphors may share the same target domain, each metaphor focuses on distinct characteristics of that domain by employing different source domains (Kövecses, 2010). This can be illustrated by the metaphors used to conceptualize ANGER: ANGER IS FIRE ('her ex-boyfriend was smoldering for days'), ANGER IS A DANGEROUS ANIMAL ('he tried to cage his temper') and ANGER IS A NATURAL FORCE ('they erupted like a volcano'). These metaphors indicate different aspects of ANGER: its intensity, its irrational nature and its uncontrollable character, respectively.

Certain cross-domain mappings have been considered 'universal' because they appear across different communities. However, the conceptual link that exists in one culture may differ significantly—or may not even exist—in another. Therefore, when examining conceptual metaphors, it is necessary to consider not only the experiential grounding but also the cultural context in which they occur. A classic example of claimed 'universality' in metaphor is the mapping between sight and intellect (UNDERSTANDING IS VISION; e.g., 'I don't see the point of enrolling in an online college'). However, many studies have shown that the dominant role of vision in cognition is not a universal, cross-cultural phenomenon (e.g., Evans & Wilkins, 2000; Majid, 2021). As Ibarretxe-Antuñano (2008) points out, while all humans use the senses to

gather and process information, the primary sense through which knowledge is accessed depends on the culture in which an individual is embedded. In the domain of PERCEPTION, she proposes framing cross-cultural relationships at a more general level—UNDERSTANDING IS PERCEPTION—rather than identifying a specific sense with a specific cognitive capability.

1.3. Polysemy

1.3.1. What is Polysemy?

Everyday language is governed by the principle of *communicative efficiency*. Levshina and Moran (2021) argue that efficiency occurs when people convey an intended message without spending unnecessary effort, thereby maximizing processing ease for the recipient. Thus, the theoretical umbrella of efficiency covers principles such as the Principle of Least Effort (Zipf, 1949), Grice's Maxim of Quantity (Grice, 1975) and the Principle of Economy (Haiman, 1983). The notion of communicative efficiency applies across all modalities of communication; however, it is especially noticeable in speech, as people generally aim to utter their ideas as briefly as possible—for example, through phonological elisions such as *wanna* instead of *want to*.

One linguistic phenomenon that derives from this natural tendency to avoid unnecessary effort is polysemy. Broadly speaking, polysemy occurs when distinct related meanings are associated to the same word or linguistic expression. Thus, rather than each meaning being linked to a distinct word form, polysemy results in a one-to-many correspondence between form and meaning.

Although this broad definition captures the essence of the phenomenon, polysemy can be more accurately and succinctly defined as the synchronic association of multiple related meanings with a single form (Sweetser, 1990). These meanings are often viewed as figurative interpretations of the prototypical sense of a polysemous lexical item; that is, they contribute

semantically to the reading of a sentence that ends up distinct from but related to its 'basic' lexical reading (Norrick, 1981). In particular, the main point of polysemy is not the multiplicity of meanings but the underlying mechanisms responsible for creating and connecting them. These non-prototypical related meanings are conceived as natural and systematic semantic extensions of the prototypical—or 'basic'—sense of a polysemous lexical item. Meaning extensions thus refer to the non-arbitrarily related senses of a word that arise through diverse cognitive strategies such as metaphorical and metonymical mappings. Overall, polysemy should be understood not only as a lexical phenomenon but also as a conceptual one, dependent on human cognitive structure.

An illustration of a polysemous lexical item is the verb *visit*. We can say: 1) 'Sam visited his parents last weekend'; 2) 'the doctor will visit physically-challenged patients' and 3) 'Emily told us to visit the museum website.' The first example reflects the prototypical meaning of the verb *visit*: 'to go to see a person and spend time with them.' This meaning specializes in the second example to indicate that the purpose of the visit is to give professional advice (in this case, medical). The last example differs from the prototypical meaning since it no longer entails physical movement but instead refers to the act of accessing or viewing a digital space.

Generally, polysemy can be subdivided into two types: regular and irregular. This distinction, introduced by Apresjan (1974), states that the polysemy of a word A is regular when the semantic relation between meanings a_1 and a_2 occurs in exactly the same way in at least one other word B of the same language, with the meanings b_1 and b_2 (e.g., several animal terms can denote either the living creature or its meat). In contrast, irregular polysemy happens when the semantic relation between meanings a_1 and a_2 is unique to a single lexical item and does not manifest in any other word of the language (e.g., senses of the verb *run* in 'run a marathon' and 'run late'). While later publications made more fine-grained distinctions within each type of polysemy (Sweetser, 1990; Pustejovsky, 1995; Taylor, 2003; Dölling, 2020), these formal

definitions led to the broad view that metonymically motivated polysemy is typically regular and metaphorically motivated polysemy is irregular. However, this widely accepted distinction is not entirely clear-cut. It has been argued that genuine regularity in polysemy should involve more than one corresponding alternative within the language (Falkum, 2015; Vicente & Falkum, 2017; Ortega-Andrés & Vicente, 2019).

Since polysemy may appear to overlap with other sources of ambiguity, it is usually explained in contrast to a similar but different linguistic phenomenon known as *homonymy*. Homonymy occurs when a form has two or more unrelated meanings, as in the case of *bank*, which may refer to a financial institution or to the land alongside a river. Occasionally, etymological information can help determine whether a lexical item is polysemous, provided its meanings share the same origin; however, this criterion is not decisive (Ibarretxe-Antuñano, 1999). Although distinguishing between polysemy and homonymy is sometimes problematic, the distinction becomes evident in online processing studies. As highlighted by Klepousniotou (2007), several studies have found differences in the way polysemy and homonymy are processed. These findings support the theoretical distinction between both types of lexical ambiguity, despite the existence of challenging positions (cf. Klein & Murphy, 2001, 2002). In summary, the main distinction between polysemy and homonymy lies in the presence or absence of a systematic semantic relationship between the word meanings. If the multiple readings of a word do not exhibit a regular semantic relation, that word is not polysemous, but a homonym instead (Norrick, 1981).

1.3.2. Main Approaches to Polysemy

Considering the inherent complexity of polysemy, several perspectives can be adopted to study it (see Falkum & Vicente, 2015, for a detailed account of the variety of proposals in interdisciplinary research on polysemy). Among the diverse theoretical proposals on the origin

and representation of polysemy, three main perspectives can be highlighted: 1) *the sense enumeration view*, 2) *the one representation view* and 3) *the sense networks view*.

The sense enumeration view claims that all the meanings of a polysemous expression are represented in our mental lexicon and, more importantly, that there is a separate representation for each of the senses of a polysemous word (Falkum & Vicente, 2015). From this perspective, polysemy and homonymy are largely treated in the same way. Although this model has been supported across several research areas, the sense enumeration hypothesis appears to offer only a limited and insufficient conception of language. In fact, this approach is thought to be "undermined by the pervasiveness of the idiomatic in language," since determining the number of meanings of a lexical item—and what these might be—fails to align with our creative understanding of language (Taylor, 2006, p. 51).

The one representation hypothesis provides an alternative to the sense enumeration view. This perspective asserts that the senses of a polysemous lexical item either belong to a single representation or depend on it. Given how problematic it seems to acknowledge that the multiple senses of a polysemous word are stored as part of a single representation, most researchers adopt a more moderate stance. They propose that the senses of a polysemous expression depend on a single representation but do not belong to it. Thus, when speakers interpret a polysemous word, they "access a semantic representation which acts as a gateway to the different senses" (Falkum & Vicente, 2015, p. 5). This is the underlying assumption behind one of the most influential approaches to the study of polysemy: Pustejovsky's Generative Lexicon (1991, 1995).

Pustejovsky's decompositional account addresses the multiplicity of word meaning, deals with lexical ambiguity and aims to determine how we can give an infinite number of senses to words using finite means (Ibarretxe-Antuñano, 1999). He proposed that related meanings can be

decomposed into semantic *primitives* (i.e., elementary pieces into which a lexical item can be broken down) as the source of meaning inferences. In other words, the semantic extensions of a polysemous lexical item are seen as manifestations of the same primitive (i.e., core meaning) occurring in different contexts. Even though this approach addresses more word meaning issues than sense enumeration models, it has also been criticized in relation to the creative use of words, the permeability of word senses and the expression of multiple syntactic forms (Ibarretxe-Antuñano, 1999).

The sense networks perspective rejects the decomposition of word meanings into primitives. Instead, it proposes that the meanings of polysemous words form a semantic network in which meanings are interconnected through conceptual associations or systematic extensions. As noted by Falkum and Vicente (2015), this view, suggesting that the full range of senses is stored in semantic memory, can be considered a radical version of the sense enumeration hypothesis. In light of the criticisms directed at the latter proposal, some authors opt for a less radical account. For example, Tyler and Evans' (2003) Principled Polysemy approach, which supports the idea of sense networks centered around a prototypical sense, distinguishes between senses stored in semantic memory and those pragmatically constructed. Broadly speaking, the sense network approach draws a distinction between linguistic knowledge and encyclopedic knowledge. As such, this approach is not only useful for analyzing lexical manifestations of polysemy but also conceptual polysemy.

As discussed in the general introduction, polysemy raises important questions about how meanings are structured and related in the mental lexicon. These issues become particularly evident in the highly polysemous domain of perception. The following section provides a detailed account of the embodied and semantically diverse meanings that characterize perceptual concepts.

1.4. Domain of Perception

1.4.1. Perception Conceptualization

Perceptual information is a constant in our lives, as it guides how we interact with the world. Therefore, the importance of studying perception lexicon is evident since all of our experiences are mediated through the senses. Winter (2019) takes this further by declaring that language would be useless if it did not allow us to express perceptual content. In fact, all languages have developed diverse strategies to express the ways in which we perceive our environment through seeing, hearing, feeling, tasting and smelling. In this regard, Majid and Levinson (2011) argue that languages are windows on the senses because they allow us to share our internal sensations. In English, speakers can convey a pleasant experience by saying 'it was a sweet moment,' but also a negative one by saying 'that moment left a bad taste in their mouths.' These examples reinforce the notion of embodiment, which is central to the domain of perception as well as to Cognitive Linguistics.

In the Western world, the domain of perception is generally divided into five senses: vision, hearing, touch, taste and smell. However, it should be kept in mind that the traditional five senses model is imperfect, as perception words reflect the interactions that occur between distinct sensory modalities. This is illustrated by the fact that "what we call tasting when we eat is [often] actually smelling," since, when describing food, speakers may say that something 'tastes a little foul' (Gibson, 1966, p. 136). In addition to this, previous works—such as Majid et al. (2018) and Van Putten (2020)—have demonstrated that the connections between each sensory domain and specific cognitive meanings vary across languages and cultures. Consequently, there is no perfect or universal mapping onto the folk model of the five senses. When this model is adopted—given that the same information can be perceived through different senses—perception should be conceived as a system of interconnected rather than

mutually exclusive sensory channels (Gibson, 1966). Despite these limitations, and given that English lexicalizes five senses, this investigation employs the traditional five-sense framework as a practical starting point for exploring perception verbs from a multimodal viewpoint.

1.4.2. Perception Verbs

Since we tend to consider perception as "something that humans [and other living beings] do" (Winter, 2019, p. 193), most studies have examined how different languages encode perception by focusing on the most basic perception verbs, such as *see*, *hear*, *touch*, *feel*, *taste* and *smell* (e.g., Sweetser, 1990; Ibarretxe-Antuñano, 1999; Evans & Wilkins, 2000). As previously stated, this thesis focuses on these verbs as their wide range of extended meanings across languages—not only in English—shows that the conceptual domain of perception is highly polysemous[1]. Perception may sometimes be regarded as a 'universal' polysemous domain, since certain non-prototypical meanings are spotted cross-linguistically. However, not all languages share the same extended meanings, due to differences in how the senses are conceptualized. Perception verbs—and much of the perception lexicon—can be used to convey information different from that associated with describing the act of perception (i.e., prototypical use). The non-prototypical uses of perception lexicon are grounded in our perceptual experiences. For instance, the connection between physical closeness and subjectivity explains why, in English, the vocabulary of taste, rather than hearing, is frequently used to express the abstract notion of personal likes and dislikes (e.g., 'the humor in that movie is an acquired taste').

[1] It has been argued that the different mappings structuring perception verbs should be understood as "polysemous senses of the conceptual domain of sense perception," and that these verbs are not inherently polysemous themselves, since they require additional semantic content from lexical items to trigger extended readings (Ibarretxe-Antuñano, 1999, p. 206). While this is an important point, in this thesis, both the domain of perception and perception verbs are labeled as polysemous—primarily for the sake of reading fluency.

As in the previous example, the non-prototypical meanings of perception verbs extend beyond the physical domain to include abstract domains. Therefore, while the motivations underlying their physical (or literal) extended meanings are primarily metonymical, those underlying their abstract (or figurative) extended meanings are essentially metaphorical. This can be illustrated by the verb *see*, whose prototypical meaning is 'to perceive something by using the eyes' ('I woke up early to see the sunrise'), which also has the physical extended meaning 'to meet' ('the CEO wants to see you in her office') and the abstract extended meaning 'to imagine' ('I can't see him as a nurse').

In particular, metaphorical mappings of the non-prototypical meanings of perception verbs are pervasive in English and in the Indo-European languages. This observation, explored in depth by Sweetser (1990), generated special interest in examining the systematic correspondences between perception verbs and the vocabulary of internal sensations used to describe mental and bodily experiences. Sweetser (1990) proposed the conceptual metaphor MIND AS BODY, motivated by the correlations between our external experience and internal states, as the cognitive mechanism structuring these mappings. Despite its usefulness in demonstrating that the pervasive metaphoric extensions of perception verbs across languages are not random, this model provides an incomplete account of the extended meanings of perception verbs: 1) it does not account for the physical extended meanings and 2) it does not specify which parts of the physical domain are mapped onto the abstract domain.

Property Selection processes (Ibarretxe-Antuñano, 1999) were proposed as a potential solution to the aforementioned issues. In simple terms, this alternative presents a typology of the fundamental properties that characterize each of the five senses, grounded in their physiology, rather than in language. Thus, Property Selection processes are presumably applicable to any language. The goal of Property Selection processes is to describe the bodily basis upon which the physical and abstract extended meanings of perception domains are based. Although the

properties in Ibarretxe-Antuñano (1999) apply specifically to the field of perception, different sets of properties could be defined for other semantic domains. In Property Selection processes, extended physical meanings involve a single step, whereas metaphorical extensions require a second step. The first step consists of selecting some of the prototypical properties from one domain and projecting them onto the other, resulting in the property selected physical meaning. The second step involves applying metaphorical processes "in order to convert the physical meaning to an abstract meaning, with the final result of a property selected, metaphorical abstract meaning" (Ibarretxe-Antuñano, 1999, p. 177).

An example illustrating how Sweetser's (1990) criticisms can be addressed in the Property Selection process applied to the sentence 'the raindrops touched the leaves, making them glisten.' In this sentence, the verb *touch* has the meaning 'to affect, physically,' shifting from the prototypical physical meaning of *touch* ('make contact with something by using your body') to a different physical domain. In this case (see Table 1), three properties from the source domain (prototypical meaning) are selected in the target domain (physical extended meaning): <contact $_{yes}$>, <closeness $_{yes}$> and <effects $_{yes}$>. This non-prototypical meaning selects the property <contact $_{yes}$> because the raindrops are on the leaves, implying some contact between both elements. Then, this contact implies <closeness $_{yes}$>, as both elements are close to each other. Finally, <effects $_{yes}$> is selected since the contact alters the original state of the leaves.

Property Selection process →	
Prototypical physical meanings	**Property selected physical meaning**
<contact $_{yes}$> <closeness $_{yes}$> <internal $_{no}$> <directness $_{yes}$> <detection $_{yes}$> <identification $_{yes}$> <voluntary $_{yes}$> <effects $_{yes}$> <briefness $_{yes}$>	<contact $_{yes}$> <closeness $_{yes}$> <effects $_{yes}$>

Table 1. Representation of the Property Selection process in 'to affect, physical' based on Ibarretxe-Antuñano (1999)

1.5. Multimodality and Gesture

1.5.1. The Importance of Multimodality in Language and Meaning

Human communication cannot be fully understood without considering multimodality, as semantic information often emerges from the combination of multiple communication channels. When we communicate, we express simultaneously meaning through different modalities, such as speech and gesture. Accordingly, language must be understood as a multimodal phenomenon. The interplay of different modalities enhances communicative efficiency, allowing people to convey more information in less time. For instance, contrastive stress can help speakers highlight an element of their message without adding an explicit explanation. In the sentence 'I didn't say I hate you,' stressing the negation (didn't) denies that the speaker made the statement, whereas stressing the object (you) specifies that the hatred is directed at someone else, not the interlocutor. The ability to convey different interpretations of the same sentence solely through the combination of spoken words and prosody demonstrates how multimodality contributes to efficient communication.

Another modality that can emphasize or extend the meaning of spoken words is gesture. When we gesture while speaking, we can use our hands to highlight either the same aspect of the

message conveyed in speech or a different one. For example, in Piagetian conservation tasks, which ask children whether the amount of liquid in a tall narrow glass is the same as in a short wide glass, children often focus their explanation on the height of the container while making a gesture representing the same dimension. However, they can also indicate the dimension of width through their hand gestures, adding new information to their spoken reasoning (Church & Goldin-Meadow, 1986).

Additionally, several studies have shown that gestures can reveal specific embodied structures that are part of our conceptual system but are not explicitly conveyed by means of speech. For instance, the multimodal nature of metaphor has been demonstrated in gesture studies (e.g., Cienki & Müller, 2008). Although there are different ways of viewing metaphor in gesture, studies on this topic have shown the dynamic and embodied essence of metaphoric meaning. This line of research underscores how significant metaphoric processes are for the motivation of the semantic meanings of gestures (Cienki & Müller, 2014). This can be illustrated by the gestural patterns found in studies on the conceptualization of time through space. Cienki (1998) found that English speakers usually refer to the events they are talking about in an imaginary timeline and move their hands from left to right. Furthermore, several authors have demonstrated that temporal gestures are not universal but culturally influenced since they reflect our internal understanding of the phenomenon of time (see Boroditsky, 2000; Núñez & Sweetser, 2006; Núñez et al., 2012). Thus, gesture analysis also proves essential when studying our understanding of abstract concepts. Overall, examining the relationship between speech and gesture can provide valuable insights into the nature of language and meaning.

1.5.2. A Multimodal Approach to Meaning Disambiguation

In everyday communicative exchanges, people typically determine the meaning of ambiguous items by appealing to context, which involves the spoken words but also additional elements

such as visual input and shared sociocultural background. Traditionally, research on meaning disambiguation has primarily focused on lexical items and syntactic structures that help interlocutors select the appropriate meaning. However, a problem arises when the ambiguous expression is 'unpredictable'—that is, when the spoken context is insufficient and there is limited access to additional pieces of information from the environment in which the message was produced. A classic example that illustrates this situation is the sentence 'John touched Mary' (Ibarretxe-Antuñano, 1999). If no context is available to interpret this highly ambiguous sentence, we are likely to contemplate at least two frequent senses of the polysemous verb *touch*: 1) 'to make physical contact' or 2) 'to affect someone—Mary, in this case—in an emotional manner.' On the whole, when ambiguous expressions are presented with insufficient contextual information, people cannot figure out their specific intended meaning.

Regarding this issue, some researchers have begun approaching this problem from a multimodal perspective, focusing on non-verbal features such as gesture. A recent study by Hinnell and Parrill (2020) proved, through a video-based experiment, that participants relied on gesture (when available) to understand the speaker's ambiguous statements. In particular, this investigation demonstrated that gesture is used by listeners to disambiguate references in discourse (namely, pronouns), when resolving ambiguous statements related to neutral and moral preferences. Although some previous studies (Holler & Beattie, 2003; Holle & Gunter, 2007) have examined the role of gesture in the interpretation of ambiguous constructions including homonyms[2], no additional research has explored whether the meaning of a polysemous lexical item can be disambiguated through the modality of gesture. Considering these findings, a multimodal approach to the study of ambiguity needs to be further explored.

[2] A more detailed discussion of these studies is provided in section 3.2.

For this reason, this thesis focuses on the potential role of gesture as an element that can be utilized to disambiguate meaning in polysemous expressions.

1.5.3. Gesture Definition

The aforementioned arguments motivated the multimodal approach of this thesis and its focus on the relationship between speech and gesture. This is, of course, not the first investigation incorporating gesture into the study of a linguistic phenomenon. Since the first modern gesture studies (Condon & Ogston, 1966, 1967; Efron, 1972), interest in the relevance of gesture has remained high—not only for the analysis of speech but also for understanding human cognition in general (see Kendon, 2007, for more information on the history of Gesture Studies). Although numerous scholars in this field have attempted to differentiate gestures from other types of bodily movement, determining what counts as a gesture can still be challenging.

One of the most well-known efforts to define the boundaries of what constitutes a gesture is *Kendon's continuum* (a term coined by McNeill, 1992). Based on the labels used by Kendon (1982), this continuum—later subdivided into four continua by McNeill (2000)—distinguishes among *gesticulation*, *pantomimes*, *emblems* and *sign languages* (see Table 2). According to McNeill (1992, p. 37), gestures—gesticulation in Kendon's terminology—are "idiosyncratic spontaneous movements of the hands and arms accompanying speech," which cannot be fully explained in purely kinetic terms, as their meaning can be freely designated by the speaker at the moment of speaking. Thus, they are subject not only to temporal variability but also to individual variability, since gesture production is culturally patterned. In simpler terms, Kendon (2004) defines gestures as visible actions that manifest deliberate expressiveness and are used as an utterance (or at least as part of one). Taking both definitions into account, the basic criteria for distinguishing gestures from other types of movement stipulate that they are produced during the communicative act of speaking and that they are not a functional act

performed on an object or person; rather, they deliberatively encode the expression of meaning (Goldin-Meadow, 2003).

	Gesticulation → Emblem → Pantomime → Sign Language			
Continuum 1: relationship to speech	Obligatory presence of speech	Optional presence of speech	Obligatory absence of speech	Obligatory absence of speech
	Gesticulation → Pantomime → Emblem → Sign Language			
Continuum 2: relationship to linguistic properties	Linguistic properties absent	Linguistic properties absent	Some linguistic properties present	Linguistic properties present
Continuum 3: relationship to conventions	Not conventionalized	Not conventionalized	Partly conventionalized	Fully conventionalized
Continuum 4: character of the semiosis	Global and synthetic	Global and analytic	Segmented and synthetic	Segmented and analytic

Table 2. Gesture continua as defined in McNeill (2000)

The gestures that convey meaning alongside verbal utterances are referred to as *co-speech* gestures. Co-speech gestures depend on the context provided by speech and, as a result, their content is coordinated with "semantically corresponding parts of speech" (Alcaraz-Carrión, 2018, p. 71). Consequently, co-speech gestures need to be interpreted within the framework provided by speech, since they are tightly intertwined with it in both timing and meaning (Goldin-Meadow & Wagner, 2005). The close relationship between co-speech gestures and verbal utterances makes these gestures a valuable object of inspection. Co-speech gestures can be used to emphasize the spoken message that it is correlated with or to add complementary or new information to the utterance (e.g., Casasanto & Jasmin, 2012). As Goldin-Meadow (2003) stated, the gestures we produce when we talk convey ideas that can be "related to, but are often different from, ideas expressed in the accompanying speech." In the latter case, these gestures may even contradict the linguistic information expressed (Norrick, 1981). For instance, when giving directions to a tourist, a speaker might say 'make a U-turn and then turn left' while simultaneously gesturing to the right side, creating a mismatch between speech and gesture.

Gestures can involve the movement of body parts, such as the head or face, different from our arms and hands. Thus, the way gestures are structured depends on the body parts involved. With respect to hand gestures—the main object of analysis in this thesis[3]—each gesture (or G-phrase in McNeill's, 1992, terminology) can be structured through one or more movement phases each. Several gestures (G-phrases) may cluster together within one gesture unit (G-unit). As explained in McNeill (1992), up to five movement phases can underlie a single hand gesture:

- *Preparation*. The limb moves away from its rest position to the area in gesture space where the stroke begins.
- *Pre-stroke hold*. The position and hand posture reached at the end of the preparation phase are held briefly until the stroke begins.
- *Stroke*. The meaning of the gesture is expressed in this phase, which corresponds to the peak of effort in the gesture. This is the only obligatory phase of a hand gesture; all the others are optional.
- *Post-stroke hold*. The position and hand posture reached at the end of the stroke phase are held briefly until the retraction begins.
- *Retraction*. The hand returns to a rest position.

[3] Despite the original goal of analyzing both facial and hand gestures in this thesis, only the latter were examined. This was due to challenges encountered during the manual analysis of the speaker's face, such as distinguishing deliberate gestures from incidental facial movements. Thus, facial gesture analysis had to be discarded due to the high variability of facial expressions and the impracticality of conducting a systematic and detailed analysis given the diverse nature of the visual data (TV recordings).

In the following example, three gesture phases can be distinguished (see Figure 3). In this case, the TV presenter Kit Hoover, while uttering ‘she’s got the sledgehammer’ (2017-01-04, KNBC Access Hollywood Live), iconically represents grabbing and raising an object—specifically, a sledgehammer—and then directing it downward. In particular, the gesture unit consists of a preparation, a stroke and a retraction phase. In the preparation phase, the speaker produces the words ‘she’s got the’ while raising both hands together until they reach her shoulder area, where she then changes the direction of the movement. This change of direction is continuous, as no pre-stroke hold is performed. The stroke phase, consisting of a downward movement away from the speaker’s body, begins simultaneously with the word ‘sledgehammer.’ The same alignment is observed at the end of the stroke, when the speaker pauses to allow her interlocutor to participate in the conversation. Finally, during the retraction phase, the speaker lowers her hands until they rest alongside her legs.

Figure 3. Example of preparation, stroke and retraction gesture phases

1.5.4. Gesture Categorization

Based on the definition of gesture, it can be concluded that “messages are auditory as well as visual,” since meaning can be transmitted across speech and the movement of different body parts (Holler & Levinson, 2019, p. 1). As previously stated, this situation poses a challenge: discerning irrelevant movements from content gestures, even when focusing exclusively on the most extensively studied group—hand gestures. Although diverse gesture typologies have been

proposed to describe different gesture functions, this thesis adopts a hybrid typology (see Table 3) that merges distinct labels from earlier classifications (Ekman & Friesen, 1969; Cohen, 1977; McNeill & Levy, 1982; McNeill, 1992). This decision is motivated by the spontaneous nature of co-speech gestures, which makes it difficult for researchers to apply a gesture typology in any fixed way[4] (Kendon, 2004). The multifunctionality of gestures, evident in everyday conversation, further illustrates this challenge. For instance, a speaker may direct their hand toward a ceiling fan to point at it while simultaneously moving it in a circular motion to represent one of its characteristics.

Representing speech content	Gesture type
Yes	**Emblem** (Ekman & Friesen, 1969) **Iconic** (McNeill & Levy, 1982) **Metaphoric** (McNeill & Levy, 1982) **Deictic** (McNeill, 1992)
No	**Beat** (McNeill, 1992) **Self-adaptor** (Ekman & Friesen, 1969)

Table 3. Hybrid gesture typology used in this thesis

Most gesture classification systems distinguish between the hand movements that represent some aspect of the content of speech and those that do not. Gestures conveying information related to the content of speech are commonly labeled as *representational* gestures (Cohen, 1977). Within this category, four distinct types are identified: *emblems* (Ekman & Friesen, 1969), *iconics* (McNeill & Levy, 1982; McNeill, 1992), *metaphorics* (McNeill & Levy, 1982; McNeill, 1992) and *deictics* (McNeill, 1992). Emblems are gestures characterized by a standardized association between form and meaning. A clear example is the thumbs-up gesture, which in Western cultures is used to express a positive response or signal approval—similar to saying *yes* or *okay*. Iconics or iconic gestures depict, through its form or manner of execution,

[4] Most gesture classification systems are rather rigid, as they assume that each gesture performs only one function—an assumption that is not always true. In this thesis, gestures are also assigned a single primary function for practical reasons; however, a hybrid gesture typology was developed to ensure a representative range of primary functions was captured during the coding process.

aspects of the act referred to in speech; for instance, when speakers say the phrase ‘grab a coffee,’ they may represent the action of drinking from a cup. Metaphorics or metaphoric gestures are similar to iconic gestures “in that they are pictorial, but the pictorial content presents an abstract idea” (McNeill, 1992, p. 14). If someone says ‘my perspective completely shifted’ while rotating their hand, the gesture metaphorically represents a change in viewpoint, even though it does not imply the literal turning of an object. Deictics or deictic gestures are used to indicate objects and events in the concrete world, but they are also used to refer to abstract concepts when there is nothing objectively present to point at. For example, speakers may point at the floor when using the phrase ‘right here.’

Among the motions that are incidental to gesture research, the hybrid typology adopted in this thesis includes two further categories: *beats* (McNeill, 1992) and *self-adaptors* (Ekman & Friesen, 1969). Beats or beat gestures follow the rhythmic pulsation of speech, but these motorically simple movements do not depict semantic content related to the spoken message (Alibali et al., 2001). Self-adaptors are the movements in which people touch their own body or manipulate their clothing (Ekman & Friesen, 1969).

Despite the fact that the principal gesture typologies—as well as this hybrid typology—pay special attention to the function of hand gestures, exclusively limiting gesture analysis to this characteristic can be problematic. Gestures that serve the same primary function may be performed in multiple ways. For instance, the action of taking a photo looks quite different when Baby Boomers and Gen Z are asked to represent it through gesture: while Boomers tend to mime holding a camera up to their eye, Gen Z typically simulate holding a smartphone in front of them and tapping the screen. In both cases, an iconic gesture is used, but the position and posture of the hands differ.

For this reason, the studies included in this thesis focus primarily on the formal aspects of hand gestures. Following Alcaraz-Carrión (2018), a simplified notation system based on Bressem (2013) was employed, given the variability of gestures that can be found in non-experimental settings—in this case, in TV recordings (see Sections 3.1.2. and 3.3.2. for more information on the corpus dataset used in this thesis). For each gesture identified in the corpus-based studies (i.e., the first and third papers), four main formal aspects were annotated: the hand involved, the handshape, the gesture axis and the gesture direction (see Table 4). Additional formal aspects, such as the gesture location, which were included in specific analyses, are described in detail in their corresponding paper sections.

Formal aspect	Coding options
Hand involved	right, left, both
Handshape	palm, grabbing, pick, index finger, fist, other, combination
Gesture axis	lateral, vertical, sagittal, punctual, other, combination
Gesture direction	rightwards, leftwards, in, towards, up, down, other, combination

Table 4. Coding scheme for the main formal aspects of hand gestures analyzed in this thesis

1.6. Thesis Overview

Following the first section presenting the theoretical framework, this thesis is organized into three additional main sections. The second section outlines the general research objectives that guided this investigation and specifies the central questions addressed in the different empirical studies. The third section includes the empirical contributions that constitute this thesis, presented in the form of three articles. The first paper—A multimodal approach to polysemy: The senses of touch—narrates the motivations underlying the origin of this investigation, which focused solely on the perception verb *touch*. The second paper—Disambiguating polysemy: How hand gestures help observers interpret the verb touch—contains three distinct studies, also focused on the verb *touch*, which were inspired by the findings of the first paper. The third paper—A multimodal approach to polysemy: The senses of taste and smell—closes the set of empirical contributions by examining whether the findings from the first article generalize to

additional perception verbs. The final section summarizes the main conclusions, discusses the broader implications of the thesis and outlines limitations and directions for future research. Overall, this structure reflects the progression from the objectives of the investigation, through the empirical research, to a synthesis of the conclusions reached.

2. Research Objectives

2.1. General Objectives

This investigation explores the multimodal characterization of the semantic extensions of a group of polysemous words. In particular, the multimodal approach adopted to address this research gap focuses on the expression of the perception verbs *touch*, *taste* and *smell* by analyzing (1) the linguistic markers surrounding these polysemous verbs and (2) the hand gestures coordinated with the speaker's utterance.

The linguistic markers examined in this research include intensifiers (i.e., verb modifiers and object quantifiers) and negation words. This decision was based on previous research (Argaman, 2009; Carrillo-de-Albornoz & Plaza, 2013), which showed that these markers are frequently associated with emotional attributes in language. These findings regarding word choice could potentially help identify the semantic extensions of polysemous verbs, such as in the case of the emotional meaning of *touch* (e.g., 'the president's speech deeply touched the audience').

With respect to co-speech hand gestures, the rationale for their inclusion is that they are an integral part of communication that cannot be ignored when analyzing utterances. Moreover, they can reveal information about the speaker's dimensions of conceptualization that cannot be accessed through the study of speech production alone (Goldin-Meadow, 2003; Kendon, 2004).

Within the general objectives of this thesis, two research questions were specifically addressed:

- Do speakers distinguish between the different meanings of polysemous expressions in speech and gesture?
- Can observers disambiguate the meaning of polysemous expressions through gesture?

Answering these research questions will contribute to addressing additional key issues in the field of Cognitive Semantics, such as how is meaning created, transformed and interpreted.

Overall, this thesis provides a new perspective on the study of polysemy by leveraging the complementary layer of analysis offered by co-speech gestures.

2.2. Do Speakers Distinguish Between the Different Meanings of Polysemous Expressions in Speech and Gesture?

The first research question guides the specific objectives of the first and third papers in this thesis. Both papers present extensive corpus-based investigations into the multimodal expression of the perception verbs *touch*, *taste* and *smell*. The specific objectives aimed at determining whether speakers distinguish between the different meanings of polysemous expressions are as follows:

- Conduct a detailed investigation into the potential strategies used in speech and gesture by English speakers to distinguish between the different meanings of *touch*, *taste* and *smell* based on large samples of naturalistic data. This includes identifying and describing any patterns that accompany the expression of these three perception verbs.
- Assess possible connections between the semantic extensions conveyed through speech and the information conveyed through gesture regarding the expression of the verbs *touch*, *taste* and *smell*.
- Evaluate the presence of specific information in gesture that reveals aspects of the conceptualization of the different meanings of the three English perception verbs under analysis.
    - If this is the case, analyze whether it occurs in parallel with the speech manifestation (showing similar aspects) or differently from the spoken words (gestures depicting new aspects of the senses expressed).
- Account for how the findings of this research can be integrated within the theoretical frameworks adopted to study the phenomenon of polysemy.

2.3. Can Observers Disambiguate the Meaning of Polysemous Expressions through Gesture?

The second research question guides the specific objectives of the second paper in this thesis. This paper presents a series of three experimental studies on the disambiguation of the physical and emotional meanings of the perception verb *touch*. Study 1 examines whether English speakers can accurately guess the meaning of *touch* based solely on the hand gestures made by other speakers. Study 2 investigates whether specific gesture features—gesture location and handshape—are associated with the physical and emotional meanings of *touch*. Study 3 measures whether observers decide the meaning of *touch* faster when a co-speech hand gesture is available, and whether reaction time varies depending on the specific gesture observed. The common and specific objectives aimed at determining whether interlocutors disambiguate the different meanings of polysemous expressions are as follows:

- Design a paradigm for differentiating the central visual component in a multimodal analysis. This involves isolating facial expressions from hand/body gestures and analyzing which body part has a greater effect on the participant's response. (Study 1)
- Assess possible connections between the formal gesture features and the meaning of *touch* interpreted. (Study 2 and Study 3)
- Evaluate the effect of co-speech gestures in reaction time when interpreting the meaning *touch*.
    - If gesture facilitates faster decision-making, assess whether this effect persists when gestures are isolated from the speaker's speech. (Study 3)
- Account for how the findings of this research can be integrated within the theoretical frameworks adopted to study both the disambiguation of meaning and the phenomenon of polysemy. (Study 1, Study 2 and Study 3)

3. Research Papers

As this thesis is presented in the form of a compendium of publications, this section comprises the three individual research papers derived from the work conducted during this doctoral project. All three papers included in this section have been peer-reviewed and accepted for publication. The articles were coauthored by Irene Bolumar Martínez (University of Murcia), Daniel Alcaraz-Carrión (University of Murcia) and Javier Valenzuela Manzanares (University of Murcia). The second paper additionally includes Susan Goldin-Meadow (The University of Chicago) as a coauthor. With respect to author contributions, Irene Bolumar Martínez was responsible for materials preparation, data collection, data analysis and writing the first draft of each manuscript. All authors contributed to the research conception and design, provided critical feedback on multiple versions of the manuscripts and approved the final versions. All three papers are reproduced here with the permission of the respective publishers.

The articles included are as follows:

- **First paper**:

  Bolumar Martínez, I., Alcaraz Carrión, D., & Valenzuela Manzanares, J. (2024). A multimodal approach to polysemy: the senses of touch. *Language and Cognition*, *16*(4), 1697-1717. https://doi.org/10.1017/langcog.2024.23 © The Author(s), 2024. Published by Cambridge University Press. Reproduced with permission.

- **Second paper**:

  Bolumar Martínez, I., Goldin-Meadow, S., Alcaraz-Carrión, D., & Valenzuela, J. (2025). Disambiguating polysemy: How hand gestures help observers interpret the verb touch. *Review of Cognitive Linguistics*. Advance online publication. https://doi.org/10.1075/rcl.00243.bol

  From Disambiguating polysemy: How hand gestures help observers interpret the verb touch/Irene Bolumar Martínez, Susan Goldin-Meadow, Daniel Alcaraz-Carrión and Javier Valenzuela. In *Review of Cognitive Linguistics*, 13 November 2025. Reprinted

with kind permission from John Benjamins Publishing Company, Amsterdam. [www.benjamins.com]

- **Third paper**:

  Bolumar Martínez, I., Alcaraz-Carrión, D., & Valenzuela, J. (2025). A multimodal approach to polysemy: The senses of taste and smell. *Cognitive Semiotics*. Advance online publication. https://doi.org/10.1515/cogsem-2025-2005

3.1. A Multimodal Approach to Polysemy: The Senses of Touch

This study investigated whether speakers use multimodal information (speech and gesture) to differentiate the physical and emotional meanings of the polysemous verb *touch*. We analyzed 302 hand gestures that co-occurred with this perception verb. For each case, we annotated (1) the meaning of *touch* (physical vs. emotional), (2) the gesture referent[5] speakers physically touched (*other-touch* vs. *self-touch*), (3) the personal pronoun following the verb and (4) if they used intensifiers and negation. There were three main findings. First, we have seen that when speakers express the physical meaning, they are likely to reach an external referent (other-touch), but when they imply the emotional meaning, they tend to touch their own body (self-touch). Second, the most frequent co-speech gesture (*chest-touching* gesture) was associated with the emotional meaning, uncovering the metaphor THE HEART IS A CONTAINER FOR EMOTIONS. Third, this study showed that the physical meaning of *touch* usually coexists with a wide variety of personal pronouns and negation words; in contrast, the emotional meaning of *touch* occurs primarily with the pronoun *me* and it is usually modified by intensifiers. Thus,

[5] Although the term gesture referent is used here to preserve the original content of the published article (Bolumar Martínez et al., 2024), it will be referred to as gesture location in subsequent papers. This refinement in terminology aims to better capture the spatial aspect of the gesture and to avoid potential confusion with the entity involved in the gesture.

speakers use both speech and gesture to differentiate the meanings of the polysemous verb *touch*.

Keywords: gesture; multimodality; perception; polysemy; touch

3.1.1. Introduction

Polysemy can be succinctly defined as the synchronic association of multiple related meanings with a single form (Sweetser, 1990). Polysemous words typically have a prototypical meaning (i.e., their most frequent semantic interpretation) to which other senses are related in a motivated way. The complex phenomenon of polysemy has been extensively researched from a variety of disciplines such as lexicography, cognitive semantics and psychology, among others (see Falkum & Vicente, 2015, for an overview of the main theories regarding polysemy). Regardless of the research focus, most authors agree that the fundamental characteristic of polysemous constructions is the natural and systematic relation of its multiple meanings. In other words, the main point of polysemy is not the multiplicity of meanings but the underlying mechanisms in charge of creating and connecting its different senses.

If the multiple readings of a particular lexical item exhibit no systematic semantic relation, that word is considered homonymous rather than polysemous. Thus, the main difference between homonymy and polysemy is the absence or presence of a systematic relation between meanings. For example, even though the word *ball* can be used to designate at least two different meanings, (1) ‘a round object used in games’ and (2) ‘a formal dance,’ it must be considered homonymous since these meanings are unrelated, that is, they show no type of connection, neither semantic nor etymological (i.e., each meaning originated from different word forms). In contrast, an example of a polysemous lexical item is the noun *taste*. It is mostly used with adjectives to describe the flavor of a food product (e.g., ‘it is a basic cake, but it has a good taste’). However, it can also be used to refer to someone’s discernment (e.g., ‘he has

bad taste in clothing'). These senses are naturally related as they both denote likes and dislikes. In this case, the latter sense reflects the metaphorical mapping of an already existing meaning from the domain of PERCEPTION to the domain of PERSONAL JUDGMENT.

One of the theories which has examined more closely the potential motivations behind the semantic extensions of polysemous lexical items is cognitive semantics. This discipline has paid a great deal of attention to the role of metaphor as a connecting mechanism between senses. Many of the senses of a polysemous word, as is the case of the discernment sense of *taste*, are often considered figurative interpretations of the 'prototypical' sense of a lexical item. These 'non-prototypical' related figurative meanings (e.g., metaphorical, metonymic, pejorative) are generally conceived as semantic extensions of the polysemous lexical item.

Thus, metaphor is one of the key notions behind the motivation of semantic extensions. For a substantial number of scholars in the field, metaphor is the mechanism that allows us to ground abstract concepts on embodied experience. This means that the way we perceive our surroundings shapes our conceptual system (Lakoff & Johnson, 1999). In many cases, there is a bodily based motivation behind the senses of a polysemous item which structures its meanings, and the most frequent structuring mechanisms are metaphor and metonymy (Ibarretxe-Antuñano, 1999). An illustration of this would be the sentence 'I see your point,' where we use the verb *see* metaphorically to convey comprehension instead of perception (UNDERSTANDING IS SEEING). This case appears to be motivated by the fact that most cultures conceive sight as the most reliable source of information that eventually leads us to the understanding of events. Therefore, this example shows how metaphor produces nonarbitrary associations between embodied experience and semantic extensions (Lakoff, 1987).

One of the areas in which the role of metaphor in semantic extension has been more thoroughly researched is the vocabulary of perception. The main reason for this is that, even though the

vocabulary of perception is used to describe how we perceive the world through our senses, it is extremely common to find perception words used metaphorically. In fact, Sweetser (1990), who pioneered research on perception verbs by diachronically exploring their multiple meanings in English, concluded that the systematic connections between our vocabularies of mind and body are essentially metaphorical in nature. For example, when we say 'I smell trouble' there is no physical act of perception; instead, we use this verb to express our intuition that something bad is going to happen. Multiple works have focused on the analysis of perception words (especially, verbs) and their semantic extensions in many languages such as English (Kövecses, 2019; Selmistraitis & Boikova, 2020; Sweetser, 1990; Winter, 2019), Spanish (Ibarretxe-Antuñano, 1999; Jansegers et al., 2015), Italian (Monachini & Roventini, 1994) and other languages (Burenhult & Majid, 2011; Evans & Wilkins, 2000; Krishna et al., 2022; Majid & Levinson, 2011; van Putten, 2020), amply showing the highly polysemous nature of this conceptual domain.

One of the fundamental works on the polysemy of perception verbs is the cross-linguistic analysis carried out by Ibarretxe-Antuñano (1999, 2002, 2006, 2019). In her work, she provides a cognitive semantic account of polysemy in English, Basque and Spanish perception verbs. The author describes the extended meanings that each of the five senses (i.e., vision, audition, touching, tasting and smelling) has in these three languages and summarizes the rich structure of the perception metaphors underlying them. For example, she explains that, besides its basic, physical sense of coming into contact with something, *touch* can be understood as 'to affect' both physically ('my boots were here, who touched them?') and metaphorically ('her story touched me deeply'). Whereas the first sentence could be rephrased as 'who changed the position of my boots?,' in the second one the verb *touch* specifies that what is affected is the emotional side of the speaker. Moreover, Ibarretxe-Antuñano (1999, 2006) shows how polysemy can sometimes be activated by the combination of a lexical item with other elements

of the sentence which are necessary to build up that meaning. In a sentence such as 'John hardly touched the food' (Ibarretxe-Antuñano, 1999, p. 69), the property of briefness is being highlighted by the adverb *hardly* (which here refers to a 'small degree of something'). Without this adverb, however, the extended meaning of 'partake' disappears completely and the resulting sentence ('John touched the food') comes to trigger the physical meaning of *touch*. In spite of this, we may also encounter utterances, such as 'X touched Y,' without any additional linguistic elements activating the correct meaning of the polysemous lexical item. When the competition between the multiple meanings of a lexical item cannot be resolved, people cannot determine which of the distinct potential representations of that word should be activated. Polysemous words may lead to this situation which is known as language ambiguity. Thus, our ability to process ambiguous constructions is also of relevance when studying polysemy.

Taking this into consideration, language ambiguity has been of great interest for polysemy studies since the phenomenon of polysemy is believed to be one of the major sources of ambiguity (Nerlich & Clarke, 2003; Norrick, 1981; Taylor, 2006). The issue of ambiguity often emerges because the different interpretations of polysemous words are manifestations of the same core meaning; however, in most polysemous expressions there is one sense that seems more appropriate in the given context (Klepousniotou, 2007). We tend to consider that a polysemous word is ambiguous when we find it hard to choose between two or more senses that can be potentially associated with that specific form. Although people normally determine the meaning of ambiguous items by appealing to context (e.g., previous speakers' utterances, common sociocultural background, visual input), it is not always possible to access those additional pieces of information. A classic example that illustrates this situation is the sentence 'John touched Mary'—'X touched Y'—(Ibarretxe-Antuñano, 1999, p. 124). With no additional context, at least two of the most frequent senses of the verb *touch* could be considered: (1) 'establishing physical contact,' or (2) 'affecting someone in an emotional manner.' In short,

when there is insufficient contextual information when uttering polysemous expressions, listeners cannot decide their most suitable meaning interpretation (Ibarretxe-Antuñano, 1999).

On the whole, previous works have mostly examined the phenomenon of polysemy from a linguistic perspective (Falkum & Vicente, 2015). However, the study of polysemy including 'non-verbal' communication modalities has not been sufficiently explored (Ibarretxe-Antuñano, 2019). Considering that meaning can be conveyed through multiple modalities such as gesture or voice pitch, it makes sense to extend the analysis of polysemy from the purely verbal level to a multimodal perspective. In contrast to the traditional pairings of specific word forms and discrete meanings, the notion of multimodality asserts that semantic distinctions generally arise from the interplay of multiple communication modalities which can be of a different nature (i.e., 'verbal' and 'non-verbal'). Hence, introducing the factor of multimodality to study the polysemy of perception verbs could be a way of expanding our previous knowledge and improving our existing models of polysemy.

One of the areas in the study of polysemy that multimodality could contribute to is the disambiguation of meaning. As mentioned before, this area has mainly been addressed by looking at the linguistic elements accompanying polysemous constructions (Hirst, 1987). For this reason, redirecting the research focus toward distinct communication modalities and methods to study language ambiguity could prove to be especially useful, since it adds another layer of analysis to complement verbal information. So far, how (or indeed, whether) people disambiguate the meanings of polysemous words in spontaneous communication using multimodal cues has received little attention. There are, nonetheless, some works that have attempted to combine distinct modalities when studying language ambiguity. Holler and Beattie (2003) report that speakers do use gesture to clarify verbal ambiguity for the listener. They found that when speakers are asked to clarify the meaning of an ambiguous word used in the course of an extended discourse (in their work, homonyms), they are more likely to use a

gesture in association with the problematic word; similar results were found by Holle and Gunter (2007). Similarly, when investigating the resolution of ambiguous statements related to neutral and moral preferences, Hinnell and Parrill (2020) showed that participants rely on gesture, when available, to understand the speaker's ambiguous statements. Particularly, gesture appeared to be used by observers to understand the speaker's opinion, since they were more likely to choose the idea accompanied by a gesture as the speaker's preference, regardless of whether the linguistic cue expressed a neutral idea or a moral view that might be considered socially unacceptable. On the whole, these three studies sustain that, apart from listening to the oral message, interlocutors tend to look for meaning in the hand movements that accompany ambiguous speech (Maricchiolo et al., 2014).

Therefore, it appears that focusing on gesture together with speech could be beneficial for the analysis of polysemy since these modalities are semantically co-expressive and may transmit a different aspect of the same idea (Goldin-Meadow, 2003; McNeill, 2013). This means that gesture can express information either on its own or when co-occurring with oral production. Co-speech gestures (i.e., gestures co-occurring with speech) may express task-relevant information not conveyed in speech; for example, sometimes children make a gesture depicting width while saying 'this cup is bigger' (Goldin-Meadow & Alibali, 2013). Co-speech gestures have also become a key aspect in language, cognition and communication studies because they can convey both concrete and abstract ideas. We can point to the restroom while saying 'Ana is in the toilet,' but we can also draw circles in the air while saying 'the song is on repeat.' Besides, co-speech gestures may be especially useful when dealing with words that have literal and figurative meanings, such as polysemous words, since they can reveal some of our covert cognitive representations (cf. Alcaraz-Carrión & Valenzuela, 2021; de la Fuente et al., 2014; Núñez & Sweetser, 2006). This is mainly because while gestures convey meaning "globally, relying on visual and mimetic imagery", speech does it "discretely, relying on codified words

and grammatical devices" (Goldin-Meadow, 2003, p. 25). If someone said 'she is wearing a hat' while moving his hand as though grasping the bill of a cap, this gesture might encourage listeners to infer that the hat referred to was particularly a baseball cap (Goldin-Meadow & Alibali, 2013). Therefore, in contrast to the traditional verbal approach, a multimodal approach focused on gesture could potentially extend our understanding of the mental processes behind polysemy.

Taking into account the aforementioned examples of polysemous perception verbs, the verb *touch* seems to be especially interesting, given the fact that it is the only perception verb closely tied to the domain of emotion (see Sweetser, 1990, for a more detailed explanation). In fact, several works (Ibarretxe-Antuñano, 2006; Jansegers et al., 2015; Lacey et al., 2012) have already explored how the semantic content of tactile expressions and its arguments contribute to the creation of meanings such as 'to affect,' 'to deal with,' and so on.

While previous studies aimed at introducing different modalities (through videos and images) in language disambiguation research, no work has paid attention to the role of specific gestures in the expression of polysemous constructions in spontaneous communication. Thus, the present article explores to which extent co-speech gestures can be used to disambiguate the meaning of polysemous words. Our hypothesis, thus, is that gesture, together with linguistic cues, could also play a relevant part in the phenomenon of polysemy.

This research adopts a multimodal perspective to explore two specific senses of the perception verb *touch*: its physical meaning and its emotional meaning. We address the following research question: do speakers use multimodal information (namely, speech and gesture) to differentiate the physical and emotional meanings of *touch*? With a view to discover whether gestures are used to disambiguate polysemous words, three specific objectives emerge: (1) to compare the co-speech gestures associated with each selected meaning of *touch*, (2) to determine the

motivation behind potential frequent co-speech gestures associated with *touch* and (3) to describe the relation between the linguistic context and the physical and emotional meanings of *touch*. In this investigation, we evidence that speakers can manifest different semantic extensions of the same lexical item through verbal (speech) and non-verbal (gesture) modalities. Based on these findings, we conclude that both speech and gesture are indeed used to distinguish the meanings of polysemous words.

3.1.2. Methodology

3.1.2.1. Dataset and Tools

The audiovisual and textual information was obtained from the NewsScape Library of TV News, a multimodal repository of television news hosted in the UCLA and CWRU libraries (which gathers data from March 2004 until today). This dataset is managed by The International Distributed Little Red Hen Lab™ which is a big data science consortium for research into multimodal communication. It contains more than 500,000 hours of television news programs as well as a 4-billion-word multi-lingual dataset formed by television subtitles of TV news. Speech and subtitles are synchronized through force-alignment tools, which enables researchers to access the exact moment in which specific linguistic expressions were uttered. The communicative exchanges found in NewsScape are set in different communicative situations like interviews, open discussions or debates.

We also employed the corpus software CQPWeb (Hardie, 2012) with a subset of the NewsScape dataset (Uhrig, 2018) in order to perform more accurate linguistic searches. Specifically, we employed the NewsScape 2016 sub-corpus (269,269,133 words) which gathers data and recordings from the year 2016.

3.1.2.2. Linguistic Search

We collected all the cases in which the verb *touch* was followed by a personal pronoun or by the words *someone* or *somebody* in NewsScape 2016. In addition, we complemented this search in the 2016 sub-corpus with a search for the specific combination *touched me* in the whole NewsScape Library dataset (time span: March 1, 2004, to May 1, 2021). This second search was made to balance the number of emotional and physical senses and to obtain a representative sample of gestures used in an emotional context. The expression searched was selected from a preliminary analysis which revealed that *touched me* seemed to be the most frequently associated construction with the emotional sense of *touch*.

3.1.2.3. Analysis

3.1.2.3.1. Data Filtering

Considering that the aforementioned corpora are built with TV recordings, the textual and visual data stored in them tend to be quite 'noisy' (e.g., hits may also include voice-over cases). After the searches were done, it was necessary to discard those cases which were not suitable for our goal. So, data were manually filtered in different steps to obtain an adequate final sample. First, we removed all the cases that presented a technical issue (e.g., audio/video not available) or presented a text-speech mismatch. We also removed tagging errors (i.e., when parts of speech were not correctly identified by the system). Second, we annotated whether the query item had a physical, an emotional or an ambiguous sense, as well as cases with other senses; here, only cases with physical and emotional meanings were kept. In order to check if the meaning of *touch* in the speakers' oral message had been accurately classified, a second coder analyzed a subset of the sample obtained ($N = 100$ gestures, 33.33% of the final dataset); there was almost perfect agreement ($\kappa = 0.94$) between coders in identifying the meaning of *touch* in each case (Cohen, 1960). Third, we noted down whether the hands of the speakers

were visible or not. In particular, we classified the cases into (1) barely or nonvisible gesture, (2) partly visible gesture and (3) completely visible gesture. Cases in which the speakers' hands were not visible (category 1) were discarded, as it could not be determined whether the speaker performed a hand gesture or not. Considering that partly visible gestures can be reconstructed by the viewer, no distinction is made between categories 2 and 3 in the analysis. Fourth, we removed repeated clips (i.e., when the exact same utterance appeared in a duplicate video) to obtain a list of unique cases. Lastly, we annotated whether the speakers performed a co-speech hand gesture or not. The end product of this filtering process was the creation of two lists of cases: one containing the gestures made with a physical sense of *touch* and another one containing the gestures made with an emotional sense of *touch*.

3.1.2.3.2. Gesture Analysis

For each of the cases that contained a co-speech gesture, we annotated the following features.

Firstly, we annotated whether the gesture was related to the semantics of the linguistic expression or not, as well as those gestures whose meaning was unclear. Specifically, we considered to be semantically related co-speech gestures those which were coherent with the speakers' utterance by adding a parallel or complementary nuance (McNeill, 1992). An illustration of this category would be any case in which speakers point at an element mentioned in their verbal message, such as when somebody says 'do not touch it' and simultaneously points to the object s/he is referring to. A gesture was considered unrelated when it was not directly associated with the speakers' words, not depicting any semantic information related to the speech. For example, in this category, we include cases such as beats which are gestures that move along with the rhythm of speech (McNeill, 1992) and self-adaptors which are the incidental movements individuals make when manipulating their own body or clothes, such as when someone adjusts his glasses (Ekman & Friesen, 1969). There were eight problematic co-

speech gestures that were labeled as unclear since they caused special controversy among the coders over whether they were related or unrelated cases. These cases were discarded from the dataset. A second coder annotated whether the co-speech gestures were related to the speaker's utterance or not, showing a substantial agreement ($\kappa = 0.71$; Cohen, 1960) with the first coder. Unrelated co-speech gestures were thus discarded from the final dataset. Then, seven co-speech gestures initially labeled as related were also eliminated from the final dataset since they were 'acted' cases (e.g., Frankie Valli appears on KABC Jimmy Kimmel Live and while singing the words 'touched me,' he extends his free arm forward and then changes his open palm handshape into a fist shape); in other words, these gestures were part of a performance and not produced in a natural communicative situation.

Secondly, we annotated whether speakers touched the referent of the gesture (e.g., body parts, other people). With regards to what speakers physically touched, two categories were distinguished: *other-touch* (whether speakers touched an object or another person) and *self-touch* (whether speakers touched their own arm, chest, leg or belly). A second coder analyzed a subset of the data ($N = 100$ gestures, 33.33% of the final dataset) in terms of what speakers touched, showing a substantial agreement ($\kappa = 0.73$; Cohen, 1960) with the first coder.

3.1.2.3.3. Linguistic Analysis

In relation to the linguistic elements, we examined the distribution of the pronouns following the verb *touch* and, taking into account that certain linguistic modifiers have been previously associated with emotional attributes in language, two additional verbal components were also annotated: the presence of intensifiers (i.e., linguistic markers which have "a heightening or lowering effect" on a lexical unit; Quirk et al., 1972, p. 376) and the use of negation (i.e., syntactic process through which a word or particle can change a positive construction into a negative one; Quirk et al., 1972). In particular, the reason for the annotation of intensifiers was

motivated by the fact that these linguistic markers have been shown to be very frequently associated with the expression of emotional experience (Argaman, 2009). Additionally, negation words were also taken into account because, in a Sentiment Analysis study, Carrillo-de-Albornoz and Plaza (2013) showed that negation was the most influential phenomenon when predicting the emotional meaning of the text. Since one of the two meanings of *touch* examined in this article is directly linked to the expression of emotion, both intensifiers and negation words could be fundamental elements that help identify the sense of the verb.

These characteristics were noted down when they modified the verb and were fewer than five words away from the query item (i.e., when they were included within the range –5:+5). The list of intensifiers and negation signals found can be seen in Table AI.1. The final dataset is publicly available at https://osf.io/v8gp3/.

3.1.3. Results

3.1.3.1. Gesture Frequency

The *touch + personal pronouns* search (in NewsScape 2016) returned 2571 matches (60.2%) and the *touched me* search (in the whole NewsScape dataset) gave 1698 matches (39.8%), amounting to a total of 4269 cases (100%). From the overall amount, we removed 417 cases that presented either a technical issue or a text-speech mismatch and, subsequently, we removed 59 tagging errors. When examining closely the meaning of *touch* in the remaining 3793 cases, we found 30 cases with an ambiguous sense and 214 cases with other senses (i.e., neither physical nor emotional), which were then excluded. After discarding the cases that were not associated with the physical or emotional meaning of *touch*, the remaining dataset (3549 cases) resulted in 2194 cases expressing the physical meaning of *touch* and 1355 cases expressing an emotional meaning. Then, 2562 cases (1528 physical and 1034 emotional) in which the speaker's hands were not visible were eliminated. After this step, 175 physical cases and 71

emotional cases (246 in total) were removed as they were repeated. As this work focus was on co-speech gestures, 265 cases (178 physical and 87 emotional) where speakers did not make any hand gesture were not included in our gesture sample. In order to obtain the definitive sample of co-speech gestures, 8 unclear gestures, 159 unrelated gestures and 7 acted gestures were eliminated. After that, 186 physical (4.4%) and 116 emotional (2.7%) semantically related gestures were kept in the final dataset. Hence, the total number of cases was reduced from 4269 cases to 302 co-speech gestures which amounts to 7.1% of the cases obtained from both linguistic searches (for more data curation details see Table AI.2).

Taking into account only the cases associated with the physical or emotional meaning of *touch* (see Figure 4), a chi-square test of independence was performed to examine the relation between the presence and absence of gesture and the meanings of *touch*. The relation between these variables was not significant, $\chi^2(1, N = 567) = 1.9124$, $p = .167$. Hence, there was no difference in the distribution of cases with related gestures and cases without gesture when co-occurring with the meanings of *touch*.

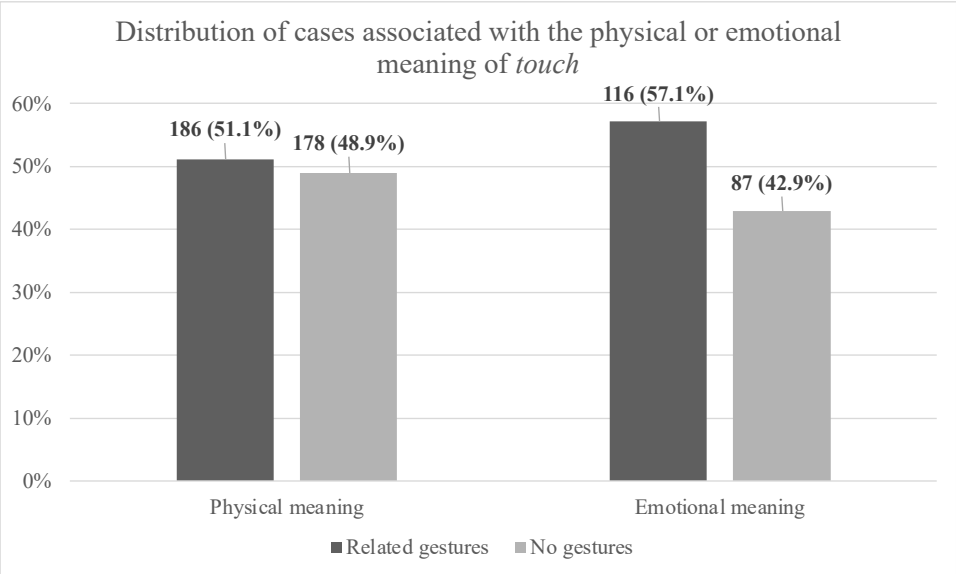

Figure 4. Distribution of cases associated with the physical or emotional meaning of *touch*

3.1.3.2. Comparing the Physical and Emotional Senses of Touch

3.1.3.2.1. Gesture Referent

In the case of co-speech gestures performed with the physical sense of *touch*, speakers touch the referent of the gesture in 21.5% of the cases ($N = 40$, out of 186). Concerning the co-speech gestures performed with the emotional sense of *touch*, speakers identify the referent by touching it 38.8% of the times ($N = 45$, out of 116). With a view to discovering whether there was a significant association in the final sample between touching the referent, or not, and the type of verb sense (i.e., physical and emotional), we calculated a chi-square test and the result, $\chi^2(1, N = 302) = 10.558$, $p = .001$, indicates a significant association between both variables. This means that speakers are more likely to touch the referent of the gesture when the sense of the verb is emotional than when it is physical.

Additionally, when speakers physically touch the referent of the gesture, two categories can be distinguished: other-touch and self-touch (see Table 5). Self-touch is found with both verb meanings (16 times, 18.8%, in physical cases and 45 times, 52.9%, in emotional cases), but the former, other-touch, can only be found when the verb *touch* implies its literal meaning (24 times, 28.2%, in physical cases). Since there were no cases in some of the cells (Table 5) we could not use the chi-square test. This is why we employed Fisher's exact test to determine if there was a significant association between the type of the meaning (i.e., physical and emotional) and the referent that was touched (i.e., other-touch and self-touch). There was a statistically significant association between the two variables ($p < .001$).

	Other-touch		Self-touch			
	Object	Other person	Arm	Chest	Leg	Belly
Physical meaning	14.1% (12/85)	14.1% (12/85)	8.2% (7/85)	5.9% (5/85)	3.5% (3/85)	1.2% (1/85)
Total physical	28.2% (24/85)		18.8% (16/85)			
Emotional meaning	- (0/85)	- (0/85)	1.2% (1/116)	48.2% (41/85)	- (0/85)	3.5% (3/85)
Total emotional	0% (0/85)		52.9% (45/85)			

Table 5. Distribution of other-touch and self-touch in the final gesture dataset

Specifically, the group of gestures associated with the physical meaning shows that speakers tend to touch an object or another person (12 times, 14.1%, in each case; see Figure 5 for an example). Referring to the group of gestures associated with the emotional meaning, speakers mostly physically touch their chests (41 times, 48.2%; see Figure 6 for an example). The data collected suggests that the latter case frequently occurs together with the emotional meaning of the *touch*. Hereafter, this specific type of co-speech gesture is referred to as *chest-touching* gesture[6].

A chi-square test was carried out to examine the relation between the distribution of gestures where speakers touch their chest and gestures where speakers touch other things and the meanings of *touch*. The relation between these variables was significant, $\chi^2(1, N = 85) = 52.7032$, $p < .001$. This result shows that speakers are more likely to physically touch their chest when the verb has an emotional meaning than when it has a physical meaning.

Concerning the cases of other-touch, they are noteworthy as they only seem to appear with the physical meaning of *touch*. The category other-touch can be easily distinguished as it involves physically touching an external element (in the following case, a different person). The other-touch example below (see Figure 5) is made while the speaker says:

1) ‘he came over and touched me’ (2010-12-07 KABC The View)

[6] Several instances of speakers pointing toward their chest were also observed. However, these gestures were not classified as chest-touching gestures and were consequently excluded from the analysis, as this paper focuses on gestures involving physical contact.

In this case, Whoopi Goldberg talks about a little boy who was behaving badly and annoying her on a train trip. The most upsetting part of that situation was that the child touched her with his sticky hands. Simultaneously with this utterance, she touches her colleague's arm with her left hand to illustrate and emphasize the fact that the naughty child dared to touch her.

Figure 5. Example 1: other-touch gesture made with a physical meaning

Regarding chest-touching gestures, the next case is a very clear example (see Figure 6):

2) 'initially it [the project] touched me' (2019-06-16 KCBS Entertainment Tonight)

In this clip, Halle Berry explains how she got involved in a documentary of patients with AIDS. When talking about the phone call she received to participate in that project, she makes reference to a personal experience in order to explain why she accepted the offer immediately. Then, while saying the sentence above, she raises her left hand from a resting position and opens the palm to touch her chest area.

Figure 6. Example 2: chest-touching gesture made with an emotional meaning

3.1.3.2.2. Linguistic Analysis

With regard to the linguistic utterances, we checked whether the meanings of *touch* could be associated with specific personal pronouns, the use of intensifiers and the use of negation words.

First, we counted the cases of each pronoun in our sample of gestures, confirming that the physical and emotional senses of *touch* might also be distinguished just by looking at the pronouns that appear after the verb (see Table 6). Taking into account only the *touch* + *personal pronouns* search in CQPweb (where pronouns were not specified), it can be seen that, when speakers express the physical meaning of *touch*, they mostly address an external referent (126 times, 79.8%) rather than refer to themselves (32 times, 20.3%). Concerning the cases associated with the emotional meaning of *touch*, speakers refer to themselves 11 times which amounts to 50% of the cases. That is, the frequency of the pronoun *me* increases when *touch* has an emotional meaning. The difference between the type of meaning (i.e., physical and emotional) and the type of pronoun that accompanied the linguistic expression (i.e., *me* and the other pronouns) is significant, $\chi^2(1, N = 180) = 9.3982, p = .002$. Thus, the utterances conveying the emotional meaning of *touch* tend to be produced with the personal pronoun *me*.

	Pronoun	CQPweb	%
Physical meaning	Me	32	20.3
	You	17	10.8
	Yourself	1	0.6
	Him	15	9.5
	Himself	1	0.6
	Her	7	4.4
	It	66	41.8
	Us	1	0.6
	Them	14	8.9
	Someone	2	1.3
	Somebody	2	1.3
Total		158	100
Emotional meaning	Me	11	50
	You	5	22.73
	Yourself	-	-
	Him	2	9.09
	Himself	-	-
	Her	-	-
	It	-	-
	Us	2	9.09
	Them	-	-
	Someone	2	9.09
	Somebody	-	-
Total		22	100

Table 6. Distribution of pronouns that appear in the *touch + personal pronouns* search in CQPweb

As far as the use of intensifiers is concerned, it is infrequent with the physical sense of *touch* and pervasive with the emotional one (Table 7). Within the group of related gestures when *touch* had a physical meaning, only one out of 186 cases included an intensifier, a mere 0.5%. In contrast, speakers used intensifiers in 56 of the cases in which they did a related gesture when *touch* had an emotional meaning; this is 48.3% of the cases.

Concerning negation words, they are commonly found when *touch* has a physical meaning but extremely rare in its emotional sense (Table 7). Negation words were present in 62 (33.3%) of the cases associated with the prototypical meaning of *touch* where speakers made a related gesture. In comparison, only two negation words (1.7%) were found in the list of cases linked to the emotional meaning of *touch* where speakers made a related hand gesture.

	Intensifiers		Negation words	
	Cases	%	Cases	%
Physical meaning	1/186	0.5	62/186	**33.3**
Emotional meaning	56/116	**48.3**	2/116	1.7

Table 7. Frequency of intensifiers and negation markers in physical and emotional meanings that co-occurred with a related gesture

Taking into account the information above (see Table 8), a chi-square test was calculated to examine the relation between the distribution of linguistic markers (intensifiers and negation words) and the meanings of *touch* among all valid cases (i.e., non-repeated cases with visible hands). The result, $\chi^2(1, N = 278) = 239.1518$, $p < .001$, shows a significant association between both variables. Thus, speakers are more likely to use negation words when the sense of *touch* is physical and they appear to use intensifiers when the sense of the verb is emotional.

	Intensifiers		Negation words	
	Cases	%	Cases	%
Physical meaning	7/491	1.4	150/491	**30.5**
Emotional meaning	118/250	**47.2**	3/250	1.2

Table 8. Frequency of intensifiers and negation markers in physical and emotional meanings among valid cases

After verifying that there was a statistically significant association between these two variables, we wondered if the meaning of *touch* could be predicted based on the use of intensifiers and negation. For this reason, a binary logistic regression (see Table 9) was calculated to predict whether the meaning was physical or emotional based on the use of intensifiers and negations among the valid cases. The effect of intensifiers and negation words on the meaning of *touch* was statistically significant ($\chi^2_2 = 329$; $p < .001$; $R^2_{McF} = .347$, $R^2_N = .497$). In particular, the presence of intensifiers—and absence of negation—was associated with expressions with an emotional meaning and the presence of negation—and absence of intensifiers—was associated with expressions with a physical meaning.

	B (SE)	**95% *CI* for *odds ratio***		
		Lower	*Odds ratio*	*Upper*
Intercept	-0.947 (0.103)			
Intensifier	4.054* (0.436)	3.20	57.6264	4.908
Negation	-3.478* (0.652)	-4.75	0.0309	-2.201

Table 9. Binary logistic regression model including Standard Estimate and 95% Confidence Interval for odds ratio

3.1.4. Discussion

3.1.4.1. Gesture Frequency

Despite the fact that the distribution of the categories related gesture and no gesture among the physical and emotional meanings of *touch* was not statistically significant, these data also uncover some interesting points. In this work, the physical meaning and the emotional meaning of *touch* appear to be equally distributed in both categories (related gesture and no gesture). Thus, the difference in meaning (physical vs. emotional) does not influence the number of gestures speakers make. A potential explanation for this could be that speakers do not vary the frequency of gesture they make when they use either the physical meaning or the emotional meaning, instead they rather change the type of gesture (i.e., its formal features) they use with each verb sense. For example, when speakers express the emotional meaning of *touch*, they tend to physically touch their chest and they never reach out to external objects—as happens when speakers convey its physical meaning.

3.1.4.2. Gesture Referent

Concerning the proportion of cases in which speakers physically touch the gesture referent, it varies depending on the meaning of *touch*. People touch the gesture referent more when the meaning of this perception verb is emotional (38.8%) than when it is physical (21.5%). Thus, speakers tend to visually indicate and specify the referent of their gesture—which might be

crucial for determining the verb sense as it is typically the same as the 'receiver of the action'—when they are expressing the emotional meaning of *touch*. In addition to this, speakers tend to physically touch different types of gesture referents depending on the sense of *touch* expressed in speech. In fact, the gesture category other-touch (i.e., speakers touch an object or another person) was only found when speakers expressed the physical meaning of the verb. Even though the category self-touch (i.e., speakers touched their arm, chest, leg or belly) could be found with both meanings, it shows greater proportion in the emotional meaning. This is due to the high frequency of chest-touching gestures in emotional cases (48.2%), which is not only higher than the frequency of chest-touching gestures in physical cases (5.9%) but also higher than the overall frequency of self-touch in physical cases (18.8%). This would suggest that, while physical experience is connected to the external elements in our surroundings, emotions are conceived as an internal phenomenon developed and contained within our bodies.

This bond between chest-touching gestures and the emotional meaning of *touch* certainly is the most interesting result obtained from the gesture analysis. The relevance of this gesture goes beyond its frequent use since it seems to represent the Western conceptualization of hearts as the seat of emotions. Chest-touching gestures could be considered a way of depicting that something or somebody reached the speakers' heart. The conceptual metaphor that seems to motivate the production of these gestures is THE HEART IS A CONTAINER FOR EMOTIONS (Kövecses, 1986; Lakoff & Johnson, 1980) which is at the same time included in the broader metaphor THE BODY IS A CONTAINER FOR THE EMOTIONS (Lakoff, 1987) that can be found in many languages (see Ibarretxe-Antuñano & Valenzuela, 2021, for an overview of this topic). For example, in English people can 'brim with excitement' and in Spanish people can be full of rage (*llenarse de rabia*). In a few words, it seems that the emotional meaning of *touch* is expressed through the embodiment of emotions by means of chest-touching gestures.

Considering that mappings can sometimes vary depending on the speakers' cultural background, this gesture is not only metaphoric but it is also likely to be culture-dependent. Due to the fact that the language explored in this research is English, it is coherent to find a conceptualization belonging to the Western world. However, as claimed in Ibarretxe-Antuñano and Valenzuela (2021), other cultures conceive the seat of emotions to be located in different organs such as the belly as is the case in Kuuk Thaayorre (Gaby, 2008), the liver in Indonesian (Sharifian et al., 2008), or even different organs as happens in Chinese, where anger is situated in the liver, fear in the kidneys, sadness in the lungs and joy in the heart (Yu, 2009). Taking this information into account, we cannot state that chest-touching gestures are associated with the expression of emotions in non-Western cultures since this path has not been researched yet. Hence, the gestures co-occurring with the narration of emotional events in languages such as the aforementioned needs to be explored in order to determine which are the gestures speakers use and whether these also imply the embodiment of emotions.

Another explanation for the propensity to use chest-touching gestures when the verb *touch* transmits an emotional meaning could be due to speakers seeking to reinforce this less frequent meaning of the verb by adding a more salient multimodal element. Considering the apparent lower frequency of use of the emotional meaning of *touch*, people appear to draw upon a more noticeable gesture—which does not only involve the hands but also a part of the speaker's body—to facilitate the recognition of the 'non-prototypical' verb sense (i.e., emotional meaning). Furthermore, this explanation does not conflict with the potential use of co-speech gestures as elements helping to disambiguate the meaning of polysemous words. The results of this work suggest that gesture may be a useful supplementary level of analysis in the field of polysemy. Considering that, when there is no access to additional linguistic information, meaning disambiguation seems an impossible goal, the modality of gesture could be valuable to solve those cases occurring in situations with insufficient context or potential ambiguity.

In fact, Ibarretxe-Antuñano (2006) declared that the emotional meaning of *touch* is easy to identify when context is provided, but quite challenging to detect when it appears 'isolated' since it could have several interpretations. This situation was illustrated in Ibarretxe-Antuñano (2006) with the sentence 'John touched Mary' (as discussed in the introduction). This is the reason why similar cases are called unpredictable cases of polysemy and their potential meanings argument-driven extensions (see Ibarretxe-Antuñano, 2006, for the detailed explanation of these concepts). Expressed differently, the meaning of this type of sentences cannot be predicted only by the arguments that the verb takes.

3.1.4.3. Pronoun Frequency

The meanings of *touch* can also be distinguished by looking at the personal pronouns used with each type of meaning. Focusing on the *touch* + *personal pronouns* search, when speakers made a hand gesture related to the physical meaning of *touch*, they mostly addressed an external referent (79.8%). However, when the hand gestures were associated with the emotional meaning of the verb, speakers referred to themselves in half of the cases. Expressed differently, when speakers convey the physical meaning of *touch*, they make use of a wider variety of personal pronouns (where *it* is the most repeated one). The reason why physical cases occur with a diverse range of pronouns is simple: this meaning is more likely to include more elements (i.e., linguistic referents) because the description of external events depends on multiple environmental stimuli rather than on one, as happens when we refer to emotional experiences. Considering that the cases related to the emotional meaning of *touch* tend to be produced with the personal pronoun *me*, it appears that speakers do not usually use the verb *touch* to describe other people's internal processes. The high proportion of the emotional meaning of *touch* being used to express our own emotions might occur due to the potential difficulty that entails reading people's genuine feelings.

On the whole, this correlation between the personal pronoun accompanying *touch* and the meaning expressed appears to support the aforementioned gesture differences (self-touch vs. other-touch). This means that the wider the variety of pronouns used, the greater the diversity of gesture referents touched by speakers. The only issue concerning the distribution of personal pronouns in our sample of gestures is the fact that sometimes it might be challenging to determine their role; expressed differently, to decide whether chest-touching gestures were motivated by the emotional component of touch or originated as an 'pronoun emphasizing strategy.' As chest-touching gestures used with the emotional meaning of touch are linked to the pronoun *me*, it could be argued that speakers touch their chest because they want to make the referent (who/what was touched) stand out. However, this is normally done through pointing gestures without the need of physically touching the referent—in this case, themselves (Cooperrider & Mesh, 2022). Nevertheless, another possibility is that these potential roles of chest-touching gestures might coexist together instead of excluding each other.

3.1.4.4. Linguistic Markers

The linguistic analysis showed that there is an interaction between the meaning of *touch* activated and the type of linguistic marker used with the verb. When the co-speech gestures are associated with the physical meaning of *touch*, speakers often include a negation word while omitting intensifiers. In contrast, when the gestures are related to the emotional meaning of the verb, speakers do not use negation particles and they usually strengthen the meaning expressed by using intensifiers. This interaction can be almost conceived as an inverse relation since when a type of linguistic marker appears with the physical meaning, it is rarely manifested in the emotional meaning—and the other way round.

In general terms, linguistic modifiers interact with each other and can be used jointly since their appearance does not exclude the possibility of using more than one type of modifier within a

construction (Carrillo-de-Albornoz & Plaza, 2013). In this research, there is also 'the exception that proves the rule'; that is, one intensifier was found in physical cases and two negative particles in emotional cases. It is worth mentioning that the two negation words found in the emotional cases precede an intensifier and they are used before a comparative structure (e.g., 'the health care debate has not really touched me as much as it has Kelly'). In this sentence, we find a great illustration of the use of negation as a diminisher of the strength of the polarity of the statement. Thus, the actual function of this negation is to emphasize that the speaker's partner was deeply affected by the health care debate. In a few words, it seems that the negation word in this example is not directly used to 'cancel out' the emotional meaning implied by *touch*, but to be specific about the speakers' level of affliction in comparison with his partner.

The association of the physical sense of *touch* with the use of negation might be due to the fact that it is felt as more 'invasive' in comparison to the emotional sense. In other words, speakers mostly use negation in order to warn people that they should not attempt to physically touch them or an external element. On the contrary, using negation words while conveying the emotional meaning of *touch* seems quite futile as individual emotions can be neither controlled nor 'restricted' by interlocutors themselves (Fernández Jaén, 2012, 2016; Jansegers, 2017). Also, physical contact could be potentially conceptualized as a binary feature [±physical contact], something either touches you or it does not.

As previously mentioned, the binary conceptualization of the physical meaning of *touch* does seem to not manifest in the emotional meaning of the verb. Actually, the fact that the emotional sense is associated with the use of intensifiers (e.g., very, really) supports the widely accepted idea that emotions are conceptualized along a gradable scale (Fontaine et al., 2013). Thus, it makes sense that intensifiers modify only emotional meaning (rather than physical meaning) because emotions are usually conceived as something which can have different degrees of intensity. Furthermore, in relation to the lack of intensifiers found when *touch* implies physical

contact, it is quite unlikely that speakers would express the intensity of the touch (perception of pressure, temperature, etc.) through modifiers expressing degree rather than using manner adjuncts (e.g., 'he touched me softly so that he would not scare me'). Concerning the degree of emotional involvement that someone might feel in a particular situation, it can also change from person to person and, as a result, it can be 'graded.' For example, an event might 'touch' someone (i.e., they found the situation moving), but that same event might 'really touch' someone (i.e., they were deeply emotionally involved). In contrast, this specificity in the level of the action expressed cannot be done with the physical sense of *touch*.

Argaman (2009) stated that there is a connection between the intensity of emotions and the lexical modalities that speakers use in their expression of the emotional experience. She concluded that linguistic markers, such as intensifiers, are objective tools for determining people's emotions since we normally give priority to a correct delivery of the content of the message rather than to a finely determined word choice, which is usually a conscious process. Our findings would confirm that speakers often use intensifiers when sharing their emotions, in our case, in oral production. All things considered, the correlation between the use of intensifiers and the expression of emotional meaning could potentially apply generally and not only in relation the perception verb *touch*.

Analyzing the linguistic markers modifying polysemous verbs could be crucial when trying to determine the meaning conveyed. "In some cases, the polysemy is due both to one of the words of the sentence, in this case the verb, and to the meaning of the other arguments that the verb takes" (Ibarretxe-Antuñano, 1999, p. 69). The connections found between linguistic modifiers and the meanings of *touch* could be much more than mere associations, as they also reveal specific ways in which speakers conceptualize these meanings. Thus, the results in this article would confirm Ibarretxe-Antuñano's claim (1999, 2006) that polysemy is not always localized

on a single word (*touch*), but also on other elements of the sentence (in our case, negation and intensifiers) which may contribute to trigger one specific meaning.

3.1.5. Conclusion

This research has adopted a multimodal perspective to explore the expression of the physical and emotional meanings of the perception verb *touch*. Our initial hypothesis has been confirmed: speakers use multimodal information (speech and gesture) to differentiate the meanings of the polysemous verb analyzed. With regards to our first objective, we have seen that while the gestures associated with the physical meaning tend to reach an external referent (other-touch), the gestures associated with the emotional meaning only involve touching the speaker's body (self-touch). Concerning our second objective, the most frequent co-speech gesture, the chest-touching gesture, was mainly used when touch had an emotional meaning. Its motivation appears to be the conceptual metaphor THE HEART IS A CONTAINER FOR EMOTIONS (Kövecses, 1986; Lakoff & Johnson, 1980); therefore, this is the structuring mechanism by which chest-touching gestures embody the emotional meaning of *touch*. In relation to our third objective, linguistic elements distinguishing both meanings (pronouns, negation and intensifiers), this work showed that the physical meaning of *touch* is characterized by its use of a wide variety of personal pronouns and of negation words; in contrast, the emotional meaning of the verb *touch* occurs mostly with the personal pronoun *me* and it is usually modified by intensifiers. These marked differences reveal two issues: the conceptualizations of each sense and the way polysemy works. While the physical meaning appears to be considered a binary feature, the emotional meaning seems to be conceived along a gradable scale. Thus, this relation between the linguistic context and the meaning of *touch* corroborates that polysemy can be distributed among several elements of the sentence and not only on the polysemous item itself.

Redirecting the research focus of polysemy studies toward additional communication modalities could certainly contribute to expand our knowledge about this phenomenon and improve future works in this field. The present investigation could generate future experimental studies, which could test whether gestures are indeed elements helping to disambiguate meanings of *touch* combining the gesture patterns detected in this work with examples of unpredictable polysemy. Apart from this, there are of course questions that we are not able to answer just yet. First, more exploratory research is necessary to firmly establish whether this multimodal dimension can be found in other body-bound metaphors and in different languages. It would also be very useful to check whether the distinction of meanings through gesture is only related to the metaphors motivating the emotional meaning of *touch* or if this also happens with other figurative meanings, regardless of their emotional component.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/langcog.2024.23.

Data availability statement. All the data collected to conduct this analysis can be found at https://osf.io/v8gp3/overview.

Funding statement. The research reported here was supported by the 21596/FPI/21 Fundación Séneca, Región de Murcia (Spain), predoctoral fellowship to I.B.M.; by a Fulbright scholarship to I.B.M. (00001/ FLB/22, Fundación Séneca, Región de Murcia-Fulbright); by the European Union – NextGenerationEU (Beatriz Galindo award) to D.A.C. and by the Ministerio de Ciencia e Innovación (Spanish State Research Agency) and Feder/UE funds (grant number PID2022-139194NB-I0).

3.2. Disambiguating Polysemy: How Hand Gestures Help Observers Interpret the Verb Touch

This paper examines whether observers use gestural information to decide the meaning of the polysemous verb *touch* in ambiguous contexts. To address this question, three studies were carried out. Study 1 tests whether observers could accurately distinguish the meaning of the verb *touch* just by looking at hand gestures. Study 2 explores which gesture location and handshape combinations are associated with the physical and emotional meanings of *touch*. Study 3 investigates whether observers decide the meaning of *touch* faster when they see a co-speech hand gesture and whether reaction time varies depending on the specific gesture combination observed. The main findings illustrate how the modality of gesture helps observers to disambiguate the meaning of a polysemous word such as the verb *touch*. Thus, this research shows that location and handshape are key components that bias the meaning of *touch* when the verbal message is ambiguous or absent.

Keywords: polysemy, gesture, meaning disambiguation, ambiguity, touch

3.2.1. Introduction

Human communication is a multimodal event. When we communicate, we convey meaning through a variety of channels. For example, when people roll their eyes as they speak, they are likely to be displaying their frustration; when they talk in a shaky voice, they could be indicating insecurity or nervousness. In both cases, meaning that goes beyond the speaker's verbal message is expressed.

One of the components of communication that interacts with our spoken words is gesture. We naturally embellish our everyday conversations with gestures that deepen the meaning and intention of our words (Goldin-Meadow, 2023). Since the pioneering investigations in the late 60s and early 70s (Condon & Ogston, 1967; Efron, 1972), gesture has been considered central

to human communication—not only in relation to language but also in the cognitive processes underlying communication. In fact, several authors argue that speech and gesture are integral parts of a single process of utterance formation (McNeill, 1992; Kendon, 1997; Goldin-Meadow, 2003; Özyürek, 2014; Hostetter & Alibali, 2019; Hagoort & Özyürek, 2024).

Gestures can be simply defined as body movements—commonly involving the hands, head, face or shoulders—that convey relevant information during communication. Kendon (2004, p. 15) defined gestures as visible actions that manifest deliberate expressiveness and are used "as an utterance or as part of an utterance," where an utterance is understood as a unit of activity treated as a communicative contribution. The unit may be constructed from visible bodily action, speech or a combination of both modalities. Gestures are distinct from other types of movements in that they do not have a direct effect on the world; rather, they affect the world indirectly by playing a role in communication (e.g., a twist gesture in the air can result in jar opening only if someone interprets it as a request that the jar lid be twisted open; Goldin-Meadow, 2003).

Although gestures can be produced without speech, our focus here is on co-speech hand gestures. The notion of co-speech gesture makes reference to the spontaneous movements of the hands and arms produced along with speech, whose meaning is determined in relation to that speech (McNeill, 1992). The meaning of a co-speech gesture depends on the context provided by the speech it accompanies, as its content must be semantically coordinated with the speech (Alcaraz-Carrión, 2018). Co-speech gestures are tightly intertwined with speech in timing, meaning and function, and are crucial communicative components, as they reflect thoughts that underlie a speaker's message (Goldin-Meadow & Wagner, 2005). Indeed, gesture can reveal aspects of a speaker's conceptualization that cannot be accessed through speech alone (e.g., Peiró Márquez & Ibarretxe-Antuñano, 2021; Alcaraz-Carrión et al., 2022; Lorson et al., 2025)—which might be one of the reasons why gestures can enhance our everyday

communicative exchanges. For example, gestures are used to establish locations or track references in discourse (i.e., cohesive gestures), which then facilitate language comprehension (Gunter et al., 2015) and influence the interpretation of ambiguous pronouns (Smith & Kam, 2012; Nappa & Arnold, 2014). Thus, co-speech gestures not only emphasize the message expressed in speech but also convey information that adds to and complements what is said, often improving understanding.

Given that gesture often provides different information from the verbal message, co-speech gestures have the potential to be helpful in resolving ambiguous communicative exchanges, particularly when the verbal channel or immediate environment does not disambiguate the message (Holler & Beattie, 2003). Gesture might then be a reliable resource for speakers and interlocutors to use to disambiguate meaning.

Language ambiguity is a frequent feature of spoken communication (Piantadosi et al., 2012). Two sources that contribute to ambiguity are homonymy (a word has distinct unrelated meanings; e.g., bat: 'animal' or 'object used to hit a baseball') and polysemy (a word has distinct related meanings; e.g., card: 'ID format' or 'playing element'). Although lexical ambiguity studies do not always explicitly distinguish between homonymy and polysemy, research has shown that this distinction is psychologically real as it is reflected in online processing. Previous studies have provided evidence for a processing advantage for polysemy and a processing disadvantage for homonymy (Klepousniotou, 2002; Rodd et al., 2002; Klepousniotou & Baum, 2007; Carston, 2021). Despite extensive exploration of homonymy and polysemy as sources of ambiguity, disambiguating the meanings of homonyms and polysemous words can be challenging, particularly when relying solely on adjacent verbal elements. In everyday interactions, this challenge is often resolved through additional context, ranging from shared conversational and cultural background to subtle strategies such as contrastive stress. However, sentences such as 'he ate his lunch down by the bank' require

additional sources of information to determine the intended meaning of *bank* (Barnard & Johnson, 2005; Barnard et al., 2006). Thus, combining information from all available modalities may be crucial for successful word sense disambiguation.

Widening the lens to include gesture has proven to be useful in the case of homonyms and pronouns (Holler & Beattie, 2003; Holle & Gunter, 2007; Hinnell & Parrill, 2020). Two decades ago, Holler and Beattie (2003) showed that speakers are more likely to gesture when producing homonyms than non-homonyms. Unlike Holler and Beattie (2003) who were interested in the production of gestures in the context of a homonym, Holle and Gunter (2007, p. 1177) asked "whether listeners make use of this gestural information in comprehension." Since the information conveyed in a speaker's gestures can bias the activation of a word meaning, gesture might help a listener disambiguate the meaning of a homonym. This is precisely what Holle and Gunter (2007) found—gesture made the less frequent meaning of a homonym more salient. More recently, Hinnell and Parrill (2020) studied the resolution of ambiguous statements related to neutral preferences (e.g., whether to take the train to a ballgame or drive) or moral preferences (e.g., human cloning is acceptable or unacceptable). In this study, participants watched stimuli where a gesture to the left or right side co-occurred with speech expressing each position. Then, in gesture-disambiguating trials, ambiguous statements (e.g., 'I agree with that,' where *that* could refer to either idea) were accompanied by a palm-up open-hand gesture to one side or the other. Hinnell and Parrill (2020) found that participants relied on gesture to understand speakers' ambiguous statements. Overall, gesture can disambiguate homonyms and pronouns in ambiguous contexts for both speakers and listeners.

However, little work has been done on the impact of gesture on ambiguous polysemous words. This gap motivates the focus of the present paper on the phenomenon of polysemy. As previously noted, polysemy is the synchronic association of several related meanings with a

single form (Sweetser, 1990). Regarding polysemous word forms, they are believed to have a prototypical meaning—typically their most frequent semantic interpretation—to which additional senses are related in a motivated way. The fact that these additional semantic extensions are non-prototypical does not mean that they are not frequently used in communicative exchanges (see Rosch, 1973, for a detailed explanation of Prototype Theory). For instance, the prototypical meaning of the verb *see* refers to visual perception (e.g., 'I can't see anything without my glasses'), but the verb is also frequently used metaphorically to convey comprehension (e.g., 'I see what your daughter means'). The latter sense is motivated by the conceptual metaphor UNDERSTANDING IS SEEING, which reflects the notion that, in many cultures, sight is considered the most reliable source of information and therefore closely associated with knowledge and understanding. This example underscores that the central aspect of polysemy lies in the underlying cognitive mechanisms that generate and connect the different senses of a word—rather than in the mere multiplicity of its meanings.

In particular, the present paper focuses on cases of polysemy that are unpredictable (i.e., uninterpretable). These cases occur when the words surrounding the ambiguous construction do not provide clues to its meaning (Ibarretxe-Antuñano, 1999). For example, the sentence, 'John touched Mary,' has two potential interpretations: (a) John made physical contact with Mary, or (b) Mary was emotionally impacted by John's actions. This unpredictable case expresses the meaning 'to affect,' which underlies both potential senses: (a) 'to affect, physically' and (b) 'to affect, metaphorically' (Ibarretxe-Antuñano, 2006). We focused on unpredictable cases of polysemy because they cannot be disambiguated by looking at immediate context, leaving a potential role for gesture.

Before determining whether gesture helps people disambiguate the meaning of unpredictable cases of polysemy, we need to see whether speakers make different gestures for the different senses of a polysemous word. This paper focuses on the verb *touch* since previous research

revealed relevant findings regarding the hand gestures co-occurring with this polysemous verb. Additionally, perception verbs have been widely studied in the field of polysemy (Ibarretxe-Antuñano, 1999, 2006, 2019; Fernández Jaén, 2012, 2016; Jansegers et al., 2015). Bolumar Martínez and colleagues (2024) asked whether speakers produce gestures that differentiate between two meanings of the polysemous verb *touch*: its physical meaning, making physical contact with somebody/something, and its emotional meaning, affecting somebody emotionally. In that research, 302 hand gestures co-occurring with the verb *touch* were annotated according to whether the speaker used the gesture to touch an object or another person (*other-touch*) or to touch their own arm, chest, leg or stomach (*self-touch*). When speakers expressed the physical meaning of *touch*, they tended to produce an other-touch gesture; when they expressed the emotional meaning of *touch*, they produced a self-touch gesture. The majority of utterances expressing the emotional meaning of *touch* were directed at the speaker's chest, indexing the speaker's heart. Gesture thus has the potential to be valuable in determining the meaning of the polysemous verb *touch* if verbal context is ambiguous.

Here, we shift the main focus from production to comprehension by examining whether observers use this relation between gesture and speech to distinguish the meaning of the polysemous verb *touch* in ambiguous contexts. We conducted three studies. Study 1 asks whether observers can correctly guess the meaning of *touch* just by looking at the hand gestures made by the speaker. Study 2 asks whether observers associate particular gestures that vary on location and handshape (e.g., point at speaker's own chest, open palm held on speaker's stomach, fingers touch listener's arm) with the physical vs. emotional meanings of *touch*. Study 3 measures whether observers decide the meaning of the word *touch* faster when a hand gesture co-occurs with it. Our goal is to determine whether gesture plays a role in helping listeners disambiguate the meaning of unpredictable polysemous expressions such as 'you touched me.'

3.2.2. Study 1

3.2.2.1. Methodology

3.2.2.1.1. Task and Stimuli

Study 1 tests whether observers can guess the meaning of the verb *touch* by looking at the hand gestures made by a speaker. The study was distributed online through Qualtrics. Each participant was presented different muted clips followed by a forced-choice question about the meaning of the verb *touch*: the physical meaning of the verb (i.e., making physical contact) or the emotional meaning of the verb (i.e., having an emotional effect). Each participant watched 40 muted clips; half were associated with the physical meaning of *touch*, half with the emotional meaning of *touch*. The common feature of these clips was that speakers uttered the phrase 'touched me'; however, the rest of the sentence could vary (e.g., 'Arnold touched me'; 'she touched me so deeply'). The clips were muted to eliminate any influence that variations in lexical or vocal features might have on the observers' performance. Participants were told that the speaker in each clip said, 'it touched me.' A short practice phase was included for participants to familiarize themselves with the type of video stimuli used.

The clips, which were obtained from Bolumar Martínez et al. (2024), were extracted from a multimodal corpus (the NewsScape Library of TV News managed by Red Hen Lab™), which features a wide variety of speakers (different genders, ethnicities and cultural backgrounds) in diverse communicative contexts. All the clips originally showed a hand gesture. Since we wanted to ask whether hand gestures, rather than facial expressions, are the key element helping participants guess the meaning of *touch*, we created three different visibility conditions of every NewsScape clip (using the software Track-Anything; Yang et al., 2023): *head*, *hands/body* and *whole speaker* (see Figure 7). All of the conditions had a black background to avoid possible effects of the visual context on the participants' decision.

Figure 7. Example of an original clip and its different visibility conditions in Study 1

Each participant was only shown two randomly assigned visibility conditions. There were three different versions of the survey so none of the participants watched the same instance twice (see Table 10). The clip order was also randomized for every participant.

Survey A	**Head**		**Hands/body**	
Total 40 clips	10 phys. meaning clips (stimuli 1-10)	10 emot. meaning clips (stimuli 11-20)	10 phys. meaning clips (stimuli 21-30)	10 emot. meaning clips (stimuli 31-40)
Survey B	**Hands/body**		**Whole Speaker**	
Total 40 clips	10 phys. meaning clips (stimuli 1-10)	10 emot. meaning clips (stimuli 11-20)	10 phys. meaning clips (stimuli 21-30)	10 emot. meaning clips (stimuli 31-40)
Survey C	**Whole Speaker**		**Head**	
Total 40 clips	10 phys. meaning clips (stimuli 1-10)	10 emot. meaning clips (stimuli 11-20)	10 phys. meaning clips (stimuli 21-30)	10 emot. meaning clips (stimuli 31-40)

Table 10. Survey versions in Study 1

3.2.2.1.2. Participants

Participants were recruited from the Roman Family Center for Decision Research (CDR) subject pool. The CDR, which is part of the University of Chicago Booth School of Business, was in charge of the recruitment and distribution of the study. Participants had to be at least 18 years old and proficient speakers of English, and they could take part in only one of the three studies. Those who agreed to participate were reimbursed with $1 every 5 minutes of study.

A total of 110 subjects were recruited. Participants were discarded: 1) if necessary information about them was missing, 2) if they had a vision problem, hearing impairment or language

disability that could affect task performance, 3) if they failed any of the attention checks included in the survey or 4) if they selected the same meaning in 75% of the questions or more. After applying these exclusion criteria, the sample was reduced to 89 participants (see Table 11).

Survey version	Recruited	Discarded	Final sample
Survey A: Head; hands/body	38	10	**28**
Survey B: Hands/body; whole speaker	37	7	**30**
Survey C: Whole speaker; head	35	4	**31**
Total	110	21	**89**

Table 11. Distribution of participants by survey version in Study 1

The mean age of participants was 31 years in the final sample; $N = 56$ (62.9%) identified as women, 28 (31.5%) identified as men, and 5 (5.6%) identified as non-conforming/non-binary.

3.2.2.2. Results

Study 1 has two objectives: (1) to determine whether hand gestures are used to distinguish the meaning of the verb *touch,* and (2) to determine whether hand gestures, rather than facial expressions, are doing the work of distinguishing the two meanings. Table 12 presents the proportion of accurate answers in each of the visibility conditions. When participants could see only the speaker's head, they answered 56.1% of the cases accurately. The proportion of accurate answers increased to 74.5% when participants could see the hands/body and to 74.7% when they could see the whole speaker. We tested whether there was a significant association between the part of the body that was visible to participants and the accuracy of their answer. The chi-square test results indicate a significant association between both variables, $\chi^2(2, N = 3560) = 125$, $p <. 001$.

Visibility condition	Participants	Right answer	Wrong answer
Head (20 clips)	Survey A: N=28	292 (24.7%)	268 (22.7%)
	Survey C: N=31	370 (31.4%)	250 (21.2%)
	Total: N=59	**662 (56.1%)**	**518 (43.9%)**
Hands/body (20 clips)	Survey A: N=28	421 (36.3%)	139 (12%)
	Survey B: N=30	443 (38.2%)	157 (13.5%)
	Total: N=58	**864 (74.5%)**	**296 (25.5%)**
Whole speaker (20 clips)	Survey B: N=30	452 (37%)	148 (12.1%)
	Survey C: N=31	459 (37.6%)	161 (13.2%)
	Total: N=61	**911 (74.7%)**	**309 (25.3%)**
Total		2437 (68.5%)	1123 (31.5%)

Table 12. Accuracy of participants' answers in Study 1

3.2.2.3. Discussion

Study 1 suggests that English speakers can distinguish the meaning of the verb *touch* just by observing the hand gestures that co-occur with the verb. When the verbal information available is not sufficient (i.e., the gesture was produced with the phrase 'touched me' without any additional verbal information), hand gestures bias observers toward the speaker's intended meaning. Participants answered at chance when only the speaker's head was visible, but they increased their correct interpretations to almost 75% when only the hands/body was visible; no additional increase was found when the whole speaker was visible (i.e., head and hands/body). Hand gestures (rather than facial expressions) appear to be the key communicative element that disambiguates the meaning of *touch* for observers.

Interestingly, the speakers in the selected clips did not appear to give their feelings away in their faces (e.g., watery eyes, crying), even when using *touch* to convey an emotional meaning. Facial expressions have been associated with emotions, interpersonal attitudes, conversational signals, among other features (Ricci Bitti, 2014). Even if the speakers in our clips had conveyed emotion in their faces, listeners were unable to use that information to determine the meaning of *touch*.

In contrast, the fact that hand gestures can be used to correctly interpret the verb *touch* suggests that gesture is a reliable resource for disambiguating polysemous lexical items. Bolumar

Martínez et al. (2024) hypothesized that observers would associate other-touch gestures with the physical meaning of *touch* and *chest-touching* gestures with its emotional meaning. Overall, this study supports this hypothesis and indicates that gestures can be used to determine the meaning of a polysemous construction that is not expressed in speech.

Study 1 also offers a methodological contribution—a paradigm for differentiating the central visual component in a multimodal analysis. By isolating facial expressions from hand/body gestures, we were able to determine which body part transmitted the distinguishing information. Our findings suggest that facial expressions, when combined with hand gestures, did not have an additive effect on performance. This methodological contribution is important as the central visual element might differ as a function of the phenomenon. We suggest that this way of isolating parts of the body could be useful in detecting instances of hand gestures, with or without facial expressions, that are transmitting the relevant information.

As an interim summary, our study contributes to our understanding of how observers can use gesture to correctly identify the physical and emotional meanings of *touch*. Speakers use hand gestures that distinguish between the physical and emotional meanings of *touch,* and observers use that information to determine the speaker's intended meaning.

3.2.3. Study 2

3.2.3.1. Methodology

3.2.3.1.1. Task and Stimuli

Our next step was to explore specific gesture features that might influence the observers' interpretation of the physical vs. emotional meanings of *touch*. This study was distributed online through Qualtrics. Participants watched 36 muted clips showing a speaker making a hand gesture next to an interlocutor. Participants had to decide the meaning of *touch* in the sentence, 'you touched me,' uttered by the speaker in all clips. Again, participants had to choose

either the physical or the emotional meaning of the verb in a forced-choice question. A short practice phase was included for participants to familiarize themselves with the type of video stimuli used.

The clips used in this study were recorded in the Goldin-Meadow Lab. They showed a speaker and an interlocutor seated with an off-white wall behind so there was no need to modify the background (Figure 8). Our goal was to control specific gesture features that appeared to be relevant for expressing the meaning of *touch* in the dataset of Bolumar Martínez et al. (2024). We included three gesture locations (the speaker's chest, the speaker's stomach and the interlocutor's arm) and three handshapes (open palm, index finger and fingertips) in the recorded stimuli. The order of clips was completely randomized for every participant.

Figure 8. Example of video stimuli in Study 2

3.2.3.1.2. Participants

Participants were recruited following the procedure used in Study 1. Participants in Study 2 did not participate in Study 1 and/or Study 3. A total of 121 subjects were recruited for this study. After applying the four exclusion criteria (same as in Study 1), the sample consisted of 86 participants.

The mean age of the final sample of participants was 30 years; N = 58 (67.4%) identified as women, 25 (29.1%) as men, 2 (2.3%) as non-conforming/non-binary, 1 person (1.2%) did not identify with any of the gender descriptions.

3.2.3.2. Results

The analysis in Study 2 addresses two objectives: (1) to determine whether specific gesture locations and handshapes are associated with the physical and emotional meanings of *touch*, and (2) to identify which combinations of these gesture features are associated with each of the verb meanings.

Out of 1032 instances that each gesture location was shown, the speaker's chest was associated with the emotional meaning of *touch* 684 times (66.3%). In contrast, the speaker's stomach was associated with the physical meaning 617 times (59.8%), and the interlocutor's arm was associated with the physical meaning 647 times (62.7%).

Out of 1032 instances that each handshape was shown, the open palm shape was associated with the emotional meaning of *touch* 661 times (64.1%). In contrast, the index finger shape was associated with the physical meaning 774 times (75%). The fingertips shape was not clearly associated with either the emotional meaning (565 times, 54.7%) or the physical meaning (467 times, 45.3%).

Figure 9 shows the frequency with which participants selected the emotional (as opposed to physical) meaning of *touch* as a function of both the gesture location and handshape they observed.

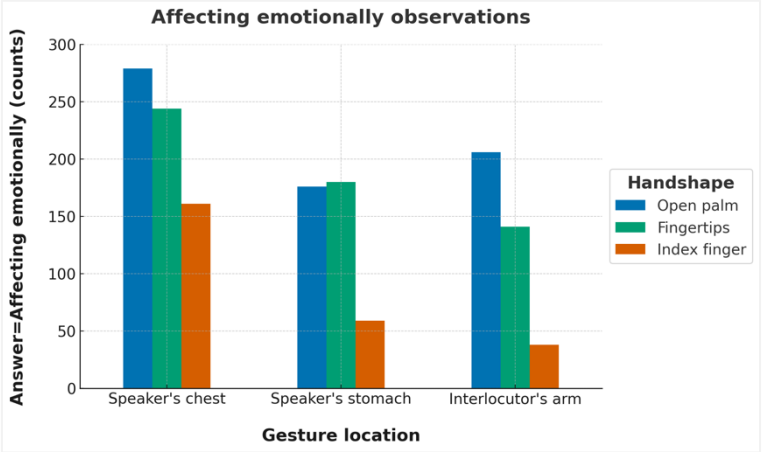

Figure 9. Emotional meaning observations in Study 2 based on gesture location and handshape

A binomial logistic regression was calculated to examine the effects of gesture location and handshape on the interpretation of the verb *touch*. The model was significant, $\chi^2_8(N = 3096) = 627$; $p <. 001$; $R^2_{CS} = .183$, $R^2_N = .245$.

The main effect of gesture location was significant, $\chi^2_2 = 64.8$, $p < .001$. The odds of selecting "affecting emotionally" were higher for the speaker's chest ($\beta = 0.80$, $SE = 0.16$, $OR = 2.22$, $p < .001$), and lower for the interlocutor's arm ($\beta = -0.46$, $SE = 0.15$, $OR = 0.63$, $p = .003$), than for the speaker's stomach (reference category).

The main effect of handshape was also significant, $\chi^2_2 = 122.7$, $p < .001$. The odds of selecting "affecting emotionally" were lower for the index finger ($\beta = -1.67$, $SE = 0.18$, $OR = 0.19$, $p < .001$) than for the fingertips (reference category). No significant difference was found between the *open palm* and the fingertips ($\beta = -0.05$, $SE = 0.15$, $OR = 0.95$, $p = .760$).

Additionally, the interaction between gesture location and handshape was significant, $\chi^2_4 = 27.5$, $p < .001$. The odds of selecting "affecting emotionally" were higher in the gesture combination involving the interlocutor's arm and the open palm ($\beta = 0.81$, $SE = 0.22$, $OR = 2.25$, $p < .001$) than in the combination involving the speaker's stomach and the fingertips (reference categories).

3.2.3.3. Discussion

Study 2 reveals that both gesture location and handshape had significant main effects on the interpretation of the meaning of *touch*. Furthermore, their interaction significantly influenced the outcome, underscoring the importance of these two gesture features in distinguishing between the physical and emotional meanings of the polysemous verb *touch*.

The results reveal that two gesture locations, speaker's stomach and interlocutor's arm, were primarily associated with the physical meaning of *touch*; the third location, speaker's chest, was associated with the emotional meaning. Only one handshape, index finger, was associated with the physical meaning of *touch* and the other two handshapes, open palm and fingertips, were primarily associated with the emotional meaning of the verb.

The gesture combinations most frequently associated with the emotional meaning of *touch* involved the speaker's chest paired with either an open palm or a fingertips handshape. The 'hand-over-heart' gesture (i.e., combination of speaker's chest with open palm) has been previously associated with the concept of honesty as people making this gesture appear more trustworthy to others than people making a control gesture (Parzuchowski & Wojciszke, 2014). Our research extends this finding by adding the aspect of emotion, showing that this particular gesture combination is often interpreted as conveying the emotional meaning of the polysemous verb *touch*. We suspect that the fingertips shape is associated with the emotional meaning of *touch* because of its similarity to the open palm shape. As discussed in Bolumar Martínez et al. (2024), the importance of chest-touching gestures lies in the conceptual metaphor that motivates them: THE HEART IS A CONTAINER FOR EMOTIONS (Kövecses, 1986; Lakoff & Johnson, 1980). In this case, this Western conceptualization is not found in the speakers' verbal message, only in gesture. In other words, although speakers neither uttered a metaphorical expression nor conveyed any emotional content verbally, the emotional meaning

of *touch* was still communicated via gesture—motivated by the conceptualization of the heart as the seat of emotions.

In contrast, when participants watched people touching their chest with an index finger, they did not infer the emotional meaning of *touch* even though they were watching a chest-touching gesture. Using the index finger transforms the interpretation of the chest-touching gesture, shifting it from a metaphoric representation to a deictic one. This difference in the interpretation of the chest-touching gesture confirms that handshape can carry semantic information that impacts the meaning of the gesture. The index finger handshape is typically used to locate items, unlike the open palm or fingertips handshapes, which are used in a broader range of contexts. As a result, the index finger encourages observers to interpret the chest-touch as a reference to the body part itself—but, interestingly, not as often as when the index finger is combined with the speaker's stomach or the interlocutor's arm, both of which primarily elicited physical meaning interpretations. In short, although the index finger is still associated with the physical meaning of *touch* when paired with the speaker's chest, the proportion of physical interpretations is likely lower due to the specific interaction between gesture location and handshape.

Note that the speaker's stomach was predominantly associated with the physical meaning of *touch* despite its similarity to the chest-touching gesture; the axis (sagittal) and the direction (inward) used to perform these two gestures are the same. We did not anticipate such a strong association between the speaker's stomach and the physical meaning of *touch*, as previous research suggests that gestures performed along the sagittal axis signal the social cognitive dimension of warmth, with inward gestures more frequently associated with descriptions of high warmth, compared to outward gestures (Lakshmi, 2024). In light of this work, we suspect that the association of the speaker's stomach with the physical meaning of *touch* is likely driven by the use of the index finger shape, which focuses attention on the body. In addition, when the

speaker's stomach is combined with the index finger handshape, it tends to be interpreted as the physical meaning of *touch* because the body part pointed at is not conceptualized in relation to emotions in Western cultures and, accordingly, lacks the conceptualization underlying chest-touching gestures. In general, participants interpreted gesture combinations involving the index finger as conveying the physical meaning of *touch*.

Overall, Study 2 illustrates how speakers interpret different meanings of *touch* based on specific combinations of gesture location and handshape. Combining speaker's chest location with an open palm or fingertips handshape tends to be associated with the emotional meaning of the verb, whereas combining speaker's stomach or interlocutor's arm locations with an index finger handshape is linked to the physical meaning. These findings highlight the importance of examining how meaning in discourse can be influenced by both the location and handshape of a gesture.

3.2.4. Study 3

3.2.4.1. Methodology

3.2.4.1.1. Task and Stimuli

Study 3 asks whether figuring out what *touch* means is faster when a gesture is presented along with the word than when the word is presented alone. This study was conducted in person at the CDR campus lab and Mindworks location. The first part of the study presented different clips to participants, followed by a question about the meaning of the verb *touch*. As in the previous studies, participants had to select either the physical or the emotional meaning of the verb *touch*. A short practice phase was also included to familiarize participants with the video stimuli and the software. The participants' reaction times (RT) were recorded by means of key presses, which were counterbalanced. RT in this study corresponds to the time elapsing between the beginning of a clip and the moment participants pressed the key. The first part of

the experiment was built using the software PsychoPy (Peirce et al., 2019). The second part of the experiment consisted of a follow-up Qualtrics questionnaire designed to obtain insight into the results. Participants were presented one of two study versions: one in which the speech on the clips was audible (through headphones), and one in which the speech was muted.

The 48 clips used in this study were a combination of the 36 clips containing gesture used for Study 2 and 12 additional clips with no gesture, recorded in the same setting. The clips showed a speaker and an interlocutor seated (see Figure 10), as in Study 2. Each participant watched 24 (of 48) clips[7] in which a speaker uttered the sentence 'you touched me': 18 Speech + Gesture (S+G) clips showed speakers making a hand gesture; 6 Speech (S) Alone clips showed speakers not producing a gesture. In the 18 S+G clips, the different gesture locations (speaker's chest, speaker's stomach and interlocutor's arm) and handshapes (open palm, index finger and fingertips) were equally distributed. The 6 S Alone clips showed different speakers uttering 'you touched me' without moving their hands. A second group of participants were shown the 18 S+G clips but without speech, that is, the Gesture (G) Alone condition (these participants also saw 6 clips that contained neither S nor G so that they would see an equal number of stimuli; responses to these clips were not included in the analyses). The presentation order of the clips was randomized for every participant.

[7] There were two sets of 24 clips which were randomly assigned to participants.

Figure 10. PsychoPy screen in Study 3

3.2.4.1.2. Participants

Participants were recruited following the same procedure used in Study 1 and Study 2, with one exception. Participants who carried out the experiment at the Mindworks location were given 100 points every 5 minutes of study; these points were redeemable for different prizes (e.g., water bottle, gift cards). Participants in Study 3 did not participate in either Study 1 or Study 2.

A total of 244 subjects were recruited for this study. After discarding those who met the exclusion criteria, the final sample consisted of 212 participants (114 who saw S+G and S Alone clips; 98 who saw G Alone clips). The exclusion criteria were the same as in Study 1 and Study 2, but we also excluded participants whose answers were not correctly saved due to an error during data collection.

In the final sample, the mean age of participants was 31 years. In relation to gender, $N = 117$ (55.2%) identified as women, 82 (38.7%) identified as men, 8 (3.8%) identified as non-conforming/non-binary, and 5 (2.4%) preferred not to respond.

3.2.4.2. Results

Before beginning data analysis, we discarded outliers who were 1.25 *SD* from the RT mean since our sample was not normally distributed. The RT mean decreased from 3.82 s to 3.26 s in the final sample.

3.2.4.2.1. Effect of Speech and Gesture vs. Speech Alone on RT

First, we asked whether participants decided on the meaning of the polysemous expression 'you touched me' faster when they heard the speech and watched the hand gestures, S+G stimuli, than when they only heard the speech, S Alone stimuli. Participants reacted 0.49 s faster to S+G stimuli (RT mean = 3.17 s) than to S Alone stimuli (RT mean = 3.66 s). In order to determine whether this difference was statistically significant, we carried out an independent sample t-test. The test verified that the participants' RT to S+G stimuli is significantly lower than their RT to S Alone stimuli ($t_{2519} = 7.5$; $p < .001$). The size of the observed effect is low (Cohen's $d = 0.351$).

When participants viewed the stimuli with speech but without gesture (S Alone), they were more likely to interpret the sentence 'you touched me' as physical rather than emotional (55.9% of responses were physical, 44.1% were emotional). Interestingly, when gesture was added to the stimuli (S+G), participants showed the opposite pattern (44.4% of responses were physical, 55.6% were emotional). Reaction Time (although overall longer for the S Alone stimuli than for the S+G stimuli, as just shown) displayed the same pattern: faster RT for physical interpretations (3.57 s) than for emotional interpretations (3.77 s) when speech was presented without gesture; slightly slower RT for physical interpretations (3.22 s) than for emotional interpretations (3.13 s) when speech was present with gesture.

Participants thus settled upon a meaning for the polysemous expression in speech more quickly when gesture was included than when it was not, suggesting that participants are indeed gleaning meaning from gesture that is speeding up their reaction to the event.

3.2.4.2.2. Effect of Speech and Gesture vs. Gesture Alone on RT

To control for the possibility that participants might be responding more quickly to a stimulus containing speech and gesture than to a stimulus containing speech alone simply because there is more information in S+G than in S Alone, we looked at RT in the 98 participants who responded to the 18 G Alone stimuli. In a between-subjects analysis, we found that RT to G Alone stimuli was 3.22 seconds, which was not significantly different from RT to S+G stimuli (3.17 s; $t_{3521} = 1.07$; $p = .283$), and significantly different from RT to Speech Alone stimuli (3.66 s; $t_{2202} = -6.57$; $p < .001$; Cohen's $d = -0.314$).

When presented with stimuli with gesture and no speech (G Alone), participations evenly split their responses between physical and emotional responses (50% vs. 50%) and reacted equally fast when they gave physical vs. emotional responses (3.22 s vs. 3.22 s). This contrasts with the previously mentioned fact that participants were more likely to respond with emotional than physical responses (56% vs. 50%), and were faster at making their emotional responses than their physical responses (3.13 s vs. 3.22 s), when gesture was accompanied by speech in S+G stimuli.

3.2.4.2.3. Effect of Gesture Location and Handshape Interaction on RT

Participants decided the meaning of *touch* faster when the stimuli included gesture, with speech (S+G) or without (G Alone). Our next step was to examine RT for each gesture combination. We conducted a linear regression, including the interaction between gesture location and handshape, to assess whether RT means differed depending on the location/handshape combination observed (see Table 13). Only responses that aligned with the expected

interpretation, as determined in Studies 1 and 2, were considered, that is, gestures involving the speaker's chest when interpreted as emotional meaning (672 of the 1185 combinations, 56.7%), and gestures involving the speaker's stomach/interlocutor's arm when interpreted as physical meaning (1603 of the 2338 combinations, 68.6%). The test effect was significant in both S+G stimuli ($F_{8,1218} = 7.64$; $p < .001$) and G stimuli ($F_{8,1039} = 5.82$; $p < .001$).

The shortest RT in response to the speaker's chest was observed when it was combined with the open palm handshape (2.41 s in S+G stimuli, 2.57 in G stimuli). The shortest RT in response to the speaker's stomach occurred when it was combined with the index finger (3.01 s in S+G stimuli, 3.08 in G stimuli). Finally, the shortest RT in response to the interlocutor's arm was found when it was combined with the fingertips in S+G stimuli (3.03 s) or with the open palm handshape in G stimuli (2.77 s). Interestingly, there were no significant differences in how quickly participants reacted to S+G stimuli (3.03 s) and G Alone stimuli (3.09 s) when their responses aligned with the expected interpretation ($t_{2273} = 1.07$; $p = .283$). These findings indicate that, although RT varied as a function of gesture location and handshape, it was not affected by whether the gesture was accompanied by speech.

Expected interpretation	RT estimated marginal means (s)						
		Open palm		**Index finger**		**Fingertips**	
		S+G	**G**	**S+G**	**G**	**S+G**	**G**
Emotional	**Speaker's chest**	2.41	2.57	3.28	3.31	3.01	3.15
Physical	**Speaker's stomach**	3.33	3.55	3.01	3.08	3.37	3.44
	Interlocutor's arm	3.12	2.77	3.24	3.20	3.03	3.04

Table 13. RT estimated marginal means of each gesture combination when the participants' answer aligned with expected interpretation in Study 3

3.2.4.3. Discussion

This study provides further evidence that hand gestures can be an important resource in disambiguating unpredictable polysemous expressions such as 'you touched me.' When deciding the meaning of *touch*, participants responded significantly faster when they heard the

speech and observed a co-speech hand gesture (S+G stimuli) than when gesture was absent (S Alone stimuli). This difference suggests that seeing gesture and hearing speech helped participants disambiguate the unpredictable polysemous expression 'you touched me' more quickly than they would have had they only heard the speech.

Importantly, there were no significant differences in RT when participants responded to S+G vs. G Alone stimuli, even when considering only responses that aligned with the interpretations we expected based on Study 2. This finding indicates that participants reached a decision just as quickly with gesture alone as with both speech and gesture. Notably, the reduction in RT (relative to RT for S Alone) was driven by the presence of gesture rather than the combination of both modalities. Thus, in the case of the polysemous verb *touch*, gesture not only provides information to disambiguate the meaning of the verb but also facilitates a faster resolution of meaning.

The second result that stands out in this study is the different trends in the distribution of selected meanings depending on the stimuli presented. When only hearing speech (S Alone stimuli), participants were slightly more likely to select the physical meaning of *touch* and reacted faster when doing so. When speech co-occurred with gesture (S+G stimuli), the opposite occurred: participants were slightly more inclined to select the emotional meaning of the verb and also responded more quickly when choosing this meaning. Interestingly, when presented only with gesture (G Alone stimuli), participants selected both meanings evenly and reacted equally fast for each. This finding suggests that when observing S Alone stimuli, the decisions participants made were based on the frequency of use of each meaning, as the physical meaning of *touch* is the prototypical and more commonly used, compared to the emotional semantic extension. The reverse trend in S+G stimuli may have originated because participants who watched these stimuli had also been exposed to S Alone stimuli. In other words, unlike with S Alone stimuli, participants sought the additional semantic information

conveyed through gesture. The variability of gesture combinations likely prevented participants from defaulting to the physical meaning of *touch*. In summary, while the physical meaning of the verb was the default interpretation when watching S Alone stimuli, the introduction of gesture (S+G stimuli) led to a shift in meaning interpretation.

Finally, we checked that, when participants' responses aligned with the expected interpretation (based on Study 2), the particular combination of gesture location and handshape had a significant effect on RT. Regardless of the presence of speech, participants reacted more quickly when observing certain gesture combinations over others. The combination of speaker's chest + open palm elicited the fastest reactions, both in S+G stimuli and in G Alone stimuli; this might be the prototypical gesture for expressing the emotional meaning of *touch* as suggested by Bolumar Martínez et al. (2024). Its relatively fast reaction time may result from the fact that it embodies the conceptual metaphor underlying the emotional meaning of *touch* (THE HEART IS A CONTAINER FOR EMOTIONS; Lakoff & Johnson, 1980; Kövecses, 1986). When individuals experience intense sadness or anxiety, they often feel (noncardiac) chest pain, which could increase the likelihood of them touching their chest.

In short, Study 3 demonstrates that observers determine the meaning of *touch* faster when a co-speech hand gesture is present than when it is not. Indeed, the time it takes to react to these gestures when presented without speech is no different from the time it takes to react to them when presented with speech. Gesture is a powerful stimulus in combination with speech and even on its own.

3.2.5. Limitations and Future Directions

While these findings are promising, they should be interpreted with caution since this research has some limitations. The actors in the stimuli did not reflect a wide range of ethnicities and ages, nor were they balanced for gender. These factors could have an impact on gesture

interpretation. Moreover, our participant sample included both native English speakers (L1) and second-language English speakers (L2). Although no significant differences in task performance were found based on the participants' mother tongue, this mixed linguistic background might still be a limitation, as it could introduce variability that is not fully accounted for.

Further research is needed to address unanswered questions. For example, future research could explore the effect of handshape when location varies as a function of axis (e.g., chest vs. stomach) in additional communicative contexts. Another question is whether the association between chest-touching gestures and emotional meaning is shared across non-Western cultures. Finally, future studies on the role of gesture in the disambiguation of meaning should extend the variety of polysemous words to build on these findings.

3.2.6. Conclusion

The three studies presented in this paper investigate whether observers use gesture to decide on the meaning of the polysemous verb *touch*. In Study 1, observers were able to accurately guess the meaning of *touch* just by looking at hand gestures. In fact, when guessing the meaning of the verb, hand gestures were the key element, rather than facial expressions. In Study 2, observers associated specific gesture location and handshape combinations with physical vs. emotional meanings of *touch*. The combinations most strongly associated with the physical meaning of the verb involved either the speaker's stomach or the interlocutor's arm combined with an index finger handshape. The combination most strongly associated with the emotional meaning of *touch* was the speaker's chest paired with either an open palm or a fingertips handshape. In Study 3, observers decided the meaning of *touch* in the polysemous expression 'you touched me' more quickly if the verb was presented with gesture. In fact, they responded just as quickly if the gesture was presented without speech, suggesting that it is the presence of

gesture itself that leads to faster responses. The results of Study 3 also revealed that RT varied depending on the gesture combination that participants watched. The gesture combination with the shortest RT mean was the speaker's chest when combined with an open palm.

The main findings of each study illustrate how gestural information can help observers to disambiguate the meaning of a polysemous lexical item such as the verb *touch*. This paper provides evidence that location and handshape are key gesture components that bias the meaning of *touch* when the verbal message is ambiguous or absent. The findings highlight the relevant role of hand gestures in disambiguating polysemous words, paving the way for future investigations examining how gesture can be exploited as an additional layer of meaning. More specifically, the main contribution of this article lies in demonstrating the potential of gesture to disambiguate the meaning of unpredictable polysemous expressions—such as 'you touched me'—by focusing on their comprehension. This study opens new avenues for future research by highlighting the importance of considering the interaction among different gestural features in meaning-making—an aspect that has often been overlooked. In doing so, it encourages a more integrative approach to gesture analysis, moving beyond the traditional focus on isolated gesture features.

Data availability statement[8]. The data collected to conduct these analyses are publicly available in the following link: https://osf.io/4ze6c/overview.

Ethics statement. The studies involving human participants were reviewed and approved by the Social & Behavioral Sciences Institutional Review Board at the University of Chicago and the Ethics committee at the University of Murcia. The participants provided their written informed consent to participate in the studies.

[8] The video stimuli in this dataset are managed by Irene Bolumar Martínez. If you wish to access the video data, please contact Irene Bolumar Martínez at irene.bolumarm@um.es for further instructions.

Acknowledgments. The research reported here was supported by a Fulbright scholarship to I. B. M. (00001/FLB/22. Fundación Séneca. Región de Murcia-Fulbright); by the 21596/FPI/21 Fundación Séneca, Región de Murcia (Spain), predoctoral fellowship to I. B. M.; by the European Union – NextGenerationEU (Beatriz Galindo award) to D. A. C.; and by the Ministerio de Ciencia e Innovación (Spanish State Research Agency) and Feder/UE funds (grant number PID2022-139194NB-I0). Also, special thanks to the CDR and Mindworks staff for their cooperation and support.

3.3. A Multimodal Approach to Polysemy: The Senses of Taste and Smell

This study explored whether co-speech gestures and linguistic markers help distinguish between the literal and figurative meanings of *taste* and *smell*, building on prior findings related to the verb *touch*. To determine whether patterns found for *touch* generalize to additional verbs, we aimed to 1) compare gestures associated with *taste* and *smell*, 2) examine the motivation behind frequent gestures associated with *taste* and *smell* and 3) describe the relation between spoken words and the meanings of *taste* and *smell*. Gesture analysis showed that while gestures help differentiate the meanings of *touch*, they do not exhibit distinct patterns between literal and figurative meanings of *taste* and *smell*. Linguistic analysis showed that only object quantifiers were frequent when conveying literal meanings of *taste* and *smell*. Although negation and verb modifiers were key in distinguishing the meanings of *touch*, they were not commonly used with either *taste* or *smell*.

Keywords: gesture; perception; polysemy; taste; smell

3.3.1. Introduction

Polysemy, which occurs when a word has multiple related meanings, has traditionally been investigated by studying the context in which polysemous words are used (namely, by looking at the adjacent verbal elements); previous research has shown how this linguistic phenomenon

can also be approached from a multimodal perspective. Bolumar Martínez et al. (2024) suggested that speakers are able to convey two different meanings of the polysemous verb *touch* by combining non-verbal (gesture) and verbal (speech) modalities. In particular, the meanings of interest for this work were the physical meaning of *touch* (i.e., literal; 'establishing physical contact') and the emotional meaning of *touch* (i.e., figurative; 'affecting someone in an emotional manner'). This recent study discovered that when speakers express the physical meaning of this perception verb, they tend to touch an external referent such as an interlocutor or an object (*other-touch* gesture). In contrast, when speakers express its emotional meaning, they are likely to touch their own chest (*chest-touching* gesture), drawing on the Western conceptualization of the heart as the seat of emotions (Nummenmaa et al., 2014, 2018; Farley et al., 2021). Regarding verbal information, this work found that the physical meaning of *touch* often coexists with negation (e.g., never), while its emotional meaning usually occurs with intensifiers (e.g., really), showing how each of the meanings of *touch* analyzed is associated with different linguistic markers.

The present paper aims to extend the investigation of Bolumar Martínez et al. (2024) by investigating whether a multimodal approach can also provide insights into the polysemy of the perception verbs *taste* and *smell*. Given that previous research has shown that gestures and linguistic markers play a meaningful role in the interpretation of *touch*, we aim to determine whether gestures and linguistic markers (negation words, verb modifiers and object quantifiers) also systematically accompany and differentiate between literal and figurative uses of *taste* and *smell*. If so, it would establish gesture as a modality that should be taken into consideration when examining the different meanings of perception words. In contrast, if no other perception verb benefits from a multimodal analysis, this would suggest that the scope of gesture information is not domain-based but rather of lexical nature (i.e., determined by properties of the specific word), and plays a key role exclusively with the verb *touch*.

Based on previous research (Argaman, 2009; Carrillo-de-Albornoz & Plaza, 2013), which demonstrated the association of linguistic markers (negation and intensifiers) with emotional attributes in language, Bolumar Martínez et al. (2024) explored the role these verbal elements had when distinguishing the physical and emotional meanings of *touch*. Although the meaning of *taste* and *smell* is most often determined by the object following the verb, linguistic markers may also be relevant in differentiating the meanings of these two perception verbs. If so, linguistic markers could function as a more 'covert' mechanism for distinguishing the meanings of perception words. Conversely, if this is not the case, it would again suggest that linguistic markers play a role exclusively with the verb *touch* and in the identification of emotional attributes in language.

Therefore, if gesture and/or linguistic markers do not function as a general meaning-distinguishing strategy across polysemous perception verbs, their relevance would appear to be verb-specific, uniquely tied to the semantics of *touch*. This would indicate that gestures do not reveal distinct conceptualizations underlying the literal and figurative meanings of all perception verbs, but rather reflect the embodied nature of *touch*, whose prototypical literal meaning (i.e., physical) inherently involves direct physical interaction, while its emotional figurative meaning is motivated by a container metaphor (THE HEART IS A CONTAINER FOR EMOTIONS; Kövecses, 1986; Lakoff & Johnson, 1980). Regarding linguistic markers, this would suggest that an emotional component is necessary for these elements to contribute to meaning distinctions. Such a finding would underscore the need to consider verb-specific constraints when investigating the role of gesture and/or linguistic markers in meaning differentiation and highlight the limitations of adopting a one-size-fits-all approach to the multimodal study of polysemy.

We chose *taste* and *smell* for three different reasons. Firstly, we wanted to explore these two perception verbs because, together with *touch*, they have been traditionally considered less

important than verbs such as *see* and *hear*, and, as a result, they have been less researched in linguistic studies (Fernández Jaén, 2012). This aligns with the fact that the senses of taste, smell and touch have been labeled as 'lower' senses because the most primitive organisms (e.g., protozoa) only have the chemical senses of taste and smell and a basic sense of touch (Classen, 1997). Secondly, these verbs (including *touch*) express highly subjective experiences due to the need for physical closeness between the perceiver and the experience perceived in order to carry out the action. This argument further contrasts these three perception verbs with those referring to the senses of vision and hearing, which are considered more objective since there is greater distance between the perceiver and the experience perceived (Fernández Jaén, 2012). It is important to note, however, that most perception verbs—regardless of the associated sense—can exhibit systematic metaphorical connections with vocabulary related to the internal self and internal sensations (Sweetser, 1990). Thirdly, we chose *taste* and *smell* since they are less frequently used than *see* and *hear*, and previous works (Pagán Cánovas et al., 2020; Alcaraz-Carrión et al., 2022) have shown that people tend to make more gestures with low-frequency expressions. Furthermore, selecting interconnected domains may provide insight into whether their interconnectedness is reflected in similar or contrasting gesture patterns.

When speakers use the verb *taste* to express a perceived experience, this verb typically means 1) to 'perceive the flavor of a substance by using your tongue' (e.g., 'I can taste a little vanilla'). This is not the only literal meaning of the perception verb *taste*, as it can also mean 2) to 'sample the flavor of food or drink by taking it into the mouth' (e.g., 'you get to taste a piece') and 3) to 'have a specific flavor' (e.g., 'a cicada would probably taste like a cricket'). While these literal meanings are similar, the primary focus varies: meaning 1 centers on the subject doing the tasting, meaning 2 on the action of tasting itself and meaning 3 on the object whose taste is being analyzed. Beyond these literal uses, *taste* has several figurative meanings such as 4) to 'experience something' (e.g., 'they can taste a little bit of what Texas is feeling'). Other

figurative meanings also include the semantic extensions of the literal senses: 5) to ‘perceive the property of something that is not food-related’ (e.g., ‘generations of Afghans who [...] tasted a new life’), 6) to ‘sample the property of something that is not food-related’ (e.g., ‘it's like tasting a cloud’) and 7) to ‘have a specific property that is not food-related’ (e.g., ‘they taste like a shoe’).

With regards to the verb *smell*, it also has several meanings that can focus on different aspects of a perceived experience. One of its literal meanings, which focuses on the subject or action, is 1) to ‘perceive/discover a substance by using the nose’ (e.g., ‘they smelled a foul odor’). Shifting the primary focus from the subject or action to the object reveals additional literal meanings: 2) to ‘have/emit an odor’ (e.g., ‘it smells a little like smoke’) and 3) to ‘have/emit an unpleasant odor’ (e.g., ‘we live in tents, it smells a little bit’). In relation to the figurative meanings of *smell*, they are not used to express ‘olfactory information’ but rather to express subjective evaluation regarding something else, such as assumptions not based on solid evidence, which can be sometimes negative. We find a variety of figurative meanings: 4) to ‘detect/suspect something by means of instinct’ (e.g., ‘I do smell a press play’), 5) to ‘give indications or be suggestive of something’ (e.g., ‘it smells a lot like pay for play’) and 6) to ‘seem untrustworthy or bad’ (e.g., ‘it smells a little fishy’).

Considering that *taste* and *smell* convey multiple figurative meanings alongside their literal uses describing acts of perception, this study explores whether speakers establish any differentiation among these meanings through co-speech gestures apart from verbal information. Expressed differently, rather than focusing solely on gestures that reflect abstract metaphorical mappings, the study also examines how gestures accompany different meanings of these verbs within the physical domain. In particular, this paper analyzes the association of the different meanings of *taste* and *smell* with linguistic information, focusing on the presence or absence of gesture and the use of specific linguistic markers. As this work aims to replicate

what was done in Bolumar Martínez et al. (2024), we have established three parallel objectives: 1) to compare the co-speech gestures associated with each of the selected groups of meanings of *taste* and *smell*, 2) to determine the motivation behind potential frequent co-speech gestures associated with *taste* and *smell*, and 3) to describe the relation between the linguistic context and the literal and figurative meanings of *taste* and *smell*.

3.3.2. Methodology

3.3.2.1. Dataset and Tools

The audiovisual and textual information was obtained from the NewsScape Library of International Television News, a multimodal repository of television news managed by the UCLA and CWRU libraries. This database is part of the Red Hen Lab which is an international cooperative for research into multimodal communication. The dataset includes more than 600,000 hours of television news programs and a 4-billion-word multilingual dataset composed by subtitles that are accurately synchronized with speech. This synchronization is crucial, as it enables researchers to view the exact moment a linguistic expression is uttered—an especially valuable feature when studying low-frequency phenomena that are difficult to elicit in laboratory settings. The communicative exchanges found in NewsScape—and in this study—are set in varied communicative situations like debates and open discussions. While the use of television recordings presents certain limitations, particularly the concern that such data may not fully represent everyday communication, the large-scale nature of the NewsScape database grants access to a wide array of communicative contexts and considerable speaker diversity (e.g., in terms of gender, ethnicity and communicative style). In this regard, we consider the use of NewsScape to be a reasonable methodological tradeoff.

In addition to this, we employed the corpus software CQPWeb (Hardie, 2012) in order to perform more complex linguistic searches. In particular, we used a sub-corpus of the whole

database called NewsScape English v5 (2,147,483,647 words), which gathers data and recordings from 2006 to 2017.

3.3.2.2. Linguistic Searches

3.3.2.2.1. Taste

We searched for two constructions: 1) *taste a* and 2) *taste like the*/*a*/*an*. The first search was expected to be followed by a greater proportion of concrete objects since the determiner *a* can be used with countable nouns. The second search was expected to provide a 'more creative' variety of objects as it already included a simile; accordingly, we expected the proportion of figurative usages to be higher. We used the NewsScape corpus (time span selected: 1st March 2004 to 3rd October 2022) in order to obtain as many cases as possible. In this search, the number of instances in which *taste* was followed by the determiner *a* was 2180. When *taste* was followed by the preposition *like* preceding the determiner *the*, *a* or *an*, we found 2092 instances. Although the number of matches was very similar for both searches, and despite our efforts to obtain sufficient figurative usages, literal usages of *taste* remained predominant in our final sample. Some of the concordances obtained were: 'they get to taste a portion of them' (WKYC Channel 3 News at 7) and 'this tastes like a shoe' (KABC The View).

3.3.2.2.2. Smell

With the verb *smell*, we followed the same approach carried out with *taste*: we searched for instances in which *smell* was followed by the determiner *a* or by the preposition *like* preceding any determiner. In this case, we decided to use the sub-corpus NewsScape English v5 (2006-2017) in CQPweb so that the sample size was comparable to that of *taste*. When we searched *smell a*, we found 2486 instances, while when *smell like* was followed by a determiner, we found 1755 instances. As with *taste*, literal usages of *smell* remained predominant in our final sample. Some of the concordances obtained were: 'you could be smelling a really good

quesadilla' (KNBC Access Hollywood) and 'I smell like a bag of garbage' (KNBC Tonight Show with Jimmy Fallon).

3.3.2.3. Analysis

3.3.2.3.1. Data Filtering

As the data stored in NewsScape originates from TV recordings, some of the cases obtained after the searches may be 'noisy.' For example, although the linguistic expression may be correctly identified, the video sometimes does not show any speaker on screen (a 'voice-over' case). This is the main reason why we needed to filter the data obtained from the different searches and discard the cases that were not suitable for our research aim. First, we deleted the cases in which no speaker appeared talking on screen. We also removed the cases which presented a tagging error (parts of speech were not correctly identified by the system), a technical issue (video/audio was not available), a text-speech mismatch or that were repeated (the utterance appeared in a duplicate video). The last step in order to obtain the group of valid cases was to keep those cases where the speaker's hands were visible and eliminate the cases where the speaker's hands were not visible. Once the group of valid cases (i.e., clips in which speakers could be seen including their hands) was established, we further annotated the cases in which a hand gesture was made. A second coder analyzed a subset of the data to determine if there was agreement in the classification of visible hand gestures and a substantial agreement was reached: $\kappa = 0.7$ for *taste* and $\kappa = 0.77$ for *smell* (Cohen, 1960).

3.3.2.3.2. Gesture Analysis

We classified the visible hand gestures into those semantically related to the linguistic expression and those non-semantically related. A gesture was classified as semantically related when it was apparently coherent with the speakers' spoken message by adding a parallel or supplementary nuance (McNeill, 1992). For example, we considered to be semantically related

the cases where speakers pointed to their mouth, when using *taste*, or their nose, when using *smell*. On the contrary, we classified a gesture as non-semantically related when it did not seem to be associated with the meaning of the speaker's utterance; that is, when it did not depict any semantic information associated with the spoken message. An illustration of this category are beats which are simple gestures that move along with the rhythm of speech (McNeill, 1992).

A second coder analyzed a subset of the data to determine if there was agreement in the semantic classification of visible hand gestures. There was almost perfect agreement ($\kappa = 0.97$) for *taste* and substantial agreement ($\kappa = 0.74$) for *smell* (Cohen, 1960). At this point, we eliminated non-semantically related gestures and the gestures labeled as semantically related that were produced in an 'acted' communicative situation (e.g., music performances). In the end, only the remaining spontaneous hand gestures that were classified as related to the verb were kept in the final dataset.

Once we obtained a list of the cases which contained a semantically related gesture, we annotated multiple features adopting a hybrid typology that merges distinct labels from earlier classifications (Ekman & Friesen, 1969; Cohen, 1977; McNeill, 1992; Bressem, 2013; Alcaraz-Carrión, 2018). The main focus of our analysis was on the formal properties of gestures.

- Function of the gesture (representational[9], emblem, other, combination)
- Hand used
- Handshape (palm, grabbing, pick, index finger, fist, other, combination)

[9] In our initial coding, representational gestures were classified as either deictic, iconic or metaphoric. However, when no meaningful patterns emerged from this distinction, we collapsed these categories into the broader term representational (McNeill, 1992; Alibali & Heath, 2001; Kita et al., 2017), as the boundaries between these types can sometimes be difficult to establish. This was the case in our study, where gestural meaning is understood as a multi-layered phenomenon in which different dimensions are expressed simultaneously (Sweetser, 2023).

- Gesture axis (lateral, vertical, sagittal, punctual[10], other, combination, N/A)
- Gesture direction (rightwards, leftwards, inwards, outwards, up, down, other, combination, N/A)
- Free hands—i.e., was the speaker grabbing anything, such as a microphone?

A second coder analyzed the final dataset with a substantial agreement in the identification of these formal features: κ = 0.78-0.95 for *taste* and κ = 0.63-1 for *smell* (Cohen, 1960).

3.3.2.3.3. Linguistic Analysis

We annotated the meaning of the verb in every valid case (where the speaker's hands were visible) by means of revising the context they were produced in. We created our meaning classification list by looking up both verbs and combining their entries in three accredited dictionaries: Cambridge, Collins and Oxford. All the authors were in charge of corroborating that the meaning classification made was accurate. The meaning options for each perception verb can be seen in Table 14.

	Meaning options for *taste*	**Meaning options for *smell***
Literal	1. Perceive the flavor of a substance by using the tongue 2. Sample the flavor of food/drink by taking it into the mouth 3. Have a specific flavor	1. Perceive/discover a substance by using the nose 2. Have/emit an odor 3. Have/emit an unpleasant odor
Figurative	4. (Have an) experience (of) something 5. Perceive the property of something that is not food-related 6. Sample the property of something that is not food-related 7. Have a specific property that is not food-related	4. Detect/suspect something by means of instinct 5. Give indications or be suggestive of something 6. Seem untrustworthy or bad
Not applicable	8. Ambiguous	7. Ambiguous 8. Idiomatic (e.g., 'smelling like a rose')

Table 14. Meaning classification options for *taste* and *smell*

[10] Punctual gestures pinpoint a point in space rather than being produced in a particular axis (Alcaraz-Carrión, 2018).

We also annotated different aspects of the message uttered. Based on previous research (Argaman, 2009; Carrillo-de-Albornoz & Plaza, 2013; Bolumar Martínez et al., 2024), we decided to note down the use of different linguistic markers, as they have been shown to be frequently associated with emotional attributes in language and this might be fundamental to help identify the sense of a polysemous verb. We paid attention to the use of negation (words or particles that can turn a positive construction into a negative one; Quirk et al., 1972). In contrast with Bolumar Martínez et al. (2024), the category 'intensifiers' (words or phrases that have a heightening or lowering effect on a lexical unit; Quirk et al., 1972) was subdivided into verb modifiers and object quantifiers. We decided to differentiate them because we suspected that the presence of the latter linguistic marker could potentially vary depending on the meaning expressed. Thus, we annotated whether speakers used a verb modifier (word or phrase that adds information to the meaning of the verb) or a quantifier (words such as little, bit, lot) before the object that followed the verb. We did not include the determiner *a*/*an* in our list of object quantifiers (even though it is a cardinal number) as it was already part of our linguistic search. The list of the negation words, verb modifiers and object quantifiers annotated can be found in Table AI.3.

3.3.3. Results

3.3.3.1. Gesture Frequency

3.3.3.1.1. Taste

The *taste* + *a* search gave 2180 matches (51.03%) and the taste + *like the*/*a*/*an* search returned 2092 matches (48.97%), amounting to a total of 4272 cases (100%). First, we discarded 1470 cases in which the speakers were not visible (voice-over cases). From the 2802 remaining cases, we eliminated 1968 cases that were classified as text-speech mismatch, technical issue, tagging error or repeated. Then, we discarded 206 cases, out of the 834 cases left, in which the

hands of the speaker were not visible. The remaining 628 cases were considered valid cases as we could only determine whether speakers were making a hand gesture or not when their hands were visible. Considering that this research focuses on co-speech hand gestures, we identified 376 cases containing a gesture and 252 cases where no gesture was made. In the group of co-speech gestures, we found 335 unrelated gestures and two acted gestures. Thus, only 39 related gestures were kept in the final dataset of *taste* (see Figure 11; for more data curation details see Table AI.4).

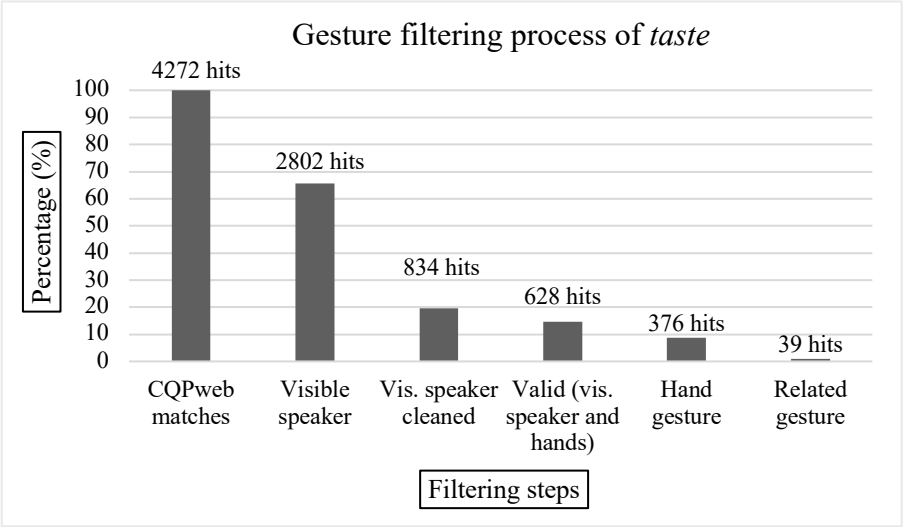

Figure 11. Gesture filtering process of *taste*

The 39 semantically related hand gestures found are mostly associated with the literal meanings of *taste* (see Table 15): four gestures (10.26%) were associated with meaning 1, five gestures (12.82%) were associated with meaning 2 and 24 gestures (61.54%) were associated with meaning 3. Five of the remaining gestures (12.82%) were associated with meaning 7, the only figurative meaning present in the final dataset. The gesture left (2.56%) was classified as ambiguous (meaning 8). In short, out of the 39 semantically related gestures, 33 (84.62%) were made while conveying a literal meaning of *taste* and 5 (12.82%) were made while expressing one of its figurative meanings.

	Semantically related co-speech hand gesture	Non-semantically related co-speech hand gesture	No co-speech hand gesture	Total
Literal	33 cases (5.27%)	276 cases (44.09%)	195 cases (31.15%)	504 cases (80.51%)
Figurative	5 cases (0.80%)	58 cases (9.27%)	56 cases (8.95%)	119 cases (19.01%)
Not applicable	1 case (0.16%)	1 case (0.16%)	1 case (0.16%)	3 cases (0.48%)
Total	39 cases (6.23%)	335 cases (53.51%)	252 cases (40.26%)	626 cases (100%)

Table 15. Distribution of the valid cases of *taste* (without two 'acted' gestures)

We performed a chi-square test of independence to examine the relation between the presence and absence of semantically related gestures and the meanings of *taste* (literal or figurative). The relation between these variables was not significant, χ^2 (1, N = 372) = 0.4293, p = .51. Thus, there was no difference in the distribution of semantically and non-semantically related gestures when co-occurring with the meanings of *taste*. However, it is important to note that the reduced sample size in some of the cells might limit the statistical power of the test. Fisher's exact test was not used because none of the expected cell counts fell below the conventional threshold of 5.

3.3.3.1.2. Smell

The *smell* + *a* search returned 2486 matches (58.62%) and the *smell* + *like* + determiner search produced 1755 matches (41.38%), amounting to a total of 4241 cases (100%). First, we eliminated 1658 cases in which the speakers were not visible (voice-over cases). From the 2583 remaining cases, we discarded 1030 cases that were classified as text-speech mismatch, technical issue, tagging error or repeated. After this, we removed 591 cases, out of the 1553 cases left of 4241, in which the hands of the speaker were not visible. The remaining 962 cases were considered valid cases as we could only determine whether speakers were making a hand gesture or not when their hands were visible. When categorizing the group of valid cases, we identified 538 cases containing a hand gesture and 424 cases where no hand gesture was made.

In the former group, we found 456 unrelated gestures and 11 acted gestures. Finally, 71 related gestures were kept in the final dataset of *smell* (see Figure 12; for more data curation details see Table AI.5).

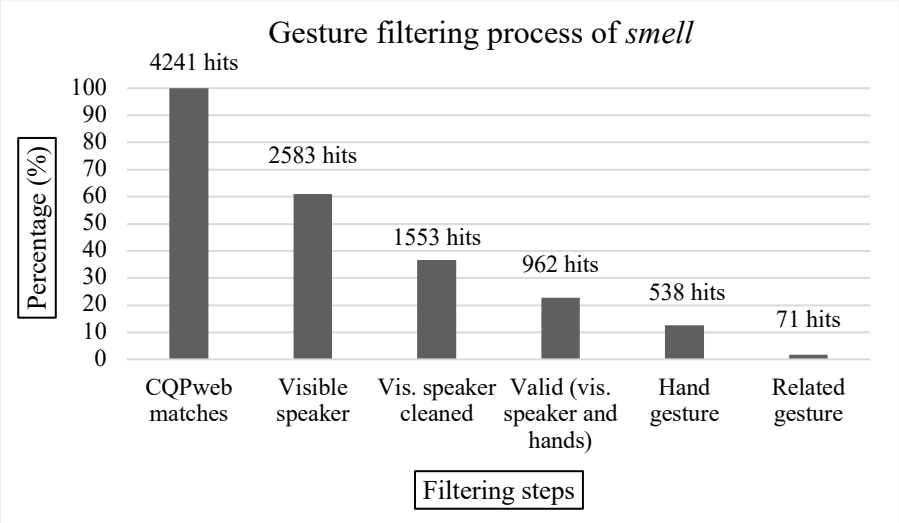

Figure 12. Gesture filtering process of *smell*

The 71 semantically related hand gestures found are mainly associated with the literal meanings of *smell* (see Table 16): 25 gestures (35.21%) were associated with meaning 1, 33 gestures (46.48%) were associated with meaning 2 and one gesture (1.41%) was associated with meaning 3. Regarding the figurative meanings of *smell*, eight gestures (11.27%) were associated with meaning 4 and three gestures (4.23%) were associated with meaning 5. The only gesture left (1.41%) was classified as idiomatic (meaning 8). None of the gestures in the final dataset were related to meanings 6 or 7. In brief, out of the 71 semantically related gestures, 59 (83.10%) were made while conveying a literal meaning of *smell* and 11 (15.49%) were made while expressing a figurative meaning.

	Semantically related co-speech hand gesture	Non-semantically related co-speech hand gesture	No co-speech hand gesture	Total
Literal	59 cases (6.20%)	276 cases (29.02%)	336 cases (35.33%)	671 cases (70.56%)
Figurative	11 cases (1.16%)	162 cases (17.03%)	74 cases (7.78%)	247 cases (25.97%)
Not applicable	1 case (0.11%)	18 cases (1.89%)	14 cases (1.47%)	33 cases (3.47%)
Total	71 cases (7.47%)	456 cases (47.95%)	424 cases (44.58%)	951 cases (100%)

Table 16. Distribution of the valid cases of *smell* (without 11 'acted' gestures)

Here, we also performed a chi-square test of independence to examine the relation between the presence and absence of semantically related gestures and the meanings of *smell* (literal or figurative). The relation between these variables was significant, χ^2 (1, N = 508) = 12.1608, p < .001. Therefore, the distribution of semantically and non-semantically related gestures varies when co-occurring with the meanings of *smell*. Speakers appear to be more likely to make a semantically related gesture when the sense of the verb is literal than when it is figurative.

3.3.3.2. Gesture Features

3.3.3.2.1. Taste

As only 39 gestures were considered to be semantically related to *taste*, the gesture analysis of this verb is not as extensive as initially expected. Based on previous findings with the verb *touch*, we expected gestures to differ depending on whether *taste* was used literally or figuratively. Specifically, we anticipated that more action-oriented gestures, such as speakers bringing their hands to their mouths, would accompany literal uses, while more abstract gestures, such as those conveying evaluative judgments, would appear with figurative uses. Although a variety of semantically related gestures was found (e.g., mouth pointing, eating mimicry), no clear gesture patterns emerged that varied depending on the meaning of the verb. In other words, speakers did not make different gestures when conveying the literal and figurative meanings of the verb *taste*.

However, two formal features stood out in the final gesture dataset. The characteristic that most semantically related gestures had in common was their shape: 31 gestures were performed using a pick shape (i.e., the thumb touches the index finger and, sometimes, the rest of the fingers). Within this group of pick shape gestures, another recurring feature was observed: most speakers rubbed their fingers together (23 out of the 31 pick shape gestures). The defining aspect of these *rubbing* gestures is that the speaker's thumb touches the fingertips of one or more of the other fingers, as if about to 'make a silent snap' (typically involving the index and middle fingers; see Figure 13 for an example).

Figure 13. Example of a rubbing gesture with *taste* (https://tinyurl.com/exampletaste)

3.3.3.2.2. Smell

Among the 71 gestures that were considered to be semantically related to *smell*, we were able to observe a wide range of co-speech gestures (e.g., nose pointing, grabbing the 'smelled element'). We initially expected gestures to differ depending on whether *smell* was used literally or figuratively. As with *taste*, our original hypothesis was that literal uses would be accompanied by action-oriented gestures, such as speakers bringing their hands to their noses while sniffing, while figurative uses might evoke more abstract gestures reflecting internal states or evaluative judgments. In our limited sample, however, speakers made similar gestures when expressing the literal and figurative meanings of the verb *smell*. After observing that this hypothesis was not confirmed for *smell*, we considered the possibility that gesture might instead be used to distinguish between meanings related to odor perception versus odor

emission. Despite this, no clear gesture pattern associated with the specific meanings of *smell* was identified.

Most gestures were classified as representational gestures made with a palm shape. The majority of the gestures were used by speakers to indicate the area where they believed the 'smelled element' was located. As shown in Figure 14, speakers employed this deictic gesture strategy not only to indicate the location of 'physical' substances (e.g., perfume) but also to represent 'figuratively perceived' objects (e.g., 'an inside job'). These results show that, when using the verb *smell*, speakers point both to concrete elements and to elements that, strictly speaking, are not physically present. When abstract deixis is used in our data, speakers seem to point to empty locations to invest them with meaning (Stukenbrock, 2014). In short, the deictic gestures in our data are primarily used to anchor words to the world and to build common ground, illustrating how fundamental deixis is for speakers in referring to different types of ideas (Cooperrider & Mesh, 2022). Another similar gesture strategy involved tracing the trajectory of the 'smelled element' as perceived by the speaker.

Figure 14. Example of a representational (deictic) gesture with *smell*

(https://tinyurl.com/examplesmell)

Rubbing gestures were observed in the *smell* dataset as well, although they were much less common: only two cases were identified, one of which was performed with a pick shape.

3.3.3.3. Linguistic Features

3.3.3.3.1. Taste

With respect to the use of specific linguistic markers with semantically related gestures, negation (which played a significant role with *touch*; see Bolumar Martínez et al., 2024) does not seem to be as relevant concerning the verb *taste*. Only two literal cases (5.13%) in which negation was used were found:

(1) 'The rye doesn't taste like the rye that you see in the rye bread'

(2) 'This doesn't taste like the waxy chocolate on top of pretzels'

If we observe the use of negation among the remaining groups of valid cases, the proportion does not increase very much. Only 23 cases (6.87%) out of the 335 non-semantically related gestures found included the use of negation. Similarly, in the group of no co-speech hand gestures there were only 14 cases (5.56%) out of the 252 were speakers used a negation word. Overall, negation is mostly used with the literal meanings of *taste* (see Table 17).

	Negation (literal)	**Negation (figurative)**	Total
Semantically related co-speech hand gesture	**2 (5.13%)**	**-**	39 (100%)
Non-semantically related co-speech hand gesture	**22 (6.57%)**	**1 (0.30%)**	335 (100%)
No co-speech hand gesture	**13 (5.16%)**	**1 (0.40%)**	252 (100%)
Total (valid cases)	37 (5.91%)	2 (0.32%)	626 (100%)

Table 17. Distribution of negation words among valid cases of *taste* (without 'acted' gestures)

In relation to the use of verb modifiers such as 'really,' speakers do not appear to use many when making a semantically related gesture. Thus, this linguistic marker does not appear to be as relevant as expected. Only three literal cases (7.69%) in which verb modifiers were used were found:

(3) ‘Your beverages actually taste a little bit better than regular ice’

(4) ‘I can just taste a hint’

(5) ‘It actually tastes a little bit like... like kidney or liver’

If we examine the use of verb modifiers among the remaining groups, the proportion shows little to no increase. Only 31 cases (9.25%) out of the 335 non-semantically related gestures found included the use of a verb modifier. Similarly, in the group of no co-speech hand gestures, there were only 19 cases (7.54%) out of the 252 where speakers used this type of linguistic marker. In short, verb modifiers are primarily used when speakers convey the literal meanings of *taste* (see Table 18).

	Vb. modifier (literal)	**Vb. modifier (figurative)**	Total
Semantically related co-speech hand gesture	**3 (7.69%)**	**-**	39 (100%)
Non-semantically related co-speech hand gesture	**28 (8.36%)**	**3 (0.90%)**	335 (100%)
No co-speech hand gesture	**16 (6.35%)**	**3 (1.19%)**	252 (100%)
Total (valid cases)	47 (7.51%)	6 (0.96%)	626 (100%)

Table 18. Distribution of verb modifiers among valid cases of *taste* (without ‘acted’ gestures)

The remaining linguistic marker we focus on appears to be the most relevant regarding *taste*. The use of quantifiers before the sentence object attains a higher proportion in the group of semantically related gestures; 19 literal cases (48.72%) and just one figurative case (2.56%) in which quantifiers were used. Some of the object quantifiers were:

(6) ‘They taste a bit stale’

(7) ‘It tastes a little grainy’

When it comes to the other groups, 109 cases (32.54%) out of the group of 335 non-semantically related gestures included the use of an object quantifier. In the group of 252 no

co-speech hand gestures, speakers used an object quantifier in 75 cases (29.76%). Overall, the use of object quantifiers appears to be prominent when expressing the literal meanings of *taste* (see Table 19).

	Ob. quantifier (literal)	**Ob. quantifier (figurative)**	Total
Semantically related co-speech hand gesture	**19 (48.72%)**	**1 (2.56%)**	39 (100%)
Non-semantically related co-speech hand gesture	**97 (28.96%)**	**12 (3.58%)**	335 (100%)
No co-speech hand gesture	**69 (27.38%)**	**6 (2.38%)**	252 (100%)
Total (valid cases)	185 (29.55%)	19 (3.04%)	626 (100%)

Table 19. Distribution of object quantifiers among valid cases of *taste* (without 'acted' gestures)

3.3.3.3.2. Smell

Contrary to the findings for *touch* (Bolumar Martínez et al., 2024), the use of negation together with semantically related gestures does not appear to be relevant in relation to *smell*. In fact, no case of negation was found in the group of semantically related gestures.

If we examine the use of negation among the remaining groups, the proportion does not increase noticeably. Only 17 cases (3.73%) out of the group of 456 non-semantically related gestures included the use of negation. In the group of 424 no co-speech hand gestures, there were only 12 cases (2.83%) where speakers used negation. If we focus on the different meaning groups, negation is mostly used with the literal meanings of *smell* (see Table 20).

	Negation (literal)	**Negation (figurative)**	Total
Semantically related co-speech hand gesture	**-**	**-**	71 (100%)
Non-semantically related co-speech hand gesture	**16 (3.51%)**	**1 (0.22%)**	456 (100%)
No co-speech hand gesture	**10 (2.36%)**	**2 (0.47%)**	424 (100%)
Total (valid cases)	26 (2.73%)	3 (0.32%)	951 (100%)

Table 20. Distribution of negation words among valid cases of *smell* (without 'acted' gestures)

Considering the use of verb modifiers, speakers do not seem to use many when making a semantically related gesture. Thus, again this linguistic marker does not appear to be a crucial signal in pointing at different verb meanings. The final dataset only contains three literal cases (4.23%) in which verb modifiers were used:

(8) 'He just smells like a champion'

(9) 'It sort of smells like the south of France over here'

(10) 'If you were to crack a piece of this off, you would actually smell a piece of gas'

If we observe the use of verb modifiers among the remaining groups, the proportion does not increase. Only nine cases (1.97%) out of the 456 non-semantically related gestures found included the use of a verb modifier. Out of 424 no co-speech hand gestures, there were only four cases (0.94%) were speakers used this type of linguistic marker. Overall, verb modifiers are mostly used with the literal meanings of *smell* (see Table 21).

	Vb. modifier (literal)	**Vb. modifier (figurative)**	Total
Semantically related co-speech hand gesture	**3 (4.23%)**	**-**	71 (100%)
Non-semantically related co-speech hand gesture	**7 (1.54%)**	**2 (0.44%)**	456 (100%)
No co-speech hand gesture	**4 (0.94%)**	**-**	424 (100%)
Total (valid cases)	14 (1.47%)	2 (0.21%)	951 (100%)

Table 21. Distribution of verb modifiers among valid cases of *smell* (without 'acted' gestures)

The remaining linguistic marker annotated once more appears to be the most important regarding *smell*. There were 16 literal cases (22.54%) in the group of semantically related gestures in which object quantifiers were used. Some of the object quantifiers were:

(11) 'We smelled a little bit of smoke'

(12) 'She smelled a lot like scotch'

Out of the 456 non-semantically related gestures, 85 cases (18.64%) included the use of a quantifier. In the group of 424 no co-speech hand gestures, speakers used a quantifier in 67 cases (15.80%). In short, the use of object quantifiers appears to be prominent when expressing the literal meanings of *smell* (see Table 22).

	Ob. quantifier (literal)	**Ob. quantifier (figurative)**	Total
Semantically related co-speech hand gesture	**16 (22.54%)**	**-**	71 (100%)
Non-semantically related co-speech hand gesture	**75 (16.45%)**	**10 (2.19%)**	456 (100%)
No co-speech hand gesture	**64 (15.09%)**	**3 (0.71%)**	424 (100%)
Total (valid cases)	155 (16.30%)	13 (1.37%)	951 (100%)

Table 22. Distribution of object quantifiers among valid cases of *smell* (without 'acted' gestures)

3.3.4. Discussion

3.3.4.1. Gesture Distribution in Touch, Taste and Smell

The distribution of the three different groups among the valid cases is quite similar in the analysis of *taste* and *smell*, but not of *touch* (see Table 23). Although the overall distribution of gesture vs. no gesture is comparable among all verbs, the difference is prominent when focusing on the distribution of the two gesture groups (semantically related and non-semantically related). While the proportion of gestures semantically related to *touch* attained 41.60% of the valid cases (Bolumar Martínez et al., 2024), *taste* and *smell* semantically related gestures reach 6.23% and 7.47%, respectively[11]. Thus, a higher proportion of semantically related gestures is found co-occurring with *touch*.

	GESTURE		**NO GESTURE**	
	Semantically related co-speech hand gesture	**Non-semantically related co-speech hand gesture**	**No co-speech hand gesture**	Total
Touch	302 cases (**41.60%**)	159 cases (**21.90%**)	265 cases (**36.50**%)	726 cases (100%)
	461 cases (63.50%)			
Taste	39 cases (**6.23%**)	335 cases (**53.51%**)	252 cases (**40.26%**)	626 cases (100%)
	374 cases (59.74%)			
Smell	71 cases (**7.47%**)	456 cases (**47.95%**)	424 cases (**44.58%**)	951 cases (100%)
	527 cases (55.42%)			

Table 23. Distribution of valid cases in the analysis of *touch*, *taste* and *smell* (without 'acted' gestures)

This notable difference could be attributed to variations in the linguistic searches. For *touch*, personal pronouns were specifically included, enabling the identification of both literal (physical) and figurative (emotional) meanings of the verb. In contrast, the linguistic search for *taste* and *smell* aimed to capture all possible uses of the verbs without restricting the object

[11] We checked whether the groups within the valid cases (semantically related gesture, non-semantically related gesture and no gesture) were equally distributed among the perception verbs *touch*, *taste* and *smell*. The result of the chi-square test was significant, χ^2 (4, N = 2303) = 439.1591, p = <.001.

being perceived or involved in the action. Nevertheless, it seems unlikely that these variations in the linguistic search fully account for the difference observed with *touch*.

Instead, we believe the higher proportion of semantically related gestures for *touch* is motivated by the notable difference between the meanings investigated. The meanings of *touch* were both associated with 'personal' experiences: one referred to a physical experience, which could range from positive to highly negative depending on the context, while the other referred to an emotional reaction, which could involve joy or sorrow. The need to distinguish between these contrasting meanings may explain why speakers are more inclined to use gestures when employing *touch* compared to *taste* and *smell*.

Another factor might be that the figurative meaning of *touch* can be easily conveyed through gesture. Unlike *taste* and *smell*, the figurative meaning of *touch* is frequently used by the speaker as a receiver of the action (Bolumar Martínez et al., 2024). Even though the emotional meaning of *touch* is abstract and metaphorically motivated, when speakers use the personal pronoun *me*, they can become the referent of the gesture through a 'whole-for-part' metonymic strategy—typically gesturing toward their chest. Thus, the emotional meaning of *touch* is embodied through chest-touching gestures, which are motivated by the container metaphor THE HEART IS A CONTAINER FOR EMOTIONS (Kövecses, 1986; Lakoff & Johnson, 1980). Interestingly, the source domain of this metaphor is not related to the sense of touch. This appears to suggest that a non-perception metaphor may play a more prominent role in gesture production than a broader perception metaphor such as AFFECTING IS TOUCHING (Ibarretxe-Antuñano, 1999) which encompasses both the physical and the emotional meanings of the verb *touch*.

3.3.4.2. Gesture Helps Distinguish the Meanings of Touch, but not Taste or Smell

Regarding the first objective of this paper (i.e., to compare the co-speech gestures associated with the meanings of *taste* and *smell*), we could conclude that the semantically related gestures

found with literal meanings of *taste* and *smell* were also found when observing the figurative semantic extensions of these verbs. Expressed differently, the literal and figurative meanings of these verbs cannot be distinguished solely by observing the hand gestures speakers make.

Although no specific gesture patterns were exclusively tied to the different meanings of *taste* and *smell*, most semantically related gestures were somewhat iconic. Speakers often represented the perceptual action, such as moving their hands toward their mouth or simulating 'moving air' toward their nose. However, gesture referents do not appear to be as relevant to distinguish between the meanings of these two verbs as in the case of *touch*. One possible explanation is the nature of the perceptual actions these verbs convey. For *taste* and *smell*, gesture referents may be less important because these actions do not require a visually present object (even less so in figurative meanings). In the prototypical meanings of *taste* and *smell*, the perceived element is typically an 'invisible' chemical substance in food or the surrounding air, such as bitterness or mustiness. While speakers can easily point to food they are tasting or locate the source of a smell, this strategy is less effective when referring to abstract notions in figurative meanings. This was further illustrated by the observation that, with *smell*, speakers seemed more likely to make a semantically related gesture when expressing a literal meaning.

In contrast, the verb *touch* inherently involves a physical element in the action, which can usually be pointed to or touched, when present. As discussed above, this extends to the figurative use of *touch*, where a 'whole-for-part' metonymic strategy is employed. Although with the verb *touch* we also found similar gestures in the literal and figurative meanings, it is true that we were able to distinguish different gesture patterns depending on the meaning of *touch* speakers conveyed. Here, the differences were evident when looking at the gesture referents: when the meaning was physical, they were likely to reach an external referent and when it was emotional, they were likely to touch their own chest. As previously suggested, we

believe the reason behind this distinction lies in the need for speakers to differentiate between the contrasting physical (literal) and emotional (figurative) meanings of the verb *touch*.

In relation to the second objective of this work (i.e., to determine the motivation behind frequent co-speech gestures associated with *taste* and *smell*), we can only discuss the potential motivations of two frequent gestures co-occurring with *taste*: pick shape gestures and rubbing gestures. Despite the fact that we analyzed the most frequent semantically related gestures of *smell*, no clear pattern emerged among them.

After revising all the pick shape gestures, we observed that when this pick shape was used 'alone' (without rubbing fingers), it appeared to be mostly related to the action of taking something and bringing it towards one's mouth. Thus, we believe that the motivation behind pick shape gestures is their iconic resemblance to the handshape people use when grabbing food.

It should be noted that although rubbing gestures were frequently used when uttering the verb *taste*, they also appeared when speakers used *smell*. At first, we believed that the rubbing gesture was somewhat related to the domain of TASTE, as speakers appeared to raise their hands to make the gesture more visible, with some even bringing their hands closer to their mouths. For this reason, the rubbing gestures observed were considered intentionally communicative and semantically related to *taste*. However, it seems more likely that rubbing gestures are particularly associated with the difficulty of describing an experience perceived. After analyzing the speakers' utterances, we noticed that these gestures were often produced when speakers struggled to describe the flavor of a food product. Elements indicating vacillation included interjections, repetitions and pauses. Taking this into consideration, rubbing gestures seem to be used when speakers find it challenging to articulate an accurate descriptor for their spoken message.

In order to explore the potential motivations of the gestures co-occurring with the verb *smell*, we examined whether there were any tendencies in the direction of gestures that distinguished the perception of smell from the emission of smell. Most gestures were directed inwards, even when the meaning of *smell* expressed the notion of emission. However, we ultimately discarded this analysis due to the high variability of gestures in the final dataset, which made consistent patterns difficult to identify. Although no motivation could be analyzed in detail, we were able to observe how specific semantic components of *smell* are represented through co-speech gestures.

If we compare the motivation behind the most frequent gesture made with *touch* (the chest-touching gesture) with the motivations behind pick shape and rubbing gestures, it is clear that the sources of these motivations are very different. While *chest-touching* gestures are believed by Bolumar Martínez et al. (2024) to be based on the conceptual metaphor THE HEART IS A CONTAINER FOR EMOTIONS (Kövecses, 1986; Lakoff & Johnson, 1980), the motivation behind pick shape gestures seems to be primarily based on an iconic representation of the action of tasting. The exact motivation behind rubbing gestures cannot be determined, as in this study it is unclear whether this peculiar gesture represents semantic features of *taste* and/or *smell* or if it serves as a discourse-related element used when speakers struggle to find an accurate descriptor.

3.3.4.3. Linguistic Markers Help Distinguish the Meanings of Touch, Taste and Smell

Concerning the third objective of this research (i.e., to describe the relation between the linguistic context and the literal and figurative meanings of *taste* and *smell*), only the use of object quantifiers attained a substantial proportion in both verbs. Although we also examined the use of negation and verb modifiers, these two types of linguistic markers were not

commonly used in the valid cases for either *taste* or *smell*. It should be noted that, due to this and the extremely low proportion of cases where speakers expressed figurative meanings of *taste* and *smell*, no model could be calculated.

With regards to object quantifiers, their use appears to be more prominent when expressing the literal meanings of *taste* and *smell*. However, no relation between their use and the particular meanings of the verbs could be identified. The association of object quantifiers with the literal meanings of both verbs should be interpreted with caution due to the low frequency of figurative meanings found in this study.

The use of object quantifiers with *taste* and *smell* suggests that both the sense of taste and the sense of smell are conceptualized along a gradable scale. This means that the perceived substance, flavor or odor, might change depending on the perceiving person and environmental factors, such as the distance to the substance source or the location where the experience occurs. In short, the subjectivity and the 'chemical nature' of the senses of taste and smell are reflected in the high proportion of object quantifiers. Furthermore, speakers often describe flavors and odors saying that a substance 'tastes/smells a bit/a little/a lot/very' + adjective. This illustrates how, through tasting or smelling, speakers can detect different components and even determine the intensity of the substance perceived. For example, the sentence 'it's milk that tastes a little sour' conveys the idea that, even though milk is expected to be slightly sweet, in this case, the sour flavor exceeds the expected level of perceived sweetness.

Unlike *taste* and *smell*, a very low proportion of object quantifiers was used with the literal (physical) meaning of *touch*. The analysis of *touch* showed that speakers used intensifiers (verb modifiers and object quantifiers) primarily when expressing the figurative (emotional) meaning of *touch*. It should be noted that regarding the distribution of intensifiers found in the analysis of *touch*, most were verb modifiers.

The use of verb modifiers and object quantifiers with *touch*, *taste* and *smell* reveals how the same experience can be perceived with different degrees of intensity. Although the frequency of the linguistic markers used with *touch* differs in comparison with the other two perception verbs, it still shows that the emotional (figurative) meaning of *touch* is also perceived along a gradable scale, likely because emotions cannot be controlled by an external being (Bolumar Martínez et al., 2024). Contrary to the analyses of *taste* and *smell* which show that the 'gradable conceptualization' is primarily present in literal meanings.

Overall, the distribution of object quantifiers highlighted the meanings that were conceptualized along a gradable scale. In the case of *taste* and *smell*, this linguistic marker was predominantly used with their literal meanings. The analysis of the verb *touch* revealed that object quantifiers were mainly used when speakers conveyed the figurative (emotional) meaning of the verb.

3.3.5. Conclusion

This paper explored whether the meanings of the polysemous verbs *taste* and *smell* could be differentiated through co-speech hand gestures and linguistic markers, extending the findings of Bolumar Martínez et al. (2024) for the verb *touch*. In short, we investigated whether the patterns observed for *touch* generalize to additional perception verbs or if such patterns are verb-specific.

The gesture analysis reveals that, while gesture plays a relevant role in differentiating the meanings of *touch*, it does not exhibit different patterns between the literal and figurative meanings of *taste* and *smell*. Although two gestures, the pick shape gesture and the rubbing gesture, often co-occurred with *taste*, they did not systematically correspond to specific meanings.

Considering that only object quantifiers were frequently associated with the literal meanings of both *taste* and *smell*, the analysis of linguistic markers indicates that these verbal features provide more discriminating information regarding the meanings of *touch*. Although the use of negation and verb modifiers did not show a meaningful distribution across meanings for either verb in this study, the frequent use of object quantifiers aligns with the conceptualization of the senses of taste and smell as gradable, where the perceived intensity of flavors and odors can vary depending on individual and external factors.

The main limitation of this work is the low frequency of semantically related gestures observed when speakers convey figurative meanings of *taste* and *smell*, making it difficult to draw strong conclusions about such gestures. However, this could also be considered a contribution, as the use of a large multimodal dataset allowed the identification of low-frequency gestures that might have otherwise been missed, such as the rubbing gesture.

Although a big data approach was used to collect gesture data, further investigation is needed to confirm our hypothesis that speakers use rubbing gestures when they experience difficulty finding verbal descriptors for experiences perceived. Additionally, future research could explore different perception verbs with well-established figurative uses, such as *see* and *hear*, to better identify which communicative modalities might help distinguish the meanings of a polysemous word.

Overall, while the findings of this paper suggest that gesture may not play a primary role in distinguishing meanings of *taste* and *smell*, the prominence of object quantifiers when expressing literal meanings of both verbs could be considered a shared meaning-differentiating pattern. These results underscore the value of multimodal data in polysemy research, as they show how the role of co-speech gestures and linguistic markers varies across different perception verbs.

Data availability statement. Data and supplementary materials are publicly available in the following link: https://osf.io/jna28/overview.

Funding statement. The research reported here was supported by a Fulbright scholarship to I. B. M. (00001/FLB/22. Fundación Séneca. Región de Murcia-Fulbright); by the 21596/FPI/21 Fundación Séneca, Región de Murcia (Spain), predoctoral fellowship to I. B. M.; by the European Union – NextGenerationEU (Beatriz Galindo award) to D. A. C.; and by the Ministerio de Ciencia e Innovación (Spanish State Research Agency) and Feder/UE funds (grant number PID2022-139194NB-I0).

4. Conclusion

4.1. General Conclusions

This thesis has explored the role of co-speech gestures and linguistic markers when distinguishing the meanings of the polysemous perception verbs *touch*, *taste* and *smell*. By adopting a multimodal approach to the study of polysemy, this is the first investigation that bridges the gap between lexical analyses of polysemy and multimodal approaches to the study of meaning, addressing an area that remains largely unexplored. This investigation focused on examining how gestures and speech content help both speakers and interlocutors manage polysemous expressions. In particular, the first research question investigated—in the first and third papers—whether speakers distinguish between the meanings of polysemous expressions in speech and gesture. The second research question examined—in the second paper—whether observers disambiguate the meaning of unpredictable polysemous expressions through gesture.

The key findings of this thesis show that the role of gesture and linguistic markers in speech varies across three distinct but comparable polysemous lexical items, highlighting the value of incorporating multimodal data into polysemy research. The first paper demonstrated 1) that speakers produce different gesture patterns when expressing the physical vs. emotional meanings of the verb (i.e., other-touch and self-touch respectively) and 2) that they use more negation words with the physical meaning but more intensifiers with the emotional meaning. The second paper helped determine how interlocutors use gesture information (specifically, location and handshape) to disambiguate the physical and emotional meanings of *touch* in unpredictable expressions. The third paper showed 1) that although some notable hand gestures (e.g., rubbing gestures) were identified, no specific gesture patterns were consistently associated with specific meanings of the verbs and 2) that while the use of object quantifiers co-occurring with *taste* and *smell* was noticeable, no linguistic markers in speech were linked to specific meanings of these verbs.

Overall, the corpus-based analyses of this thesis indicate that gesture and linguistic markers in speech seem particularly informative in the case of *touch*. The experimental part of this investigation further supports findings from previous lexical ambiguity studies (Holler & Beattie, 2003; Holle & Gunter, 2007; Hinnell & Parrill, 2020), showing that interlocutors look for meaning in gestures co-occurring with ambiguous expressions. In the case of *touch*, this appears to be possible because the gesture patterns identified (especially, chest-touching gestures) reveal aspects of the speaker's conceptualization that are not accessible through lexical information in speech.

The differing results for *touch* compared to *taste* and *smell* make it challenging to integrate these findings within a single theoretical framework in polysemy research. Considering the diverse multimodal manifestations of the three perception verbs—and the fact that similar gestures were observed with both prototypical and non-prototypical meanings of *taste* and *smell*—the sense network approach to polysemy appears best suited to integrate the results of this thesis. This framework, which proposes that the meanings of polysemous words are interconnected through conceptual associations, can also accommodate a multimodal approach to the study of polysemy, allowing for gesture 'forms' associated with either one or several related meanings. Since this thesis introduces a novel perspective on the multimodal expression of polysemous lexical items, further investigation across different polysemous domains is needed to confirm which theoretical approach to polysemy can most effectively incorporate multimodal manifestations.

The following points of this section outline the specific conclusions regarding the two main research questions addressed in this thesis.

4.1.1. Speakers Distinguish Between the Meanings of Touch, but not Taste or Smell

To address the first research question—whether speakers distinguish between the different meanings of polysemous expressions in speech and gesture—a total of 12782 clips featuring speakers uttering the verbs *touch*, *taste* or *smell* were collected.

After researching potential strategies used in speech and gesture by English speakers to distinguish between the meanings of these three perception verbs, distinct patterns were identified in each analysis. However, the patterns detected in speech and gesture did not, at a general level, confirm the initial hypothesis that gesture, together with linguistic markers, plays a relevant role in distinguishing the meanings of polysemous words. This hypothesis only appears to hold true for the verb *touch*.

The first clear conclusion regarding the verb *touch*, in comparison with *taste* and *smell*, is that negation words occur more frequently. When looking at the linguistic markers found with the three perception verbs (see Tables AI.1 and AI.3), the variability of markers modifying speech content is fairly comparable. Thus, this difference seems to arise from semantic components specific to the physical meaning of *touch*. Contrary to our initial expectation that negation would help predict the emotional meaning—based on previous findings (Carrillo-de-Albornoz & Plaza, 2013)—it appears that, as discussed in the first paper, the 'invasive' component associated with the physical meaning of *touch* accounts for the greater use of negation with this verb.

The second conclusion is that gestures related to the verb *touch* are produced more frequently. This may stem from the need to distinguish the contrasting meanings selected to study polysemy in *touch*. Another potential contributing factor is the ease with which the emotional meaning of *touch* can be represented through chest-touching gestures. In comparison with *taste*

and *smell*, the figurative meaning of the verb *touch* conceptualizes the speaker as the receiver of the action.

A related explanation for this higher gesture frequency concerns the nature of chest-touching gestures themselves. These gestures do not occur entirely in parallel with spoken words, as they express culture-dependent aspects of the emotional meaning conveyed. Through this gesture, which embodies the speaker's emotions, the Western conceptualization of the heart as the organ containing emotions is revealed. Recent research has identified this gesture[12] as a physiologically motivated—warmth in the chest—nonverbal signal of 'heartfelt' connection to others and as an emblem of empathy (Farley et al., 2021). However, as will be discussed in the last part of this section, the use and interpretation of the chest-touching gesture require further examination across cultures, communities and language modalities.

4.1.2. Observers Disambiguate the Meaning of Polysemous Expressions through Gesture

To address the second research question—whether observers disambiguate the meaning of polysemous expressions through gesture—an experimental approach to the study of *touch* was developed by adapting naturalistic data and creating a specific set of stimuli.

First, a subset of the hand gestures co-occurring with *touch* (from the first paper) was used in a paradigm designed to isolate the central visual component in a multimodal analysis. By separating facial expressions from hand gestures, it was possible to identify which body part exerted a stronger effect on observers' interpretation. Using this method, hand gestures were

[12] Previous studies have called this gesture as the 'heartfelt' or 'hand-over-heart' gesture (e.g., Farley et al., 2021; Garrah et al., 2024). The term chest-touching is adopted here as a more accurate label, reflecting both prior descriptions and the data analyzed in this thesis. While the exact hand placement may vary, it is most frequently centered on the speaker's chest; thus, since the gesture is not restricted to the left side, emphasis on the heart is unnecessary.

found to convey key semantic information, as observers' accuracy increased when these gestures were available.

Once observers' interpretation accuracy was confirmed to be higher when hand gestures were present, a new dataset was created in which English speakers performed controlled gestures while uttering the ambiguous phrase 'you touched me.' These stimuli were first used to identify the specific gesture components influencing observers' interpretation. An interaction between gesture location and handshape was found when analyzing the observers' interpretation of the stimuli. Regarding the specific gestures associated with the physical meaning of *touch*, they involved an index finger handshape reaching either the speaker's stomach or the interlocutor's arm. The strongest association with the emotional meaning, by contrast, was found for gestures involving contact with the speaker's chest, performed with either an open palm or a fingertips handshape.

After identifying the gesture components influencing observers' interpretation, the same dataset of controlled gestures was used to evaluate the effect of hand gestures on reaction time when interpreting the meaning *touch*. Gesture was shown to facilitate faster decision-making when participants interpreted an unpredictable case of polysemy ('you touched me'). Since the effect persisted even when gestures were completely isolated from speech, it appears that the presence of gesture itself contributed to faster responses.

Collectively, the experimental research on the verb *touch* revealed how interlocutors use gesture information—specifically, location and handshape—to interpret the physical and emotional meanings of *touch*. This approach showed that information conveyed through the modality of gesture can complement a decontextualized verbal expression and, ultimately, that hand gestures can help in the disambiguation of an expression whose meaning cannot be inferred from the surrounding words alone.

The combination of experimental paradigms developed in this thesis has broader applicability, as it can be used in other gesture studies addressing different research questions. The paradigms designed to answer the second research question allow for the precise identification of relevant sources of information. Stimuli that show only specific body parts—such as hands without the speaker's head, which might otherwise provide cues through facial expressions or lip reading—have proven highly effective in isolating the variables involved in the communicative process. The same applies to stimuli that depict communicative events solely through gesture and without speech. These approaches, which involve the manipulation of gesture features, not only provide a robust means of identifying the *locus* of formal components in human communication but also represent a transferable methodology for future gesture research.

4.2. Limitations and Future Directions

While the findings of this thesis are relevant, several limitations should be acknowledged. Although a comparable number of hits was obtained for each perception verb, a notable limitation lies in the smaller co-speech gesture dataset for *taste* and *smell* (particularly the limited gesture dataset for *taste*), which are less than half the size of the *touch* gesture dataset. This requires interpreting the findings related to *taste* and *smell* with slightly greater caution. Another limitation that could partially affect the comparison among verbs concerns the range of meanings examined. While only two distinct meanings were considered for *touch*, seven and six distinct meanings were analyzed for *taste* and *smell*, respectively. Furthermore, given the nature of the audiovisual data, it should be noted that the wide range of camera shots in TV recordings prevents the coding of all gestures. In other words, any potential gestures occurring outside a closed camera shot cannot be accounted for. Finally, a major limitation is that, at the time of this thesis submission (November 2025), the NewsScape Edge search engine ceased to function. Although snippets of the audiovisual data remain available, the replicability of the searches has been partially affected.

In addition to these considerations, this thesis leaves several questions open for future research. The need for further investigation of ambiguous and polysemous expressions—beyond perception verbs—to strengthen the empirical basis of this thesis is evident. Several avenues for future research concerning the verb *touch* could also complement the findings of this thesis. First, exploring whether speakers distinguish additional meanings of *touch* in speech and gesture would provide a more comprehensive picture of the multimodal profile of this verb. Second, incorporating additional layers of analysis—such as prosodic patterns—when comparing physical vs. emotional meanings of *touch* would allow for a more precise multimodal description (a step that should eventually be taken in similar research as well). Furthermore, several gesture studies could be developed to address remaining questions. In order to determine the real use of chest-touching gestures and rubbing gestures, a 'reversed' search would be necessary: first looking for the gesture and then analyzing its context of use. This is now feasible with the datasets collected in this thesis and with pose detection software such as MediaPipe Studio. Moreover, such a search methodology would enable exploration of how these gestures are used across different languages and cultures. Overall, an important direction for future research would be to extend this line of inquiry cross-linguistically.

References

Alcaraz-Carrión, D. (2018). *Time gestures: Time conceptualisation in English with evidence from gestures in a multimodal corpus* [Doctoral dissertation, Lancaster University]. Lancaster Eprints. https://eprints.lancs.ac.uk/id/eprint/132638/?template=search

Alcaraz-Carrión, D., Alibali, M. W., & Valenzuela, J. (2022). Adding and subtracting by hand: Metaphorical representations of arithmetic in spontaneous co-speech gestures. *Acta Psychologica*, *228*, Article 103624. https://doi.org/10.1016/j.actpsy.2022.103624

Alcaraz-Carrión, D., & Valenzuela, J. (2021). Distant time, distant gesture: Speech and gesture correlate to express temporal distance. *Semiotica*, *2021*(241), 159–183. https://doi.org/10.1515/sem-2019-0120

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, *44*(2), 169–188. https://doi.org/10.1006/jmla.2000.2752

Argaman, O. (2009). Linguistic markers and emotional intensity. *Journal of Psycholinguistic Research*, *39*(2), 89–99. https://doi.org/10.1007/s10936-009-9127-1

Barcelona, A. (1997). Cognitive Linguistics: A usable approach. *Cuadernos de Filología Inglesa*, *6*(2), 7–32.

Barnard, K., & Johnson, M. (2005). Word sense disambiguation with pictures. *Artificial Intelligence*, *167*(1-2), 13–30. https://doi.org/10.1016/j.artint.2005.04.009

Barnard, K., Yanai, K., Johnson, M., & Gabbur, P. (2006). Cross modal disambiguation. In J. Ponce, M. Hebert, C. Schmid, & A. Zisserman (Eds.), *Toward category-level object recognition* (pp. 238–257). Springer. https://doi.org/10.1007/11957959_13

Bates, E. (1979). *The emergence of symbols: Cognition and communication in infancy*. Academic Press. https://doi.org/10.1016/C2013-0-10341-8

Bolumar Martínez, I., Alcaraz-Carrión, D., & Valenzuela, J. (2024). A multimodal approach to polysemy: The senses of touch. *Language and Cognition*, *16*(4), 1697–1717. https://doi.org/10.1017/langcog.2024.23

Bolumar Martínez, I., Alcaraz-Carrión, D., & Valenzuela, J. (2025). A multimodal approach to polysemy: The senses of taste and smell. *Cognitive Semiotics*. Advance online publication. https://doi.org/10.1515/cogsem-2025-2005

Bolumar Martínez, I., Goldin-Meadow, S., Alcaraz-Carrión, D., & Valenzuela, J. (2025). Disambiguating polysemy: How hand gestures help observers interpret the verb touch. *Review of Cognitive Linguistics*. Advance online publication. https://doi.org/10.1075/rcl.00243.bol

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, *75*, 1–28.

Bressem, J. (2013). A linguistic perspective on the notation of form features in gestures. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body - language – communication: An international handbook on multimodality in human interaction* (Vol. 1, pp. 1079–1098). De Gruyter Mouton. https://doi.org/10.1515/9783110261318.1079

Burenhult, N., & Majid, A. (2011). Olfaction in Aslian ideology and language. *The Senses and Society*, *6*(1), 19–29. https://doi.org/10.2752/174589311X12893982233597

Carrillo-de-Albornoz, J., & Plaza, L. (2013). An emotion-based model of negation, intensifiers, and modality for polarity and intensity classification. *Journal of the American Society for Information Science and Technology*, *64*(8), 1618–1633. https://doi.org/10.1002/asi.22859

Carston, R. (2021). Polysemy: Pragmatics and sense conventions. *Mind & Language*, *36*, 108–133. https://doi.org/10.1111/mila.12329

Casasanto, D., & Jasmin, K. (2012). The hands of time: Temporal gestures in English speakers. *Cognitive Linguistics*, *23*(4), 643–674. https://doi.org/10.1515/cog-2012-0020

Classen, C. (1997). Foundations for an anthropology of the senses. *International Social Science Journal*, *49*(153), 401–12. https://doi.org/10.1111/j.1468- 2451.1997.tb00032.x

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, *23*(1), 43–71. https://doi.org/10.1016/0010-0277(86)90053-3

Cienki, A. (1998). Metaphoric gestures and some of their relations to verbal metaphoric expressions. In J. P. Koenig (Ed.), *Discourse and cognition: Bridging the gap* (pp. 189–204). Center for the Study of Language and Information.

Cienki, A., & Müller, C. (Eds.). (2008). *Metaphor and gesture*. John Benjamins Publishing Company. https://doi.org/10.1075/gs.3

Cienki, A., & Müller, C. (2014). Ways of viewing metaphor in gesture. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body - language - communication: An international handbook on multimodality in human interaction* (Vol. 2, pp. 1766–1781). De Gruyter Mouton. https://doi.org/10.1515/9783110302028.1766

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cohen, A. A. (1977). The communicative functions of hand illustrators. *Journal of Communication*, *27*(4), 54–63, https://doi.org/10.1111/j.1460-2466.1977.tb01856.x

Condon, W. S., & Ogston, W. D. (1966). Sound film analysis of normal and pathological behavior patterns. *Journal of Nervous and Mental Disease, 143*(4), 338–347. https://doi.org/10.1097/00005053-196610000-00005

Condon, W. S., & Ogston, W. D. (1967). A segmentation of behavior. *Journal of Psychiatric Research*, *5*, 221–235.

Cooperrider, K., & Mesh, K. (2022). Pointing in gesture and sign. In A. Morgenstern & S. Goldin-Meadow (Eds.), *Gesture in language: Development across the lifespan* (pp. 21–46). De Gruyter Mouton. https://doi.org/10.1515/9783110567526-002

de la Fuente, J., Santiago, J., Román, A., Dumitrache, C., & Casasanto, D. (2014). When you think about it, your past is in front of you: How culture shapes spatial conceptions of time. *Psychological Science*, *25*(9), 1682–1690. https://doi.org/10.1177/0956797614534695

Dölling, J. (2020). Systematic polysemy. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, & T. E. Zimmermann (Eds.), *The Wiley Blackwell Companion to Semantics* (pp. 1–27). John Wiley & Sons. https://doi.org/10.1002/9781118788516.sem099

Efron, D. (1972). *Gesture, race and culture: A tentative study of the spatio-temporal and "linguistic" aspects of the gestural behavior of eastern Jews and southern Italians in New York City, living under similar as well as different environmental conditions*. Mouton.

Ekman, P., & Friesen, W. (1969). The repertoire of nonverbal behavior: Categories, origins, usage and coding. *Semiotica*, *1*, 49–98.

Evans, N., & Wilkins, D. P. (2000). In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language, 76*, 546–592.

Falkum, I. L. (2015). The how and why of polysemy: A pragmatic account. *Lingua*, *157*, 83–99. https://doi.org/10.1016/j.lingua.2014.11.004

Falkum, I. L., & Vicente, A. (2015). Polysemy: Current perspectives and approaches. *Lingua*, *157*, 1–16. http://dx.doi.org/10.1016/j.lingua.2015.02.002

Fauconnier, G. (1997). *Mappings in thought and language*. Cambridge University Press.

Farley, S. D., Akin, K., & Hedgecoth, N. (2021). Exploring the meanings of the heartfelt gesture: A nonverbal signal of heartfelt emotion and empathy. *Journal of Nonverbal Behavior*, *45*(4), 567–585. https://doi.org/10.1007/s10919- 021-00371-5

Fernández Jaén, J. (2012). *Semántica cognitiva diacrónica de los verbos de percepción física del español* [Doctoral dissertation, University of Alicante]. Repository University of Alicante. https://rua.ua.es/entities/publication/185032c4-3f6d-4b6f-9555-64eb1e5bf6c5

Fernández Jaén, J. (2016). Usos modales y epistémicos del verbo sentir. *Revista de Investigación Lingüística*, *19*, 199–226.

Fontaine, J. J. R., Scherer, K. R., & Soriano, C. (2013). *Components of emotional meaning: A sourcebook*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199592746.001.0001

Fillmore, C. (1982). Frame Semantics. In Linguistic Society of Korea (Ed.), *Linguistics in the morning calm: Selected papers from SICOL-1981* (pp. 111–137). Hanshin Publishing Company.

Gaby, A. (2008). Gut feelings: Locating intellect, emotion and life force in the Thaayore body. In F. Sharifian, R. Dirven, N. Yu, & S. Niemeier (Eds.), *Culture, body, and language: Conceptualizations of internal body organs across cultures and languages* (pp. 27–44). De Gruyter Mouton. https://doi.org/10.1515/9783110199109.2.27

Garrah, K., Breeden, M., & Farley, S. D. (2024). The heartfelt gesture positively affects perceptions of personality. *Current Psychology*, *43*(45), 34904–34913. https://doi.org/10.1007/s12144-024-06932-3

Geeraerts, D. (2009). *Theories of lexical semantics*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198700302.001.0001

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. The Belknap Press of Harvard University Press.

Goldin-Meadow, S. (2023). *Thinking with your hands: The surprising science behind how gestures shape our thoughts*. Basic Books.

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, *64*(1), 257–283. https://doi.org/10.1146/annurev-psych-113011-143802

Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in Cognitive Sciences*, *9*(5), 234–241. https://doi.org/10.1016/j.tics.2005.03.006

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan. (Eds.), *Syntax and Semantics: Speech acts* (Vol. 3, pp. 41-58). Academic Press.

Gunter, T. C., Weinbrenner, J. E. D., & Holle, H. (2015). Inconsistent use of gesture space during abstract pointing impairs language comprehension. *Frontiers in Psychology*, *6*, Article 80. https://doi.org/10.3389/fpsyg.2015.00080

Hagoort, P., & Özyürek, A. (2024). Extending the architecture of language from a multimodal perspective. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12728

Haiman, J. (1983). Iconic and economic motivation. *Language*, *59*(4), 781–819. https://doi.org/10.2307/413373

Hardie, A. (2012). CQPweb – combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, *17*(3), 380–409. https://doi.org/10.1075/ijcl.17.3.04har

Hinnell, J., & Parrill, F. (2020). Gesture influences resolution of ambiguous statements of neutral and moral preferences. *Frontiers in Psychology*, *11*, 1–10. https://doi.org/10.3389/fpsyg.2020.587129

Hirst, G. (1987). *Semantic interpretation and the resolution of ambiguity*. Cambridge University Press. https://doi.org/10.1017/CBO9780511554346

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, *19*(7), 1175–1192. https://doi.org/10.1162/jocn.2007.19.7.1175

Holler, J., & Beattie, G. (2003). Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener? *Gesture*, *3*(2), 127–154. https://doi.org/10.1075/gest.3.2.02hol

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. https://doi.org/10.1016/j.tics.2019.05.006

Hostetter, A. B., & Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review*, *26*(3), 721–752. https://doi.org/10.3758/s13423-018-1548-0

Ibarretxe-Antuñano, I. (1999). *Polysemy and metaphors in perception verbs: A cross-linguistic study* [Doctoral dissertation, University of Edinburgh]. Edinburgh Research Archive. https://era.ed.ac.uk/handle/1842/22334

Ibarretxe-Antuñano, I. (2002). MIND-AS-BODY as a cross-linguistic conceptual metaphor. *Miscelánea: A Journal of English and American Studies*, *25*, 93–119.

Ibarretxe-Antuñano, I. (2006). Cross-linguistic polysemy in tactile verbs. In J. Luchjenbroers (Ed.), *Cognitive Linguistics investigations: Across languages, fields and philosophical boundaries* (pp. 235–253). John Benjamins Publishing Company. https://doi.org/10.1075/hcp.15.16iba

Ibarretxe-Antuñano, I. (2008). Vision metaphors for the intellect: Are they really cross-linguistic? *Journal of the Association of Anglo-American Studies*, *30*(1), 15–33.

Ibarretxe-Antuñano, I. (2019). Perception metaphors in Cognitive Linguistics: Scope, motivation, and lexicalisation. In L. J. Speed, C. O'Meara, L. S. Roque, & A. Majid (Eds.), *Perception metaphors* (pp. 43–64). John Benjamins Publishing Company. https://doi.org/10.1075/celcr.19.03iba

Ibarretxe-Antuñano, I., & Valenzuela, J. (2021). *Lenguaje y cognición*. Editorial Síntesis.

Jansegers, M. (2017). *Hacia un enfoque múltiple de la polisemia: Un estudio empírico del verbo multimodal «sentir» desde una perspectiva sincrónica y diacrónica*. De Gruyter. https://doi.org/10.1515/9783110476972

Jansegers, M., Vanderschueren, C., & Enghels, R. (2015). The polysemy of the Spanish verb sentir: A behavioral profile analysis. *Cognitive Linguistics*, *26*(3), 381–421. https://doi.org/10.1515/cog-2014-0055

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. The University of Chicago Press.

Katz, J. J. (1972). *Semantic theory*. Harper & Row.

Katz, J. J., & Fodor, J. A. (1963). The structure of semantic theory. *Language*, *39*(2), 170–210.

Kendon, A. (1982). The study of gesture: Some observations on its history. *Recherches Sémiotiques/Semiotic Inquiry*, *2*(1), 45–62.

Kendon, A. (1997). Gesture. *Annual Review of Anthropology*, *26*, 109–128. https://www.jstor.org/stable/2952517

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press. https://doi.org/10.1017/CBO9780511807572

Kendon, A. (2007). On the origins of modern gesture studies. In S. D. Duncan, J. Cassell, & E. T. Levy (Eds.), *Gesture and the dynamic dimension of language: Essays in honor of David McNeill* (pp. 13–28). John Benjamins Publishing Company. https://doi.org/10.1075/gs.1.03ken

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, *124*(3), 245–266. https://doi.org/10.1037/rev0000059

Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language*, *45*, 259–282.

Klein, D. E., & Murphy, G. L. (2002). Paper has been my ruin: Conceptual relations of polysemous senses. *Journal of Memory and Language*, *47*, 548–570.

Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain and Language*, *81*, 205–223.

Klepousniotou, E. (2007). Reconciling Linguistics and Psycholinguistics: On the psychological reality of linguistic polysemy. In M. Rakova, G. Pethö, & C. Rákosi (Eds.), *The cognitive basis of polysemy: New sources of evidence for theories of word meaning* (pp. 17–46). Peter Lang.

Kövecses, Z. (1986). *Metaphors of anger, pride and love: A lexical approach to the structure of concepts*. John Benjamins Publishing Company. https://doi.org/10.1075/pb.vii.8

Kövecses, Z. (2010). *Metaphor: A practical introduction.* Oxford: Oxford University Press.

Kövecses, Z. (2019). Perception and metaphor: The case of smell. In L. J. Speed, C. O'Meara, L. S. Roque, & A. Majid (Eds.), *Perception metaphors* (pp. 327–346). John Benjamins Publishing Company. https://doi.org/10.1075/celcr.19.16kov

Krishna, P. P., Arulmozi, S., & Mishra, R. K. (2022). Do you see and hear more? A study on Telugu perception verbs. *Journal of Psycholinguistic Research*, *51*(3), 473–484. https://doi.org/10.1007/s10936-021-09827-7

Lacey, S., Stilla, R., & Sathian, K. (2012). Metaphorically feeling: Comprehending textural metaphors activates somatosensory cortex. *Brain and Language*, *120*(3), 416–421. https://doi.org/10.1016/j.bandl.2011.12.016

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. The University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. The University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. Basic Books.

Lakshmi, A. (2024). *Gesture and stereotype content: Inquiries on perception, activation, and spontaneous production* [Doctoral dissertation, The University of Chicago]. Knowledge UChicago. https://knowledge.uchicago.edu/record/12843?v=pdf#files

Levshina, N., & Moran, S. (2021). Efficiency in human languages: Corpus evidence for universal principles. *Linguistics Vanguard*, *7*(s3). https://doi.org/10.1515/lingvan-2020-0081

Lorson, A., Macuch-Silva, V., Hart, C., & Winter, B. (2025). Gesture size affects numerical estimates in quantifier comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *51*(3), 391–424. https://doi.org/10.1037/xlm0001372

Majid, A. (2021). Human olfaction at the intersection of language, culture, and biology. *Trends in Cognitive Sciences*, *25*(2), 111–123. https://doi.org/10.1016/j.tics.2020.11.005

Majid, A., Burenhult, N., Stensmyr, M., de Valk, J., & Hansson, B. S. (2018). Olfactory language and abstraction across cultures. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1752), 1–8. http://dx.doi.org/10.1098/rstb.2017.0139

Majid, A., & Levinson, S. C. (2011). The senses in language and culture. *The Senses and Society*, *6*(1), 5–18. https://doi.org/10.2752/174589311X12893982233551

Maricchiolo, F., de Dominicis, S., Ganucci Cancellieri, U., di Conza, A., Gnisci, A., & Bonaiuto, M. (2014). Co-speech gestures: Structures and functions. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body - language - communication: An international handbook on multimodality in human interaction* (Vol. 2, pp. 1461–1473). De Gruyter Mouton. https://doi.org/10.1515/9783110302028.1461

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. The University of Chicago Press.

McNeill, D. (2000). Introduction. In D. McNeill (Ed.), *Language and gesture* (pp. 1–10). Cambridge University Press. https://doi.org/10.1017/CBO9780511620850.001

McNeill, D. (2013). Gesture as a window onto mind and brain, and the relationship to linguistic relativity and ontogenesis. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body - language - communication: An international handbook on multimodality in human interaction* (Vol. 1, pp. 28–54). De Gruyter Mouton. https://doi.org/10.1515/9783110261318.28

McNeill, D., & Levy, E. T. (1982). Conceptual representations in language activity and gesture. In R. J. Jarvella & W. Klein (Eds.), *Speech, place, and action: Studies in deixis and related topics* (pp. 271–295). John Wiley & Sons.

Monachini, M., & Roventini, A. (1994). Italian audition verbs: A corpus- and frame-based analysis. In W. Martin, W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg, & P. Vossen (Eds.), *Euralex 1994 proceedings* (pp. 117–127). Vrije Universiteit.

Nappa, R., & Arnold, J. E. (2014). The road to understanding is paved with the speaker's intentions: Cues to the speaker's attention and intentions affect pronoun comprehension. *Cognitive Psychology*, *70*, 58–81. https://doi.org/10.1016/j.cogpsych.2013.12.003

Nerlich, B., & Clarke, D. (2003). Polysemy and flexibility: Introduction and overview. In B. Nerlich, Z. Todd, V. Herman, & D. Clarke (Eds.), *Polysemy: Flexible patterns of meaning in mind and language* (pp. 3–30). De Gruyter Mouton.

Norrick, N. R. (1981). *Semiotic principles in semantic theory*. John Benjamins Publishing Company. https://doi.org/10.1075/cilt.20

Nummenmaa, L., Glerean, E., Hari, R., & Hietanen, J. K. (2014). Bodily maps of emotions. *Proceedings of the National Academy of Sciences*, *111*(2), 646–651. https://doi.org/10.1073/pnas.1321664111

Nummenmaa, L., Hari, R., Hietanen, J. K., & Glerean, E. (2018). Maps of subjective feelings. *Proceedings of the National Academy of Sciences*, *115*(37), 9198–9203. https://doi.org/10.1073/pnas.1807390115

Núñez R, Cooperrider K, Doan D, Wassmann J. (2012). Contours of time: Topographic construals of past, present, and future in the Yupno valley of Papua New Guinea. *Cognition*, *124*(1), 25–35.

Núñez, R., & Sweetser, E. (2006). With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science*, *30*(3), 401–450. https://doi.org/10.1207/s15516709cog0000_62

Ortega-Andrés, M., & Vicente, A. (2019). Polysemy and co-predication. *Glossa: a journal of general linguistics*, *4*(1). https://doi.org/10.5334/gjgl.564

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 1–10. https://doi.org/10.1098/rstb.2013.0296

Pagán Cánovas, C., Valenzuela, J., Alcaraz-Carrión, D., Olza, I., & Ramscar, M. (2020). Quantifying the speech-gesture relation with massive multimodal datasets: Informativity in time expressions. *PLoS ONE*, *15*(6), Article e0233892. https://doi.org/10.1371/journal.pone.0233892

Parzuchowski, M., & Wojciszke, B. (2014). Hand over heart primes moral judgments and behavior. *Journal of Nonverbal Behavior*, *38*(1), 145–165. https://doi.org/10.1007/s10919-013-0170-0

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peiró Márquez, L., & Ibarretxe-Antuñano, I. (2021). ¿Cómo incide la gestualidad en la codificación del movimiento? *E-AESLA*, *7*, 19–36. https://cvc.cervantes.es/lengua/eaesla/pdf/07/02.pdf

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291. https://doi.org/10.1016/j.cognition.2011.10.004

Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics*, *17*(4), 409–441.

Pustejovsky, J. (1995). *The generative lexicon*. The MIT Press. https://doi.org/10.7551/mitpress/3225.001.0001

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1972). A grammar of contemporary English. *ELT Journal*, *29*(1), 83–88. https://doi.org/10.1093/elt/29.1.83

Radden, G., & Kövecses, Z. (1999). Towards a theory of metonymy. In K. Panther & G. Radden (Eds.), *Metonymy in language and thought* (pp. 17–59). John Benjamins Publishing Company. https://doi.org/10.1075/hcp.4.03rad

Ricci Bitti, P. E. (2014). Facial expression and social interaction. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body - language - communication: An international handbook on multimodality in human interaction* (Vol. 2, pp. 1342–1349). De Gruyter Mouton. https://doi.org/10.1515/9783110302028.1342

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*, 245–266. https://doi.org/10.1006/jmla.2001.2810

Rosch, E. (1973). Natural categories. *Cognitive Psychology*, *4*, 328–350.

Sharifian, F., Dirven, R., Yu, N., & Niemeier, S. (2008). Culture and language: Looking for the "mind" inside the body. In F. Sharifian, R. Dirven, N. Yu, & S. Niemeier (Eds.), *Culture, body, and language: Conceptualizations of internal body organs across cultures and languages* (pp. 3–26). De Gruyter Mouton. https://doi.org/10.1515/9783110199109.1.3

Selmistraitis, L., & Boikova, R. (2020). Source domains of smell related metaphorical collocations: Study based on corpus of contemporary American English. *Respectus Philologicus*, *38*(43), 11–24. https://doi.org/10.15388/RESPECTUS.2020.38.43.54

Smith, W. G., & Kam, C. L. H. (2012). Knowing 'who she is' based on 'where she is': The effect of co-speech gesture on pronoun comprehension. *Language and Cognition*, *4*(2), 75–98. https://doi.org/10.1515/langcog-2012-0005

Stukenbrock, A. (2014). Pointing to an 'empty' space: Deixis am Phantasma in face-to-face interaction. *Journal of Pragmatics*, *74*, 70–93. https://doi.org/10.1016/j.pragma.2014.08.001

Sweetser, E. (1986). Polysemy vs. abstraction: Mutually exclusive or complementary? *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*, 528–538.

Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge University Press.

Sweetser, E. (2023). Gestural meaning is in the body(-space) as much as in the hands. In T. Janzen & B. Shaffer (Eds.), *Signed language and gesture research in Cognitive Linguistics* (pp. 157–180). De Gruyter Mouton. https://doi.org/10.1515/9783110703788-007

Taylor, J. R. (2003). *Linguistic categorization*. Oxford University Press.

Taylor, J. R. (2006). Polysemy and the lexicon. In G. Kristiansen, M. Achard, R. Dirven, & F. J. Ruiz de Mendoza Ibáñez (Eds.), *Cognitive Linguistics: Current applications and future perspectives* (pp. 51–80). Mouton de Gruyter. https://doi.org/10.1515/9783110197761.1.51

Tyler, A., & Evans, V. (2003). *The semantics of English prepositions: Spatial scenes, embodied meaning, and cognition*. Cambridge University Press. https://doi.org/10.1017/CBO9780511486517

Uhrig, P. (2018). NewsScape and the Distributed Little Red Hen Lab – A digital infrastructure for the large-scale analysis of TV broadcasts. In A. Zwierlein, J. Petzold, K. Böhm, & M. Decker (Eds.), *Anglistentag 2017 in Regensburg: Proceedings* (pp. 99–114). Wissenschaftlicher Verlag Trier.

van Putten, S. (2020). Perception verbs and the conceptualization of the senses: The case of Avatime. *Linguistics*, *58*(2), 1–38. https://doi.org/10.1515/ling-2020-0039

Vicente, A., & Falkum, I. L. (2017). Polysemy. In M. Aronoff (Ed.), *Oxford research encyclopedia of Linguistics*. Oxford University Press. http://doi.org/10.1093/acrefore/9780199384655.013.325

Winter, B. (2019). *Sensory linguistics: Language, perception and metaphor*. John Benjamins Publishing Company. https://doi.org/10.1075/celcr.20

Yang, J., Gao, M., Li, Z., Gao, S., Wang, F., & Zheng, F. (2023). *Track Anything: Segment Anything meets videos*. arXiv. https://doi.org/10.48550/arXiv.2304.11968

Yu, N. (2009). *The Chinese heart in a cognitive perspective: Culture, body, and language*. De Gruyter Mouton. https://doi.org/10.1515/9783110213348

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley Press.

Appendix I

Touch			
Negation		**Intensifiers**	
Almost	Cannot/can't	(A) little	(A) lot
Couldn't	Did not/didn't	Barely	Completely
Do not/ don't	Hasn't	Deeply	Enormously
Have not/haven't	Never	Enough	Especially
No one	Nobody	Greatly	Incredibly
Not	Not going to	Kind of	More
Stop	Without	Much	Profoundly
Won't	Wouldn't	Quite	Really
		So	Sort of
		Such	(The) most
		Totally	Truly
		Very	

Table AI.1. List of linguistic markers in Bolumar Martínez et al. (2024)

			Filtering process				Discarded gestures			Final dataset
	Category	**Hits rendered**	**Instances out of interest**	**Non-visible hands**	**Repeated clips**	**No gesture**	**Unclear gesture**	**Unrelated gesture**	**Acted gesture**	**Related spontaneous gesture**
Touch + p.pr./someone/somebody	Physical	1629	-	1144	118	112	4	88	5	**158**
	Emotional	331	-	280	9	8	2	10	0	**22**
	Ambiguous	16	16	-	-	-	-	-	-	**-**
	Tagging error	57	57	-	-	-	-	-	-	**-**
	Other	210	210	-	-	-	-	-	-	**-**
	Tech. issue/ t-s mismatch	328	328	-	-	-	-	-	-	**-**
	TOTAL	2571	611	1424	127	120	6	98	5	**180**
Touched me	Physical	565	-	384	57	66	1	27	2	**28**
	Emotional	1024	-	754	62	79	1	34	0	**94**
	Ambiguous	14	14	-	-	-	-	-	-	**-**
	Tagging error	2	2	-	-	-	-	-	-	**-**
	Other	4	4	-	-	-	-	-	-	**-**
	Tech. issue/ t-s mismatch	89	89	-	-	-	-	-	-	**-**
	TOTAL	1698	109	1138	119	145	2	61	2	**122**

Table AI.2. Raw data of *touch* filtering process in Bolumar Martínez et al. (2024)

Taste			*Smell*		
Negation	**Verb modifiers**	**Object quantifiers**	**Negation**	**Verb modifiers**	**Object quantifiers**
Almost	Actually	Bit	Almost	Actually	Bit
Didn't	Basically	Heck of a lot	Can't	Clearly	Bunch
Does not/doesn't	Clearly	Little (more)	Could not	Essentially	Little (more)
Don't	Just	Little bit (more)	Did not/didn't	Just	Little bit (more)
In no way	Kind of	Little extra	Does not/doesn't	Kind of	Lot (more)
Never	Literally	Little too	Don't	Literally	Whole lot
None	Really	Lot (more)	Never	Quite	
Not	Really just	Million	No way	Really	
Nothing	Slightly	Slight	Not	Sort of	
Will not		Thousand	Should not		
		Whole lot	Without		

Table AI.3. List of linguistic markers in Bolumar Martínez, Alcaraz-Carrión et al. (2025)

			Filtering process			Discarded gestures			Final dataset
	Category	**Hits rendered**	**No visible speaker**	**Mismatch/tech. issue/ tagging error/repeated**	**Non-visible hands**	**No gesture**	**Unrelated gesture**	**Acted gesture**	**Related gesture**
TASTE (all searches)	Literal					195	276	2	**33**
	Figurative					56	58	0	**5**
	Not applicable					1	1	0	**1**
	TOTAL	4272	1470	1968	206	252	335	2	**39**

Table AI.4. Raw data of *taste* filtering process in Bolumar Martínez, Alcaraz-Carrión et al. (2025)

			Filtering process			Discarded gestures			Final dataset
	Category	**Hits rendered**	**No visible speaker**	**Mismatch/tech. issue/ tagging error/repeated**	**Non-visible hands**	**No gesture**	**Unrelated gesture**	**Acted gesture**	**Related gesture**
SMELL (all searches)	Literal					336	276	7	**59**
	Figurative					74	162	4	**11**
	Not applicable					14	18	-	**1**
	TOTAL	4241	1658	1030	591	424	456	11	**71**

Table AI.5. Raw data of *smell* filtering process in Bolumar Martínez, Alcaraz-Carrión et al. (2025)