

Universitat de València
Facultat de Filologia, Traducció i Comunicació
Departament de Filologia Espanyola



**ANÁLISIS MULTIMODAL EN LA
DISTINCIÓN ENTRE
INTERVENCIÓN Y TURNO:
EFECTOS EN LA SEGMENTACIÓN DE
CONVERSACIÓN DESDE EL MODELO
VAL.ES.CO.**

Tesis doctoral presentada por:

Sandra Cabanes Pérez

Director:

Salvador Pons Bordería

Estudios Hispánicos Avanzados
Valencia, Junio 2023



Esta tesis ha contado con una ayuda para la Formación del Personal Investigador (referencia BES-2017-080738) y con el apoyo del proyecto *Unidades discursivas para una descripción sistemática de los marcadores del discurso en español* (referencia FFI2016-77841-P, investigador principal Salvador Pons Bordería), financiado por el MINECO, la AEI y los fondos FEDER.

A mis compañeros doctorandos. Sin vosotros, nada.

Agradecimientos

De esta etapa de mi vida tengo que agradecer a Salvador Pons, mi tutor, la oportunidad de trabajar en un grupo de investigación de prestigio como es el Grupo Val.Es.Co. y el aprendizaje que me llevo sobre el mundo académico. También doy las gracias al Grupo y al Departamento de Filología Española por acogerme, especialmente a los profesores y las profesoras con los que he compartido asignatura: Ana Llopis, Marta Pilar Montañez, Virginia González, Mustapha Mohamed-Lamin, David Giménez, Santiago Vicente, Dorota Kotwica, Karolina Grzech; así como a Marta Albelda por considerarme en primer lugar como becaria y las palabras de apoyo en malos momentos.

De fuera de la Universidad de Valencia, tengo que agradecer enormemente a Inés Olza el trato amabilísimo y sus grandes consejos en la Universidad de Navarra y en sus visitas a Valencia. I also want to thank Cornelia Müller for her methodological guidance and the dedication with which she observed the data from my corpus. My thesis would not be what it is today without her help. I would also like to express my gratitude to her students, Clara Kindler and Jana Junge, for welcoming me to the European University Viadrina.

De los compañeros y las compañeras del Doctorado es de donde me llevo el mejor recuerdo de esta etapa. De los primeros Valesquitos, que han volado primero, Elena Pascual, Shima Salameh y Adrià Pardo, y de los que estamos desplegando las alas ahora: Tomás Córcoles, Alejandra Hernández, Guadalupe Espinosa, Danny Lanza, Carlos Castelló, Sara Badia, Amparo Alemany. Solo nosotros sabemos lo que es esto, lo bueno y lo malo que hemos pasado; sobre todo, nos hemos cuidado y nos hemos dado apoyo. No hay mucha gente con la que yo sienta que puedo abrirme lo suficiente como para decir que soy yo misma, pero en ellos he encontrado un lugar en el que estoy a gusto. Espero que parte de esto no se pierda con la distancia y el tiempo.

Entrando en el círculo personal, gracias a Joan por todo el apoyo; quizás el único amigo que se lea esta tesis, aunque sufra un poco leyendo sobre Pragmática. Muchas gracias a Lorena por la ilustración para la portada. En realidad, mi familia y amigos más cercanos han participado como informantes en el corpus de la tesis, así que nombrándolos a cada uno los desanonimizaría. Gracias por darme permiso para grabaros y analizar vuestros gestos (¿qué mayor muestra de apoyo hay?) y por formar así parte de esta tesis. Pero, sobre todo, gracias por ser esa ventana por la que asomarme y recordar que hay vida más allá del escritorio.

Resumen

La interacción multimodal de la conversación se ha trabajado generalmente desde estudios etnometodológicos multimodales (Goodwin 1981, Mondada 2006, 2007, 2013a, 2013b, Deppermann 2013), pero también desde un enfoque lingüístico, que considera la integración de los gestos en el enunciado, se han abordado las funciones pragmáticas de los gestos en la interacción (Kendon 2004, Müller 2013, en prep., Payrató y Teßendorf 2014). Puesto que los modelos de segmentación del discurso —Análisis del Discurso (Sinclair y Coulthard 1975, 1992), Escuela de Ginebra (Roulet et al. 1985, 2001), Coenunciación (Morel y Danon-Boileau 1998) y Grupo Val.Es.Co. (Briz *et al.* 2003, Pons Bordería 2022)— han servido para sistematizar el funcionamiento de los fenómenos lingüísticos, pragmáticos y sociales que ocurren en la conversación; en este estudio buscamos la incorporación de los elementos multimodales en esa estructura y que estos elementos contribuyan en la descripción de las unidades conversacionales y las relaciones que se establecen entre ellas.

Para aplicar este enfoque pragmático estructural a la interacción multimodal, utilizamos el modelo jerárquico de Val.Es.Co. (Briz y Grupo Val.Es.Co. 2003, Pons Bordería 2022) y nos centramos en el nivel dialógico de la estructura conversacional, que comprende la intervención como unidad estructural y el turno como unidad social vinculada al intercambio de roles de hablante y oyente. El objetivo del estudio es examinar los elementos multimodales que afectan a la segmentación de la conversación coloquial en español en el nivel dialógico y de qué manera le afecta.

Para ello, se ha grabado un corpus de conversaciones coloquiales y se han comparado las segmentaciones realizadas a partir del audio y el video. Respecto a la preparación de los datos, se ha desarrollado una metodología para la grabación de conversaciones coloquiales que mantengan la naturalidad propia de este género, así como una metodología de anotación de los gestos en la transcripción de la conversación mediante programas informáticos y en papel. A partir de este material, se realizan dos análisis: el primero valora la cantidad y los tipos de cambios que se producen en la segmentación de intervenciones y turnos entre la grabación oral y la visual, así como las causas de estos cambios; en el segundo análisis, se observan las características de los gestos que forman parte de esta segmentación multimodal, teniendo en cuenta su posición en la estructura de la intervención y del intercambio, su forma y su función.

Abstract

The multimodal interaction of conversation has been studied primarily from multimodal ethnomethodological approaches (Goodwin 1981, Mondada 2006, 2007, 2013a, Deppermann 2013), but also from a linguistic perspective, which considers the integration of gestures in utterances and explores the pragmatic functions of gestures in interaction (Kendon 2004, Müller 2013, in prep., Payrató and Teßendorf 2014). Discourse segmentation models —such as Discourse Analysis (Sinclair and Coulthard 1975, 1992), the Geneva School (Roulet et al. 1985, 2001), Co-enunciation (Morel and Danon-Boileau 1998), and the Val.Es.Co. Group (Briz et al. 2003, Pons Bordería 2022)— have been used to systematize the functioning of linguistic, pragmatic, and social phenomena occurring in conversation. Likewise, this study aims to incorporate multimodal elements into the conversation structure and explore how these elements contribute to the description of conversational units and the relationships established between the units.

To apply this pragmatic-structural approach to multimodal interaction, we utilize the hierarchical model of Val.Es.Co. (Briz and Val.Es.Co. Group 2003, Pons Bordería 2022) and focus on the dialogical level of conversational structure. This level encompasses the *intervention* as a structural unit and the *turn* as a social unit related to the exchange of speaker and listener roles. The objective of the study is to examine the multimodal elements that influence the segmentation of colloquial conversation in Spanish at the dialogical level and how the segmentation is affected.

For this purpose, a corpus of colloquial conversations has been recorded, and segmentations based on audio and video have been compared. Regarding data preparation, a methodology has been developed for recording colloquial conversations that maintains the naturalness of this genre, as well as an annotation methodology for gestures in conversation transcriptions using both software and paper-based methods. Two analyses have been conducted using this material. The first one assesses the quantity and types of changes that occur in the segmentation of interventions and turns between oral and visual recordings, as well as the causes of these changes. The second analysis examines the characteristics of gestures that are part of this multimodal segmentation, considering their position in the structure of the intervention and the exchange, their form, and their function.

ÍNDICE

SUMMARY	17
CAPÍTULO 1. INTRODUCCIÓN	57
1.1. MOTIVACIÓN DEL ESTUDIO	59
1.2. OBJETIVOS DE LA INVESTIGACIÓN	60
1.3. ESTRUCTURA DEL TRABAJO	61
BLOQUE I	63
CAPÍTULO 2. UN ENFOQUE ESTRUCTURAL Y SOCIAL DEL TURNO	65
2.1. INTRODUCCIÓN	67
2.2. LA CONVERSACIÓN: ENFOQUES PARA SU ESTUDIO	67
2.1.1. Conversación como conducta social.....	68
2.1.2. Conversación como negociación	70
2.1.3. Conversación como estructura.....	72
2.1.4. Estructuras jerárquicas de la conversación	74
2.2. EL TURNO	79
2.2.1. El turno y su construcción social	80
2.2.2. La relación pragmática entre turnos	84
2.2.3. Turnos de hablante y turnos de oyente	91
2.2.4. Ratificación del turno	100
2.3. RECAPITULACIÓN: LA MULTIMODALIDAD EN UN MODELO ESTRUCTURAL DE LA CONVERSACIÓN	106
CAPÍTULO 3. LA INTERACCIÓN MULTIMODAL.....	113
3.1. INTRODUCCIÓN	115
3.2. LA INTERACCIÓN MULTIMODAL Y SUS ENFOQUES	115
3.2.1. La multimodalidad y los modos	115
3.2.2. Modos para una interacción multimodal	119
3.2.3. Enfoques sobre la multimodalidad	122

3.2.4. Conclusión	129
3.3. ORIENTACIÓN INTERACTIVA: POSTURA Y MIRADA	131
3.3.1. Compromiso con el diálogo.....	131
3.3.2. Gestión del turno	135
3.4. PARÁMETROS PARA LA CLASIFICACIÓN DE GESTOS.....	139
3.4.1. Grado de convencionalización.....	139
3.4.1.1. <i>Gestos espontáneos y emblemas</i>	140
3.4.1.2. <i>Gradación de la convencionalización: gestos recurrentes</i>	142
3.4.2. Dependencia temporal con el habla.....	145
3.4.2.1. <i>Temporalidad del gesto</i>	145
3.4.2.2. <i>Sincronía de gesto y habla</i>	147
3.4.2.3. <i>Sincronía de gesto y turno</i>	148
3.4.2.4. <i>Gestos no coverbales</i>	150
3.4.3. Función	151
3.4.3.1. <i>Gestos referenciales y gestos pragmáticos</i>	154
3.4.3.2. <i>Gestos referenciales</i>	156
3.4.3.3. <i>Gestos pragmáticos</i>	159
3.4.3.3.1. Gestos rítmicos y discursivos	160
3.4.3.3.2. Gestos modales.....	162
3.4.3.3.3 Gestos performativos.....	164
3.4.3.3.4 Gestos interactivos	171
3.5. RECAPITULACIÓN	181
CAPÍTULO 4. METODOLOGÍAS PARA UN CORPUS MULTIMODAL	183
4.1. INTRODUCCIÓN	185
4.2. INCORPORACIÓN DEL VÍDEO EN EL ANÁLISIS DE LA INTERACCIÓN	185
4.3. CORPUS DE CONVERSACIONES COLOQUIALES EN VÍDEO.....	188
4.4. ANOTACIÓN DE GESTOS	192
4.4.1. Sistema de anotación MUMIN	196

4.4.2. Sistema de anotación LASG.....	198
4.4.3. Sistema de transcripción multimodal de Mondada	203
4.5. RECAPITULACIÓN	204
BLOQUE II.....	207
CAPÍTULO 5. METODOLOGÍA.....	209
5.1. OBJETIVOS Y METODOLOGÍA.....	211
5.2. DISEÑO DEL CORPUS	214
5.2.1. Preservación de los rasgos situacionales de la conversación y criterios éticos	214
5.2.2. Características de las conversaciones	216
5.2.3. Equipo de grabación	218
5.3. TRANSCRIPCIÓN Y SEGMENTACIÓN DE UNIDADES CONVERSACIONALES	220
5.3.1. Transcripción	221
5.3.2. Segmentación	226
5.4. ANÁLISIS DE CAMBIOS EN LA SEGMENTACIÓN	229
5.4.1. Variable cambio.....	230
5.4.2. Variable causa.....	232
5.4.3. Criterios de segmentación de gestos.....	240
5.4.4. Anotación de cambios en la segmentación.....	249
5.5. ANÁLISIS DE LOS GESTOS	252
5.5.1. Variable posición.....	257
5.5.2. Variable forma.....	258
5.5.3. Anotación de gestos: la forma	259
5.5.4. Variable función	271
5.5.5. Anotación de gestos: la función.....	273
CAPÍTULO 6. RESULTADOS	277
6.1. CAMBIOS EN LA DIMENSIÓN ESTRUCTURAL Y EN LA DIMENSIÓN SOCIAL	279

6.1.1. La información visual es necesaria en la segmentación dialógica de la conversación.	279
6.1.2. La información visual enriquece la dimensión social de las intervenciones.	281
6.1.3. Causas visuales de cambios en la segmentación estructural	287
6.1.4. Causas visuales de cambios en la segmentación social	291
6.1.5. Resumen	295
6.2. ANÁLISIS DE GESTOS INDEPENDIENTES Y COVERBALES	297
6.2.1. Función	298
6.2.1.1. <i>Gestos pragmáticos performativos</i>	300
6.2.1.2. <i>Gestos pragmáticos interactivos</i>	306
6.2.1.3. <i>Gestos pragmáticos modales</i>	318
6.2.1.4. <i>Gestos pragmáticos discursivos</i>	321
6.2.1.5. <i>Gestos referenciales</i>	322
6.2.1.6. <i>Gestos referenciales-pragmáticos</i>	323
6.2.2. Función y posición	323
6.2.3. Forma y función	330
6.2.4. Posición y forma	334
6.2.5. Resumen	334
CAPÍTULO 7. CONCLUSIONES	337
CHAPTER 7. CONCLUSIONS	345
BIBLIOGRAFÍA	353
ANEXO I	387
ANEXO II	390
ANEXO III	392

SUMMARY

CHAPTER 1

In Chapter 1, the motivation for this dissertation (§1.1), its objectives (§1.2), and its structure (§1.3) are explained.

Conversation is a genre that typically takes place in the physical presence of its interlocutors (Briz 1996), and the elements of conversation cannot be studied without considering this reality. Various approaches have examined conversation and the modes of communication that are part of it, beyond words, such as visually captured bodily and material resources. These also form part of the sequential order of social interaction (Mondada 2014: 98).

Among the two predominant approaches in the study of gestures, the Interactional and the Cognitive ones, this study falls within the former, as it conceives of multimodality as a tool for social interaction and organizing sequential order. Additionally, an observational and qualitative methodology is employed. Especially in recent decades, ethnomethodological studies linked to Conversation Analysis have developed this aspect of multimodal interaction (Goodwin 1981, Mondada 2006, 2007, 2013a, 2013b, Deppermann 2013, Goodwin 2018; Streeck 2009a, 2017), which includes gestures and other bodily actions in the social organization and sequential structure of conversation.

However, gestures have also been studied from a Linguistic perspective, focusing on their semantics, the relationship between form and meaning, and their interpretation in the communicative context (McNeill 1992, Kendon 2004, Müller 2004, Cienki and Müller 2008). Chapter 3 elaborates on these linguistic characteristics of gestures, including their degree of conventionality and their temporal dependency on speech.

Furthermore, with the integration of gestures into utterances, both referential and pragmatic functions of gestures in interaction have been addressed (Kendon 2004, Müller 2013, in prep., Payrató and Teßendorf 2014). Pragmatics has extended the concept of speech acts to bodily movements, either combined with speech or not. These have been given various names, including *visible actions* (Kendon 2004) or *composite utterances* (Enfield 2009).

Taking into account this framework of multimodal interaction, these ideas can be connected with other pragmatic studies that have proposed models for segmenting

discourse around the speech act unit into a series of hierarchical units that explain the structure of conversation—Discourse Analysis (Sinclair and Coulthard 1975, 1992), the Geneva School (Roulet et al. 1985, 2001), and the Val.Es.Co. Group (Briz et al. 2003, Pons Bordería 2022). The hierarchical units of these models and their positional relationship have allowed for a systematic description of linguistic, pragmatic, and social phenomena occurring at the dialogical level of conversation (the relationship established between turns, forming exchanges and larger units), but they have not been applied to describe multimodal behavior. They have not considered visual information in their segmentation of conversational units. The Co-enunciation model (Morel and Danon-Boileau 1998) did include multimodality in its hierarchy at a level below the turn, but a study that combines hierarchical segmentation of interventions at the dialogical level with multimodality has not yet been conducted. Therefore, his study aims to find novel results by exploring conversation from this theoretical framework.

To apply this structural pragmatic approach to multimodal interaction, the Val.Es.Co. segmentation model (Briz and Val.Es.Co. Group 2003, Pons Bordería 2022) is used because, unlike the others, it explains the development of conversation from two dimensions: the structural dimension and the social dimension. At the dialogical level, this creates two units: *intervention* and *turn*. While an intervention is an utterance, generally associated with a change of sender, which is or provokes a reaction, typically linguistic (in our case, also gestural); a turn is a type of intervention recognized or accepted by another interlocutor or interlocutors. This distinction helps to describe in greater detail what occurs sequentially and socially, helping the identification of patterns of gestures in this system.

Starting from a linguistic conception of gesture, gestures could be adapted to the Val.Es.Co. model hierarchy because they are governed by structural and semiotic principles similar to language, allowing to find linguistic functions within the utterance in collaboration or exchange with verbal information. Therefore, in this study, the incorporation of multimodal elements into that structure is sought. These elements could contribute to the description of conversational units and the relationships established between them. In other words, this structural approach could enrich our understanding of multimodality, and in turn, multimodality could complement structural studies of conversation, which are biased because they do not consider all the verbal and non-verbal elements that are part of communication.

The overall aim of the research is to describe how multimodality affects the dialogical structure and social dimension of colloquial conversations in Spanish. This leads to a series of specific objectives listed below:

1. Describe the multimodal elements that affect structural and social dialogical segmentation. This includes the formal, functional, and positional description of gestures, as well as the identification of visible contextual information that may change the interpretation of structural or social interventions.
2. Incorporate the annotation of multimodal elements into the Val.Es.Co. transcription (Pons Bordería 2022)—whose layout reflects dialogical segmentation—both on paper and in video annotation software.
3. Describe the effects of multimodal elements on structural and social segmentation.
4. Evaluate which multimodal elements have the greatest impact on structural and social dialogical segmentation.
5. Demonstrate that segmentation based on verbal elements of conversation partially reflects the reality of this type of discourse and that the incorporation of multimodal elements is necessary for a complete dialogical-level description.

The first block of this work includes chapters 2, 3, and 4 and establishes the theoretical framework of the study. It focusses on the characteristics of the type of discourse and the type of units which the multimodal approach is incorporated into, as well as the characteristics of multimodality itself and how real data can be obtained for analysis. The second block, comprising chapters 5, 6, and 7, presents the study conducted: its methodology, results, and conclusions.

CHAPTER 2

Chapter 2 defines the discursive genre in which the object of study is situated, colloquial conversation (§2.2), and the turn as a dialogical unit (§2.3) that is part of a hierarchical structure and depends on a social system for its joint construction between the speaker and the listener. It justifies the use of the conversational segmentation model developed by the Val.Es.Co. Group as the foundation for this work, as well as the advantages of applying it to the study of multimodality (§2.4).

The discursive genre of conversation is characterized by immediate, face-to-face interaction; a non-predetermined turn-taking system; dynamism, where the alternation of turns fosters dialogical tension; and feedback, where cooperation with the conversation topic and the intervention of the other are prominent (Val.Es.Co. Group and Briz 1995: 28-29). Specifically, colloquial conversation exhibits several features. Firstly, it lacks planning (or is planned on the go), has an interpersonal or socializing purpose, and maintains an informal tone as primary traits (Val.Es.Co. Group and Briz 1995: 29-30). Secondly, colloquializing features characterize prototypical colloquial conversation and have the ability to make a non-colloquial conversation sound colloquial. These features include the equality of interlocutors (due to their social roles or the situation itself), a sense of shared knowledge and experience, a familiar interaction context (everyday physical space and specific participant relationships), and non-specialized topics (Val.Es.Co. Group and Briz 1995: 30-32).

Conversation can be studied from different approaches that can complement each other: 1) as social behavior, 2) as negotiation or argumentation, and 3) as structure (Briz 1998: 43).

1) The first approach is led by ethnomethodologists in Conversation Analysis (CA). They present a methodology and theory that blend sociology and linguistics but also draw from anthropology and psychology due to their focus on speech as a social action (Schegloff 1991:46). CA seeks interactive patterns in conversation organization, which are explained by sequential expectations oriented by participants (Levinson 1983: 323). Sacks, Schegloff, and Jefferson (1974: 700-701) observed real conversations to create 14 rules describing this type of interaction in which participants coordinate to construct turns whose content, size, and order are not predetermined. A notion of great impact in studies dealing with oral interaction is the concept of the *local management system*: the social process by which conversation participants manage turn-taking moment by moment—see the system in detail in §2.3.1. This system explains phenomena specific to conversation that do not typically appear in more planned dialogical genres: overlapped turn transitions, overlapped collaborations, conversational crosses and splits, restarts, self-interruptions, repairs, competitive overlaps, and so on. CA differs from other theoretical paradigms in that the emphasis is not on speech acts but on how these are framed within sequences of social action (Raymond and Olguín 2022: 83).

2) From a pragmatic perspective, conversation as negotiation implies that its construction requires cooperation and strategy to achieve the other's acceptance. The Geneva School (Roulet, Trognon, Auchlin, Rémi-Giraud, Moeschler, Zenone) is based on a concept of discourse as negotiation between interlocutors, constructed cooperatively and expressing arguments to reach an agreement (Roulet et al. 1985, Moeschler 1985). According to their model, each exchange aims to achieve *complétude interactionnelle* (Roulet et al. 1985) or *complétude dialogique* (Roulet et al. 2001). This means that the initiating intervention receives the preferred or desired reaction that the speaker intended, and a double agreement is reached to close the negotiation. Additionally, negotiation also aims for *complétude interactive* (Roulet et al. 1985) or *complétude monologique* (Roulet et al. 2001), ensuring good message transmission during the exchange and respecting the interlocutors' image. If negotiation fails to meet these constraints, a secondary negotiation is generated to achieve the desired reaction or clarify the intervention, creating a recursive hierarchical organization.

3) Regarding the structure of conversation, both the social and negotiator/pragmatic approaches have formulated more or less complex structural descriptions. Pragmatic proposals have developed hierarchical discourse structures to explain their functioning. These hierarchies resemble those of linguistic units, where lower units form higher ones. Generally, they use the speech act as the minimal unit. Unlike the inductive perspective of CA, this hierarchical segmentation is deductive, so once previous categories are created, conversation phenomena are fitted into them (Cortés Rodríguez 2002: 175). In Romance Linguistics, discourse segmentation has been of great interest since 1980 (Pons Bordería 2014).

This chapter mentions some segmentation models, but it particularly explains those that have developed dialogical units: Discourse Analysis (DA) (Sinclair and Coulthard 1975, 1992), the Geneva School (Roulet et al. 1985, 2001), and the Val.Es.Co. Group (Briz et al. 2003, Pons Bordería 2022). It dedicates a subsection to comparing their ranges and the subordination system of their units.

The concept of the *turn* is described in various ways by different approaches, so in this section, it is explained through the contrast or similarities of the mentioned hierarchical models. The subsections cover: 1) the turn and its social construction, focusing on its temporal segmentation, which depends on its social construction; 2) the pragmatic relationship between turns, marked by the projectability and reaction

relationships they have with each other; 3) speaker turns and listener turns, as interlocutors alternate these roles to sustain the dialogical tension of the conversation, with a particular focus on the turns emitted by the listener to sustain sequence structures; and 4) turn ratification, as a tool used by the Val.Es.Co. Group to explain the link between the structure of exchanges and the social roles of the interlocutors.

Regarding the pragmatic relationship between turns, CA focuses on what the TCUs themselves project as action and their effect on the sequence. This relationship between the turns of speakers is based on the concept of the *adjacency pair*. The pair consists of statements produced by different speakers in an adjacent (not necessarily contiguous) manner, which are ordered as a first part and a second part thanks to *conditional relevance* (Schegloff 1972): given a first pair part (FPP), a second pair part (SPP) is immediately relevant and expected.

Generally, the temporal boundaries of a turn are based on a change in the speaker (Jaffe and Feldstein 1970: 19). However, for the Val.Es.Co. Group (2014), this is a prototypical characteristic, but in colloquial conversation, there are often also discontinuous interventions produced in overlap with those of another speaker. Therefore, while other analysts base the intervention on the physical boundary of a voice change, the main and defining characteristic of the Val.Es.Co. unit is "the maintenance of the same illocutionary force linked to a single strategy" (*trans.* Briz and Val.Es.Co. Group 2003). For example, if a speech emission (A) overlaps with that of another speaker (B), (A) will be considered one or two interventions depending on whether it reacts to (B) or not. If it reacts, there are two interventions because they respond to different communicative strategies. If it doesn't react, it is a single intervention because the original strategy continues.

Regarding the structure of the turn, CA describes its construction as a process (Sacks et al. 1974): when the speaker begins to speak, they have planned to complete at least one *turn-constructive unit* (TCU); however, they do not know how the turn will finish because it depends on the other participant. When the projected ending is fulfilled, a *transition relevance place* (TRP) occurs (Sacks et al. 1974). This phased construction of the turn provides an essential feature of the turn: the length of the turn is decided during the interaction. In parallel to CA theory, from semiotics, Duncan and his collaborators (Duncan and Fiske 1977) developed more signals that activate the turn distribution mechanism at an TRP. In Spanish, Cestero (2000a) characterizes turn-taking based on two

variables: the end of the clause and the overlap, based on the model of the interactive construction of turns by Wilson, Wiemann, and Zimmerman (1984).

Both CA and DA share the use of categories such as *requests, invitations, greetings*, etc. (Levinson 1983), but CA has used them to describe the relationship between turns, and DA uses them to name internal components of the turn with those functions. In DA, the relationship is established between the moves of each speaker, which form an exchange, not between complete turns. An exchange consists of, at a minimum, one move in the initiation position and another in the response position. These positions are also taken into account by the Geneva hierarchical module and the Val.Es.Co. model. According to these models, the exchange is formed, at a minimum, by two interventions with different illocutionary functions (Roulet 1985:25): initiatives, which request a reaction, and reactives, which respond to a previous intervention. For the Val.Es.Co. model, these positions characterize the intervention as a unit: "The intervention is the maximum structural monologic unit, associated with a change of speaker, characterized by being or provoking a reaction, prototypically linguistic" (*trans.* Briz and Val.Es.Co. Group 2003: 11). The criteria for classifying interventions as initiatives or reactives in each segmentation model are explained in the chapter, as well as the relationships between a group of successive interventions.

The structure of the exchange and the pragmatic requirements of its interventions are influenced by the roles of the interlocutors because participants take on and exchange the roles of speaker and listener to facilitate turn-taking management. In the end, it is an activity that is socially managed among the participants. Therefore, in this chapter, the roles themselves are explained, the characterization of certain turns as belonging to the listener's role, and their structural function.

The interlocutors in a conversation continuously emit and receive information; they are permanent senders and receivers (Briz and Val.Es.Co. Group 2003: 16-17; Padilla 2004). Furthermore, the construction of a turn does not depend solely on the speaker; it requires active participation from the listener for its development. Through *recipient design*, the actions or inaction of the listener can lead to substantial modifications in the statement the speaker is producing (Sacks et al. 1974, Goodwin and Heritage 1990). The speaker monitors during their utterance whether their message is appropriate in the reaction of the other person, which allows him/her to modify it if necessary, and silence

after the turn is not problematic because it has already been ratified as relevant (Goodwin 1981: 103).

So, during the issuance of a turn, the speaker continues to be a recipient of stimuli and, at the same time, the listener receiving the turn provides feedback information that can modify the turn or, at the very least, confirm reception (Briz and Val.Es.Co. Group 2003: 16), either verbally or through other means (eye contact, gestures, non-verbal sounds). If the emission of words and silence are not sufficient to characterize the role of each interlocutor at a given moment, it becomes necessary to distinguish what types of interventions demonstrate that one interlocutor has the floor and which do not. Some authors propose a distinction between the turn as the actual thematic emission, with referential content, "endowed with propositional content, responsible for advancing the conversation's theme" (*trans.* Gallardo 1998: 34), and another turn that has a meta-conversational or metadiscursive function. The former conveys the information that the interlocutors exchange during the conversation, while the latter enables this communication through active follow-up to the main turn. Each author uses different terms, but Edelsky (1981) renames these two types of turn as "turn" and "floor," which Gallardo (1996) translates as "turno" and "palabra." In this way, one can have the turn from a structural perspective without having the floor.

One type of emission that is often considered a turn that only has a regulatory function in the conversation is the *back-channel*, whose identification is attributed to Yngve (1970). Yngve conceives it as the utterance of a person who occasionally takes the turn to show understanding or ask clarification questions while another person is speaking. Although there are prior mentions with other terms (Duncan and Fiske 1977, Fries 1952, Kendon 1967, Dittmann and Llewellyn 1967, 1968) or that reproduce the idea that these elements do not take the floor and merely ratify that of the speaker (Duncan and Fiske 1977, Edmonson 1981, Owen 1981, André-Larochebouvy 1984, Kerbrat-Orecchioni 1990, Gallardo 1996). Through this feedback, the listener avoids assuming the role of the speaker, so there is no true turn transition. Back-channels do not constitute a turn, following Yngve (1970), which means that the speaker retains the turn during the emission of the listener's back-channel and after it. Back-channels end when they stop being emitted, in contrast to turns, which are understood to continue and do not end until the next speaker takes the turn. Therefore, the overlap of the back-channel with another turn does not generate a simultaneous turn but rather simultaneous speech (Duncan and

Fiske 1977: 178). Duncan and Fiske (1977: 201-202) provide a detailed description of turns that can be considered listener back-channels, as do other authors (Oreström 1983, Stubbs 1987, Kerbrat-Orecchioni 1990, Gallardo 1996, Cestero 2000b).

As for the functional typology of the continuation, Kerbrat-Orecchioni (1990: 188-189) suggests that it is problematic to distinguish whether certain utterances are turns or a specific type of continuation (which she calls *regulators*): reformulation, evaluation, or commentary on the preceding intervention. In both CA (Conversational Analysis) and DA (Discourse Analysis), reactions are predictable in adjacent pairs, whether they are formed by turns or by movements in each model. In contrast, comments that do not explicitly demand a response, such as the *inform* movements in DA, may or may not receive *acknowledgment*, creating an exchange if they do. As Kerbrat-Orecchioni (1990: 201-202) also explains, there are certain acts in turns that demand a reaction from the interlocutor more than others.

The role of the speaker and listener is fluid, so the proof that the speaker is indeed so is based on the confirmation of their turn by the listener. In other proposals of hierarchical conversational units, this social distinction is not considered. In those, the intervention or turn are the same and are determined by the change of the speaker. However, Val.Es.Co. follows the path of studies that classify turns based on their propositional or metacommunicative value.

Through the social dimension, the Val.Es.Co. unit system brings a new perspective to the structure of conversation and indicates the role of the interlocutors at each moment. The interventions that characterize the speaker are the turns, whose main feature is that they are recognized or accepted by another interlocutor or interlocutors. "The consequence of this communicative view is that the listener is the true architect of the turns, who will always have an active role in the construction of the turn" (*trans.* Val.Es.Co. Group 2014). The confirmation depends on the relationship with subsequent turns by other speakers. The relationships of ratification between the types of Val.Es.Co. interventions are described in the chapter.

Regarding the incorporation of multimodality into this conversation structure, AC has indeed developed multimodal studies that consider posture, gaze, or gestures in turn-taking and sequential development (§3.2.3). However, structural segmentation models, which have the advantage of using a hierarchy of units to systematically describe the placement of phenomena in conversation, have not considered other visual modalities and

their effects on this hierarchy. The only exception is the Co-enunciation model, which has integrated gaze and manual gestures into its system, containing demarcative and enunciative functions complementary to prosody (Bouvet and Morel 2002, Morel 2010a, 2010b, Morel and Vladimirska 2014). However, this model focuses on internal units of intervention and does not classify interventions regarding their place in interaction or dialogue, as the other models developed in this chapter do.

The Val.Es.Co. conversational unit system contemplates the analysis not only of verbal interventions but also non-verbal ones. "We can find interventions that are gazes, gestures or body movements and paralinguistic expressions, such as changes in tone of voice, coughs, or laughter" (*trans.* Val.Es.Co. Group 2014: 18). So far, Val.Es.Co.'s only multimodal approach at the dialogic level has been the incorporation of paralinguistic elements, "a sign system complementary to the linguistic" (*trans.* Pons Bordería 2022), in the Val.Es.Co. 3.0 Corpus, consisting of audio-recorded conversations.

In a preliminary study (Cabanes Pérez 2020) to the current one, the analysis of a 13-minute conversation was conducted first with only auditory information and then a second analysis was performed incorporating visual information to compare structural and social changes at the dialogic level: 32 new interventions were added to the 261 interventions from the auditory analysis, and 16 new turns were added to the 184 turns. These changes occurred due to improvements in transcription and the identification of speakers through lip-reading, the identification of receivers through body orientation, and the creation of new exchanges through interventions that consisted solely of gestures. In total, there were 25 gestural interventions, 3 of which were turns, and 6 of them confirmed verbal turns that were not validated in the auditory analysis.

These results support the current research because, by using the Val.Es.Co. conversation segmentation model—which not only identifies interventions but also distinguishes socially ratified turns—, it can be observed how gestures and other multimodal information reveal new interventions and exchanges that remain hidden in the auditory analysis. As well as changes in social ratification that can have a global impact on how the development of the conversation and the relationship between the interlocutors are interpreted.

The goal of this chapter is to review studies on multimodal interaction and the multimodal elements that can be part of interaction: manual gestures, facial expressions, gaze, actions in the physical context, etc. This information aims to complement the review in §2 about the social and structural characteristics of conversation and turn-taking, so that multimodality is seen as an approach that contributes to a more comprehensive description of what happens in conversation. Additionally, this review also helps extract elements that contribute to the dialogic segmentation of conversation at both structural and social levels and that are part of the analysis in §5.

This chapter, first, explains what multimodality is (§3.2.1), what modes should be considered in multimodal interaction (§3.2.2), and the different approaches from which visual elements of interaction have been studied (§3.2.3). From this information, it explains, on one hand, the interactive functions of posture and gaze (§3.3), and on the other hand, the semiotic, temporal, and functional characteristics of other kinesic elements (§3.4).

Multimodality is a multidisciplinary term for any medium in which multiple modes of communication are used. Communication is constructed through various modalities such as verbal, paralinguistic, kinesic, proxemic, etc., which complement each other. The definition we adopt for our study is from Mondada (2013b: 580): "the integrated study of all the relevant linguistic, embodied, and material resources participants mobilize for organizing social interaction in an audible-visible intelligible way."

Certain types of movements have been generally discarded as communicative (self-adaptors or actions such as smoking, eating, knitting, or manipulating an object) (Kendon 2004: 8-9). However, any physical action can serve a function in conversation, so no physical movement or form can be ruled out directly without considering the context (Bavelas et al. 2016, Streeck 2009a: 18, Poyatos 1994: 134). Other features for distinguishing a gesture with communicative intention, in addition to context, include movement orientation, effort in execution, or synchrony with speech (Enfield 2009). Gesture and speech are also linked through synchrony, assuming that they are part of a unified, unique, and directed meaning (Kendon 2004).

Currently, there are two major frameworks in gesture studies, one of cognitive orientation and the other of interactive orientation (Cooperrider 2017, Vajrabhaya and Pederson 2018). Both are functional because they start from the premise that the structure of language reflects the objectives of language users. Most functional approaches agree

that language simultaneously serves a cognitive function, organizing and representing thought, and a social function, enabling the coordination of one's behavior with that of others (Kok et al. 2016).

In the cognitive approach to gestures, they accompany speech as an involuntary window into the mind, allowing an explanation of the cognitive process governing speech (McNeill 1985, 1992, 2005). The interactional approach focuses on gestures as a basic tool for carrying out social action (Goodwin 1981, Mondada 2006, 2007, 2013a, 2013b, Deppermann 2013, Goodwin 2018, Streeck 2009a, 2017).

In this chapter, these approaches are explained, as well as the linguistic approach, which is a minority perspective that considers issues of meaning and functions within the utterance, as well as the pragmatic contributions of gestures (Kendon 2004, Enfield 2009, Müller 2013, in prep., Payrató and Teßendorf 2014). In the chapter, some studies in the Hispanic context are also mentioned, which are more oriented towards teaching as a foreign language or the cognitive approach.

Among the mentioned visual elements, a distinction is established both in the literature review in this chapter and in the analysis itself: on one hand, posture and gaze; on the other hand, the rest of the kinesic components. This division is due to the formal, temporal, semiotic, and functional characteristics of each type of movement. In fact, these are not elements that are typically analyzed with the same parameters or studied together. With the aim of identifying body movements that can be considered turns on their own, manual, facial, or head gestures approximate the characteristics of verbal utterances: brief temporal limitations and a specific illocutionary force. In contrast, gaze and posture are movements that begin at a specific moment but extend over several verbal utterances or sequences, so it is not clear its segmentation as interventions. They also do not have a clear meaning, although it could be said that they have a metacommunicative function that affects two levels. At a macro level, they establish the space of interaction, the participating interlocutors, and their degree of commitment to the activity. These orientational and postural configurations of bodies toward each other in space for interaction have been called *ecological huddle* (Goffman 1961) or *F-formation* (Kendon 1990). Posture and gaze are the fundamental elements for indicating this orientation (Goffman 1963). Through displays of engagement, such as gaze and bodily orientation, actions become relevant for conversation (Goodwin 1981: 95-97). This behavior also differs both between speaker and recipient and among primary and secondary recipients

(Kendon 1973, Lerner 2003, Battersby 2011, Jokinen et al. 2013). At a micro level, gaze and posture contribute to the local management of each turn, along with other parts of the body. Regarding the regulatory function of gaze, Kendon (1967) provides an explanation of gaze direction that has served as the basis for microanalysis of gazes in conversational interaction. In the chapter, this is contrasted with other related (Goodwin 1981) and current studies (Bavelas, Coates, and Johnson 2002, Oertel et al. 2012, Jokinen et al. 2013, Rossano 2012, Streeck 2014). Changes in posture are also observed associated with both the turn and the sequence and the entire interaction.

The complexity of describing the entire formal variety of gestures with the same theoretical basis is the reason for this division between gaze/posture and the rest. In §3.4 and subsequent sections, the characteristics and functions of the rest of the gestures (manual, facial, etc.) are explained. Gestures have certain semiotic, temporal, and functional peculiarities that allow for different classification systems, which are often confused. Therefore, instead of explaining the typologies of the main researchers separately, we focus on the parameters by which gestures are described: the degree of conventionalization (§3.4.1), dependence on speech (§3.4.2), and functions (§3.4.3).

Based on their degree of conventionalization, we distinguish between:

1. Non-conventional or spontaneous gestures (Efron 1970 [1941], Ekman and Friesen 1969, McNeill 1992, Kendon 2004), which are more dependent on the sequential or verbal context and the iconicity of their form to understand how they contribute to the meaning of the verbal utterance.

2. Conventional gestures or emblems (Efron (1970 [1941], Ekman and Friesen 1969, Poggi 2002, Payrató 2003, Kendon 1981, 1988, 2013), which are lexically encoded and do not depend on speech for their comprehension, as their form allows for the identification of their meaning.

3. Recurrent gestures (Müller 2010b, 2017, Ladewig 2014a), which are in an intermediate degree of conventionalization due to their less strict and schematic formal articulation and their recurrence in similar contexts.

The more or less conventional relationship between the form and meaning of gestures is of interest because it relates to the identification of the gesture's function in interaction and its ability to appear accompanying verbal utterances or on its own.

Regarding the temporal dependence of the gesture on speech, various aspects are discussed in the chapter, such as the temporal phases of the gesture (Kendon 2004), its synchronization with speech (McNeill 1985, 1992, Kendon 1972, Enfield 2009), its synchronization for turn management (Goodwin 1981, Müller 1998, Streeck 2009b, Bohle 2007, Li 2014, Mondada 2015), and non-co-verbal gestures (Kendon 2004, Ladewig 2014a, Payrató 2016), which do not accompany speech. The temporal relationship between the gesture and the verbal turn is important to consider its structural position in the hierarchical model used for conversation analysis in this study.

In terms of functions, it is assumed that the functional perspective is the most suitable for understanding the role the gestures play in interaction. Due to the multifunctional nature of gestures, criteria such as their form and sequential context must be considered for their classification within a function; there is no closed repertoire. The functions are explained based on the classification by Kendon (2004) and Müller (1998, 2013, in prep.), inspired by structural linguistics. However, previous authors they are based on and subsequent ones are mentioned. They have used different nomenclatures, added specific characteristics, or opposed in various ways.

First, a distinction is made between referential functions and pragmatic functions. Referential gestures have semantic content and contribute to the meaning of the utterance by repeating the meaning of words, restricting it, adding other meanings, or representing an object. In contrast, pragmatic gestures do not have referential meaning or propositional content but have a procedural use that "embody the illocutionary force or the communicative action" (Müller 2008: 225). Pragmatic gestures are subcategorized into discursive gestures (or parsing), modal gestures, and performative gestures. We also add a fourth type, interactive gestures, because their function, as described by other authors (Ekman and Friesen 1969; Bavelas et al. 1992, Cosnier and Vaysse 1997, Streeck 2005, Bohle 2013, Wehling 2017), does not fit into Kendon's or Müller's typology.

Discursive gestures reveal how the speaker organizes the content of their discourse, shaping the flow of speech without referring to objects or ideas. They often align with prosodic rhythms (Müller 1998, Kendon 2004). These gestures can take various forms, including manual gestures like "batons" (Efron 1970 [1941]) or "beats" (McNeill and Levy 1982, McNeill 1992), the "Palm Up Open Hand" (PUOH) gesture used to present

an abstract object (Müller 2004), facial expressions that emphasize, underline, or modulate the content (Ekman 1976), head movements (McClave 2000), and more.

Modal gestures contribute to the interpretation of speech as hypotheses, assertions, quotations (Kendon 2004: 159), or they qualify it as obvious (Müller 2004), irrelevant, precise (Müller in prep.), humorous, mistaken (Bavelas and Chovil 2018), uncertain (McClave 2000), among others.

Performative gestures can accompany a verbal act to contribute to the interpretation of its illocutionary force (Kendon 2004: 159) or be the act itself if they do not accompany a verbal act (Kendon 2004: 225). Some gestures or gesture families that have been described as performative in the literature include the "Open Hand Supine" (Kendon 2004: 265), the "Away family" (Bressem and Müller 2014a, 2014b), the cyclic gesture (Ladewig 2014b), the shrug (Givens 1977, Morris 1994, Streeck 2009a, Debras 2017), head nods from left to right and up and down (McClave 2000), and certain facial expressions (Bavelas and Chovil 2006), among others.

Interactive gestures are related to turn management in conversation but do not carry semantic content. Some identifying features include their orientation towards the listener (Bavelas et al. 1992) and their placement at relevant transition points (Streeck and Hartge 1992). The chapter subdivides the explanation of this interactive function into the following functions: self-selection, turn maintenance, turn yielding, interruption, elicitation of the listener by the speaker, and listener feedback gestures.

Self-selection typically involves preparatory facial movements, but also manual deictics (Mondada 2007) and others that signal the initiation of speech to claim a turn (Duncan 1973, Kendon 1973, Schegloff 1996). It occurs just before the possible completion point and not only secures the right to speak but sometimes also anticipates the semantic and/or pragmatic meaning of the act (Streeck 2009b). The speaker achieves turn maintenance by keeping the initiated gesture to retain their speaking rights (Bohle 2014) or by repeating the gesture to demonstrate that a sequence is not yet complete (Streeck 2009b). Turn yielding can be indicated by the cessation of manual gestures, relaxation from a tense hand position (*speaker turn signal*, Duncan and Fiske 1977: 183), or self-adaptors (Li 2014: 158-162). Interruption is common and often signaled by the Open Hand Prone 'vertical palm' (OHP-VP) (OHP-VP) (Kendon 2004, Bressem and

Müller 2014a, 2014b1, Kamunen 2018). This gesture involves an open hand with the palm facing upward directed towards the recipient, and the back of the hand facing the gesturer. Elicitation from the listener typically requests a relevant response to the sequence or feedback from the listener. Visual forms used for this purpose include pointing, eye gaze, the aforementioned POUH, the speaker's nod of approval, facial gestures, and more. Feedback gestures are quite common in the listener. Nodding the head up and down are the most frequent form, however, other forms include citation gestures (Bavelas et al. 1992), PUOH (Bavelas et al. 1992, Kendon 2004, Müller 2004) or facial expressions (Bavelas et al. 1992). Feedback can even include content that contributes to the speaker's turn (Bolden 2003, Hayashi 2005, Iwasaki 2009).

This chapter helps formulate proposal for visual elements that are part of the visual segmentation of conversation in Chapter 5.

CHAPTER 4

Before explaining the methodology of this study in Chapter 5, this brief Chapter 4 addresses some issues related to data collection methodology and the multimodal annotation used for the field of gestures. It contextualizes the use of video in the study of conversation (§4.2), mentions some existing corpora (§4.3), discusses the challenges of formal annotation of gestures, and presents some annotation and transcription systems that serve as a basis for this study (§4.4).

The aim of study pertains to the structure of colloquial conversation, so it is necessary to allow the interlocutors the freedom to engage in conversation without predetermined conditions. Naturalistic data, recorded in the same environment where everyday activities take place, are particularly advocated by Ethnomethodology and Conversation Analysis. This is because the actions of the participants and their timing depend heavily on the context and the frame of participation, which should not be orchestrated or disrupted by researchers (Mondada 2013c: 987).

There are not many corpora of video-recorded Spanish conversations (see Table 10). Until recently, the only freely accessible one was the *Corpus de conversaciones para E/LE* from Columbia University (Ruiz Fajardo 2016), which was created for didactic purposes. One of these conversations was used in a preliminary study (Cabanes Pérez 2020) to conduct an initial analysis of the structure of a video-recorded conversation. In

2022, the GestINF corpus (Rodríguez Rosique and Cifuentes Honrubia 2022, Rodríguez Rosique 2021) was published online, which contains eight one-hour conversations recorded with high-quality video and audio in a laboratory setting. Three other corpora of Spanish conversations in video should be mentioned, although they are not freely available. These include the Corpus del Grupo de Investigación de Lingüística Aplicada (COGILA) (Barros 2012), the Nijmegen Corpus of Casual Spanish (NCCSp) (Torreira and Ernestus 2012), and the Freiburg SofaTalks (FST) (Pfänder). The COGILA corpus from the University of Granada includes ten conversations among Andalusian students. The NCCSp corpus is part of the Language Archive at the Max Planck Institute for Psycholinguistics and involves 52 Spanish speakers in casual conversations, totaling 30 hours of high-quality video and audio recordings. The FST corpus, developed at the Romanistics Seminar of the University of Freiburg, contains over 200 audio and video recordings in various languages, including different varieties of American Spanish.

Lastly, in Annex I, the characteristics of other video-recorded conversation corpora in different languages created for multimodal interaction analysis are available. Spanish video corpora share many characteristics with colloquial conversation, such as interpersonal purpose, experiential relationship, and non-specialized themes. The Columbia ELE corpus (Ruiz Fajardo 2016), COGILA (Barros 2012), and FST (Pfänder) were recorded in everyday spaces to promote colloquiality. In contrast to foreign corpora recorded in a laboratory setting, these Spanish corpora use monitoring equipment of lower quality, including fewer cameras capturing interaction from various angles and head microphones, among other devices. None of the Spanish corpora have annotated gestures with a formal annotation system (unlike some foreign corpora), which means that researchers must undertake this task themselves.

Regarding the annotation of gestures, non-verbal elements that occur simultaneously are usually annotated with one or more lines parallel to the transcription. Transcriptions are inherently selective and partial. By imposing a written format on language, some temporal relationships can be lost. However, they are clearly useful for identifying elements of interest and presenting them in written formats like this one. Annotation systems created for this purpose can be more descriptive/global or more interpretive/restrictive. Depending on the level of detail or granularity of gesture transcription, we can talk about annotation if it only includes certain characteristics of interest for the study's objectives. Concerning their temporality and phases, some

annotation systems include all three levels of gesture —gesture phrase, gesture phase, and gesture unit— because their study involves the synchronization of speech with gesture. Some coding schemes opt for a "phonological" perspective, based on the significant articulatory components of sign languages. In contrast, others describe the formal features that are meaningful for the study of interpersonal communicative functions, such as the MUMIN coding (Nordic Network for MultiModal Interfaces) (Allwood et al. 2007), primarily used in the NOMCO corpus.

Our system (§5.5.3.) is primarily based on MUMIN because is also interested in interpersonal functions. The formal coding is carried out using 21 attributes that describe the behavior of different parts of the body (face, eyebrows, eyes, gaze, mouth, head, body, shoulders, hands) (Table 11). Only expressions with specific communicative value are annotated, such as those providing feedback or used in turn management. In addition to formal coding, MUMIN includes attributes and values for interpersonal functions, semantic gesture typology (based on Peirce's theory: indexical, iconic, symbolic), and the relationship with verbal content. However, this study only take the formal coding for its own system.

Because the formal description of manual gestures is not as exhaustive in MUMIN, the system used in this study also uses codes from LASG (Linguistic Annotation System for Gestures) (Bressem et al. 2013, Bressem, Ladwig, and Müller 2013). The LASG annotation system follows a linguistic-semiotic approach that assumes a heuristic separation between form, meaning, and function. It focuses on a systematic analysis of gesture form, which is independent of the utterance to avoid description through intuition-based paraphrases. The coding is divided into hand configuration, gesture orientation, gesture movement, movement qualities, and gesture position.

Lastly, the chapter explains Mondada's system (2018), which does not establish specific codes for each body movement but proposes a way to organize annotation in a written transcription to relate the different modes that synchronize in interaction. It can be applied to any behavior (gesture, gaze, posture, object manipulation, etc.) and coding system.

CHAPTER 5

This studio also supposes a methodological contribution because the methodology for collecting a corpus of real conversational videos and annotating them can serve as a guideline for future multimodal studies interested in this genre, at this structural level, or in these gestures.

Chapter 5 develops the methodology used to achieve the objectives, justifies decisions regarding the creation of a corpus of real colloquial conversations in video and the variables for the two analyses, and explains the annotation of the material for analysis and the work phases performed.

To verify which multimodal elements affect the dialogic segmentation of the conversation, the analysis is based on the comparison between the segmentation of audio-recorded conversations and the segmentation of the same video-recorded conversations. The only difference between the two materials is the modality: one is purely auditory, the other includes audio and video. In this way, visual information generates a different interpretation of what happens in the conversation at the dialogic level.

From the segmentation of the material, two separate analyses are carried out:

1. The analysis of the structural and social changes that occur when incorporating visual elements: it is verified that the series of visual elements proposed in this chapter produce changes in dialogic segmentation; it is assessed which ones have the greatest effect on segmentation and what type of effect; the effects of multimodal elements on structural (creation, modification, or elimination of interventions) and social (creation, modification, or elimination of turns) segmentation are described; the relevance of visual elements in dialogic segmentation is demonstrated by the relationship between the changes produced between the first and second segmentation (without and with video).
2. The analysis of gestures that affect dialogic segmentation: the functional characteristics of gestures are described and related to their form (body part and configuration) and their position in the structure of the intervention; for this purpose, a formal annotation of gestures is proposed for the Val.Es.Co transcription system (Briz and Val.Es.Co Group 2002a), both on paper and in video annotation software.

To conduct this study, we follow the following work phases, which correspond to the order of sections in this chapter:

1. Creation of the corpus of conversations (§5.2)
2. Transcription and segmentation (§5.3)
 - a. Oral segmentation (audio without video)
 - b. Verbal and sound transcription
 - c. Segmentation into interventions and turns
 - d. Visual segmentation (audio with video)
 - e. Updated verbal transcription
 - f. Segmentation into interventions and turns, taking into account all visual elements
3. Analysis 1: Causes of change between verbal and visual segmentation (§5.4)
 - a. Visual elements causing change
 - b. Segmentation of independent and co-verbal gestures
 - c. Annotation of changes
 - d. Analysis
4. Analysis 2: Form and function of independent and co-verbal gestures (§5.5)
 - a. Form annotation
 - b. Macrofunction annotation
 - c. Specific function annotation
 - d. Analysis

In this research, focused on the structure of colloquial conversation, it is important a data as close as possible to the prototypical colloquial conversation. To preserve the naturalness of the data, the ideal situation is for the informants to be unaware that they are being recorded. However, concealing the recording equipment raises ethical issues and affects data quality because visible cameras and microphones cannot be used. Although conscious recording is generally more practical, the Val.Es.Co. Group has created a protocol for secretly recording audio conversations (Pons 2022) for its corpora (Val.Es.Co. and Ameresco).

To address the ethical issue, the researcher informs the speakers that they may be recorded in the coming weeks, obtaining their prior consent. However, at the time of recording, the participants are unaware of it. They sign two authorizations, one before recording and another afterward to validate the release of the material after listening to it

(Annex II). Since there is no corpus of real secret video-recorded colloquial conversations in Spanish, which is necessary to achieve the goal of this work, one has been created. The same secret recording protocol as Val.Es.Co.'s has been applied in this research, but with the disclosure that it is an audiovisual recording. Due to practical complications, the researcher participates in all conversations. The solution to prevent data contamination is to separate her gestures and the changes they have produced in the segmentation from the analysis.

The familiar or friendly relationship among the participants and the location in familiar spaces have facilitated the colloquial features of the encounters. Creating our own corpus does not allow for an extensive sample of speakers. The corpus consists of eleven informants (names anonymized), in addition to the confederate, divided into five conversations ranging from 16 to 36 minutes (totaling 131 minutes). Sociolinguistic characteristics of the participants were not considered in the analysis due to the small corpus size. It is composed of approximately 28,300 words of transcription and 453 transcribed gestures.

In order to ensure that all elements of the conversation are recorded, careful attention has been paid to the positioning of the speakers and the angles that the cameras can capture. The selection of equipment has been limited to the best quality options within cameras that are small or discreet enough so that participants do not perceive them.

The verbal transcription and the conversational unit segmentation system is an adaptation of Val.Es.Co.'s transcription system (Briz and Val.Es.Co. Group 2002a) and their unit system (Briz and Val.Es.Co. Group 2003) to the ELAN annotation software, following the steps of the Val.Es.Co. Corpus 3.0 (Pons Bordería 2022). This transcription and segmentation work is done twice for each conversation, first from the audio recording and then from the video recording.

Regarding the analysis methodology, the first analysis focuses on the structural and social changes that occur when incorporating visual elements. Its main objective is to confirm that the changes produced by multimodal segmentation are sufficient to justify the inclusion of visual elements in describing dialogical structure and its social dimension. The objectives derived from this are to describe the changes between oral segmentation and visual segmentation:

1. To verify that the series of visual elements proposed in this chapter produce changes in dialogical segmentation, both at the structural and social levels.

2. To describe the effects of multimodal elements on structural segmentation (creation, modification, or elimination of interventions) and social segmentation (creation, modification, or elimination of turns).
3. To assess which elements have the greatest effect on segmentation and the types of changes they produce.

The hypotheses are based on a preliminary study in which the same methodology was used to compare the segmentation of a 13-minute conversation using Val.Es.Co.'s dialogical units from both an audio recording and a video recording of the same conversation (Cabanés Pérez 2020).

The dependent variables in the first analysis of this study are the change in the structural dimension, the change in the social dimension, and the scope of the change. The independent variables are the causes of these changes.

Structural change involves the emergence of new interventions and the change in the type of an existing verbal intervention in the audio segmentation (the verbal intervention changes to initiative, reactive-initiative, reactive, or independent). Change in the social dimension is related to the validation of the intervention, i.e., whether an intervention becomes validated in the visual segmentation and becomes a turn or, conversely, an intervention considered a turn is no longer so. However, there may be interventions that change type without changing their validation; for example, if an Iind (independent intervention) changes to Ir (reactive intervention), it remains unvalidated. Additionally, new interventions that appear in the segmentation also belong to the social dimension and can be validated or not, so they can represent new turns in the conversation. Therefore, the two values of the structural change variable (new intervention and modification of intervention) are divided into the six values of the social change variable (Table 16).

Regarding the scope variable, interventions are not isolated entities. The actions taken in one intervention can modify the segmentation of another intervention with which they form an exchange. Thus, a visual element can cause changes in the intervention in which it appears, in another intervention, or in both. The scope variable is applied to each value of the structural and social change variables to create new values (Table 17).

The independent variables in this study are a series of causes that create structural and social changes, both in the intervention in which they appear and in other adjacent

verbal interventions. Based on the preliminary study, the literature, and the contrast between the oral and visual segmentation of the recorded conversations, the variables identified include independent gestures, coverbal gestures, body orientation in verbal interventions, extracommunicative actions, and lip reading. There can also be combinations of these causes.

Extracommunicative actions refer to physical activities such as eating or picking up an object that can modify the structure of the conversation without being gestures with semantic or procedural meaning. Lip reading can also help understand an unintelligible auditory fragment or correct a misinterpretation, affecting the segmentation of verbal interventions and their structural and social relationship with other interventions.

A significant methodological decision has been the distinction between gestures and body orientation. As previously discussed in the theoretical framework, when gestures are referenced in this analysis, they are movements of the head, face, shoulders, and hands that, due to their communicative effects, are included in Val.Es.Co.'s system of units, rather than gaze and posture. It is essential to understand that we do not deny that gaze and posture have communicative functions in interaction (as discussed in §3.3, they serve interactive functions related to participants' engagement in the conversation and turn-taking management). The distinction from other bodily movements is a structural matter.

While other bodily gestures can be categorized using Val.Es.Co.'s position schema (as seen in Table 19), gaze and posture have unique characteristics. Firstly, they involve movements with a much broader temporal scope. While manual, facial, or head gestures align with the temporal constraints of verbal interventions —having brief temporal limitations and specific illocutionary force—, gaze and posture are movements that initiate at a specific moment but extend across multiple verbal interventions or sequences. This makes it less clear how to segment them as separate interventions. Additionally, it is debatable whether gaze or posture, unaccompanied by a verbal intervention, have the illocutionary force to qualify as acts, unlike other independent gestures that can stand alone as separate interventions. It is worth noting that, unlike movements involving other body parts, the eyes serve as the means of perception through which individuals observe each other's behavior, and gaze also functions as an open channel of communication. Two people can look at each other and decide not to start a conversation or initiate a new

dialogue. While there is indeed communication happening within this mode, it lacks content. These gazes serve as scaffolding for initiating linguistic interaction.

A simple gaze from an interlocutor can indicate attention, that the speaker is being heard, but it does not validate the speaker's contribution as relevant content in the conversation. In contrast, more explicit reactions, such as verbal content or explicit gestures involving the head, face, shoulders, or hands directed at the first speaker, particularly towards the end or near the end of the intervention, do validate it. These elements react to the turn and validate it because they attend to it "in a manifest and simultaneous manner" (*trans.* Briz 2000, Briz and Val.Es.Co. Group 2002a, Val.Es.Co. Group 2014). Perhaps gaze and posture represent a lower level of validation, not as strong as verbal validation. After all, if no attention, not even a gaze, is received, the speaker stops talking or somehow requests the listener's attention. However, passive listening cannot carry the same weight as an explicit reaction. As Val.Es.Co. explains, "it should be considered that such unaccepted interventions have been physiologically perceived (heard) by the other interlocutor/s, that is, there have been receivers, but by not accepting it, they are not considered listeners" (*trans.* Val.Es.Co. Group 2014: 28). In other words, merely providing space for an intervention to be uttered and heard does not validate it. Thus, gaze can be considered a tool of listening that, by itself, does not validate. As Schegloff (1982: 78) points out, if maintaining eye contact were considered sufficient for attention and validation, there would be no need for additional phatic expressions like "uh-huh" and nods. Towards the possible end of a turn, expressions of understanding and posture-related responses to what has been said are relevant (Schegloff 1982: 81), not achievable through gaze alone.

Therefore, gaze and posture do not fit into the Val.Es.Co. units like other movements because they function differently, and this is evident in their temporal segmentation. This difference is not problematic attending to the multimodal nature of communication. Just as prosody does not have a segmentable meaning and relies on verbal content, gaze and posture also work in parallel with and are linked to interaction.

This distinction does not mean that information about interlocutors' orientation has been disregarded in the analysis; it is considered another cause that can alter the interpretation of an intervention as a turn. Body orientation is a modality that establishes the interaction space, the participants in the conversation, and their level of engagement in the activity (§3.3.1). Therefore, for this segmentation, it is an indicator of the direction

of interventions between the speaker and the listener. While speaking, people orient their bodies toward the listeners. Hence, in this segmentation, the speaker's orientation can sometimes change the interpretation of the verbal intervention, determining who is part of the exchange when only auditory context might make it unclear. This value pertains to the orientation of interlocutors during a verbal intervention since orientation during a gesture is a parameter of the gesture's form itself (§5.5.2.). It is not evaluated separately as a cause for changes in segmentation.

On the other hand, one of the key proposals of this dissertation is the distinction between independent gestures and coverbal gestures. The Val.Es.Co model has analyzed discourse units, primarily discourse markers, based on their position within it. This same type of analysis could be applied to gestures to examine how they affect the structure of the conversation or its social dimension (Table 19). There are gestures that do not accompany any type of verbal content from the same interlocutor (§3.4.2.4), so they themselves constitute the sole act of the intervention. It is an act because it contains illocutionary force linked to the form of the gesture and its use in context. It is also an intervention because it responds to the interactive dynamics of being or provoking a reaction. To address this gesture occurring independently of the intervention during the analysis, the terms "independent gesture", in contrast to the "coverbal gesture," or "gestural intervention," in contrast to the "verbal intervention", have been coined.

The term "coverbal gesture" refers to the gesture that is part of an intervention that includes verbal content. Observing Val.Es.Co's position schema, a gesture could be an act that occupies the initial, medial, or final position of an intervention or a subact in the initial, medial, or final position of an act. The difference between an act and a subact lies in the interpretation of the element as possessing its own illocutionary force or as a smaller informational constituent that requires others to form the act. However, even though gestures and subacts may fulfill similar functions that complete the content of the act at a semantic or pragmatic level (substantive subact - referential gesture; textual adjacent subact - discursive gesture; modal adjacent subact - modal gesture; interpersonal adjacent subact - interactive gesture), it cannot be claimed that they are two modalities of the same category because the relationship that occurs between verbal information and between verbal and gestural information, due to the nature of each modality, is different. The gesture can also be performed, and often is, simultaneously with speech, just like intonation. As two different modes of communication, they can occur at the same time,

breaking the linearity limitation of speech. Therefore, a new position is proposed: simultaneous position.

To simplify the analysis, it is only considered the distinction between independent gestures, i.e., gestural acts in an independent position from the intervention, and coverbal gestures, i.e., gestures in a coverbal position of the intervention or act, whether they occur during, before, in the middle, or after other subacts or verbal acts. The internal positions of the constituents of verbal intervention are not a priority for analyzing changes at the dialogic level; it is enough to note that they are coverbal.

Since in this analysis only gestures that affect the segmentation of interventions in the conversation are segmented, all independent gestures are segmented because they themselves are new interventions, but only some coverbal gestures are segmented, those that modify the structural interpretation of the intervention (such as iniciative intervention, Ii, reactive intervention, Ir, reactive-interactive intervention, Ir-i, or independent intervention, Iind) or its social interpretation (intervention or turn, I or T. In section §5.4.3, the decisions made in some cases where the identification between independent and coverbal gestures required additional criteria are presented: reaction criteria, duration criteria, and intensity criteria, and the gestural intervention with two independent gestures.

In the first analysis, several visual factors are considered to explain the structural and social changes in conversation segmentation. Some of these visual factors are not conversational units (lip reading, body orientation during verbal intervention, and extralinguistic actions); however, independent gestures and coverbal gestures can be part of conversational units: independent gestures as acts and interventions, and coverbal gestures as acts or subacts (§5.4.2). However, the position characteristic alone does not explain how these gestures modify segmentation. The pragmatic function they fulfill and the chosen bodily form to carry it out are variables that help us characterize gestures that affect dialogic segmentation.

Therefore, in the second analysis of this dissertation, the position of the gesture in the intervention, the function, and the form of the gestures are related, and in the case of independent gestures, the type of intervention they create is examined. To determine the functions of gestures, "Methods of Gesture Analysis" is used, which is based on the form of the gesture to systematically reconstruct the meaning of the gesture (Bressemer, Ladewig, and Müller 2013, Müller 2010b; Müller, Bressemer, and Ladewig 2013; Müller,

Ladewig, and Bressemer 2013). Each aspect of the form of the gesture is treated as potentially meaningful and non-random. This method is especially useful for the analysis of recurring gestures (§3.4.1.2.), most of which are pragmatic. It is divided into four phases (Bressemer, Ladewig, and Müller 2013, Ladewig and Bressemer, 2013):

1. Form analysis,
2. Analysis of the sequential structure of gestures in relation to speech and other gestures,
3. Analysis of the local context of gesture use, i.e., its relationship with syntactic, semantic, and pragmatic aspects of speech, and
4. Distribution of gestures in different usage contexts.

The first phase involves determining gesture units as signs motivated by their semantic-cognitive characteristics, temporal limits, the connection between gestures, and a detailed description of their form to detect patterns, without considering speech (Bressemer 2021: 14-15). In this case, while all independent gestures have been analyzed, not all coverbal gestures are included in the analysis. The selection is based on the visual information that modifies the segmentation of conversational units performed with auditory information, so it is necessary to listen to the verbal utterances accompanying the gestures. The form of manual gestures is described using the annotation system also used by the Methods of Gesture Analysis (MGA); Linguistic Annotation System for Gestures (LASG) (Bressemer 2013b; Bressemer, Ladewig, and Müller 2013) (§4.4.2) adapted (§5.5.4); other body parts are annotated using MUMIN (§4.4.1) adapted (§5.5.3), as it is an annotation already created for describing interactive gestures.

In phase 2, gesture analysis is sensitive to context. It observes how gestures are temporally related to speech. In this case, it is an advantage the fact that the position of gestures as independent or coverbal has already been annotated, as well as the type of intervention that the independent gesture generates structurally (Ii, Ir, Ir-i, Iind) and socially (turn/intervention). Then, in phase 3, the semantic, pragmatic, or syntactic functions that gestures have in the utterance in each particular case are analyzed (function selection in §5.5.4). Subsequently, in phase 4, the function of these gestures in different usage contexts is compared to distinguish the semantic core of the gesture from its local meaning and contextual variations (Bressemer and Müller 2014a: 1577-8).

Finally, once the most common positions, forms, and functions have been established, these variables can be related: the most common general and specific

functions in independent or coverbal positions, the most common functions based on their validation (turns or interventions), and the most common forms for these functions.

In the first analysis of this study (§5.4), the position of gestures generates two causes for changes in segmentation: independent gestures and coverbal gestures. Since in this second analysis the form and function of these gestures are being sought, the position becomes the dependent variable. Each independent gesture is segmented in ELAN as a conversation intervention, and the type of intervention it is (Ii, Ir-i, IrT, Ir, Iind) is annotated (see Table 23).

The head, face, shoulders, and hands are the four body parts considered as formal variables of gestures, along with the orientation of the gesture. Annotating the form of the gesture requires two main parameters: temporality and configuration. In ELAN, the boundaries of the gesture are placed temporally from the start time of the gesture to its end, following changes in the image. However, in a paper format, there is no temporal marker indicating when interventions or gestures begin and end. Overlapping symbols indicate the synchrony between interventions or modes, regardless of temporal measurement. In both formats, it has been decided not to fragment the gesture into its phases (§3.4.2.1.) because the goal is not related to semantic or prosodic synchrony between speech and gesture; it is sufficient to identify that the gesture is present.

For the objective, the attributes and values proposed by MUMIN (Allwood et al. 2007) are the most suitable, and an adaptation is proposed (Table 25). MUMIN has several attributes to describe manual gestures, but its values are concise because are designed for interactive gestures. Therefore, the LASG system (Bressemer et al. 2013) is adapted for hands because it is more detailed and can be used for referential and pragmatic gestures as well (Table 26).

The gesture transcription system explained is also inspired by Mondada's (2018) system. The main modifications to adapt it to the Val.Es.Co. transcription (Briz and Grupo Val.Es.Co 2002a) are:

- The use of black and gray colors to separate oral and visual information.
- Simplification of speech and gesture overlap symbols in square brackets only.
- To differentiate between independent and coverbal gestures, numbering the former as regular interventions with more spacing.
- For the succession of verbal and gestural emissions in the same turn, they are on the same line in different colors.

- For the temporal alignment of gestures by the same speaker, the same intervention numbering and less spacing are used.
- The \Rightarrow symbol from the PerLa Corpus (Gallardo and Sanmartín 2005: 49) is adopted for body orientation ($\Rightarrow A$, $A \Rightarrow \emptyset$, $\Rightarrow \emptyset$).

The functions of gestures are based on the classification previously presented in Chapter 3: first, the distinction between referential and pragmatic gestures, and within the pragmatic category, there are discursive, modal, performative, and interactive gestures. This typology is suitable for the purpose of the study because there are parallels with Val.Es.Co.'s conversational units that make up the intervention (Table 30).

Although these general functions are the first classification, it is descriptively useful identifying the specific functions of gestures in each specific context. Once the gestures are functionally classified, they are grouped, patterns are looked for in particular sequential contexts, and terminology is standardized based on the literature. So, these specific functions are not predefined now but are discovered through analysis. Similar to CA, cases have been observed without pre-established categories to avoid constraining interpretation to predefined molds.

Throughout this chapter, it is explained how all this methodological work of multimodal transcription, segmentation into conversational units, and annotation of variables has been recorded in ELAN (Table 31), as well as its visualization in a traditional paper transcription. This task already fulfills Objective 2 of this dissertation: incorporating the annotation of multimodal elements into Val.Es.Co. transcription (Pons Bordería 2022).

CHAPTER 6

Chapter 6 is divided into the results of the two analyses. In the first analysis, changes in the conversation are quantitatively analyzed between the audio recording and the video recording to assess the impact of visual elements on the analysis, determine the most frequent types of changes, and identify the elements that promote these changes. In the second analysis, the characteristics of the gestures that have led to changes in segmentation are presented qualitatively, considering their function, their position in the interaction, and their form.

CHAPTER 6.1.

The initial results pertain to the analysis comparing the segmentation of interventions with and without visual elements, considering changes in both the structural and social dimensions, and identifying the causes of these changes.

In the new visual segmentation, 2773 interventions remain the same as in the oral analysis, but 803 underwent a change (Graph 1). These changes occur either because the type of intervention or its validation have changed, or because new interventions emerge due to visual information. Since conversations vary in duration and the number of participants, the total number of interventions varies greatly in each conversation. However, in Graph 1, changes account for a percentage ranging from 13% to 31% within each conversation. This observation is not part of the initial analysis but raises a hypothesis for future research: the more speakers in a conversation, the greater the changes in visual dialogic segmentation. The hypothesis that there would be more changes due to new interventions than modification of the original ones is confirmed, as there are 539 new interventions (NI) and 264 modified interventions (MI).

In the social dimension of interventions, a distinction is made between validated interventions (turns) and unvalidated interventions (interventions). Using only auditory information, 3096 interventions are found, with 1783 being turns and 1313 not validated. However, in the segmentation based on visual information, there are 1971 turns and 1590 interventions. This represents an increase of 21% in interventions (Graph 3) and 10.5% in turns (Graph 4). The hypothesis that most new interventions are not validated is confirmed (Graph 5). Furthermore, interventions that change their type of intervention generally also change their social status to turn or non-turn (58%), while fewer only change their structural type of intervention (42%).

To explain the causes of the changes (independent gestures, coverbal gestures, body orientation during verbal intervention, lip reading, extralinguistic action), the focus is on changes that affect the structure of the conversation, i.e., the type of intervention based on its place in the interaction. It is distinguished whether the resulting change is a modification of a known intervention in auditory segmentation or the creation of a new intervention. It is also considered whether the change affects the speaker who performs the gesture/orientation/lip movement/action or another interlocutor who is affected.

In the preliminary analysis, there were more changes caused by independent gestures (57%) than by the rest of the causes (43%). In this larger corpus, the positions are inverted but also closer (independent gestures 49%, rest 51%). Thus, the hypothesis that it is the cause that provokes the most changes holds, but collectively, the rest of the causes contribute slightly more changes. It also remains true that new interventions are mostly generated by independent gestures (373), compared to 197 due to the rest of the causes. Notably, lip reading is the second most significant cause (124).

The observation that independent gestures generally result in fewer modifications to existing verbal interventions still holds, with only 17% of changes attributed to independent gestures (-7.701661 in the chi-square residue). This proportion is lower than what was found in the preliminary study (37%), which only considered modifications in social validation. Among the other causes, body orientation, followed by lip reading, has the most significant impact on modified interventions. It is worth noting the low number of extralinguistic actions (21).

These structural changes can be attributed to either the sender making the gesture in their own new or modified intervention, or these changes in one intervention structurally affecting another sender's intervention (Graph 7). The most frequent changes involve new own interventions (*nueva intervención propia*, NI-P), especially due to independent gestures, with lip reading in second place. Rarely (13 cases), independent gestures reveal new interventions by others. However, independent gestures are the cause of more modifications in other's interventions (*modifica intervención ajena*, MI-A) than the rest of the visual elements separately. Regarding coverbal gestures, there is a greater distinction between new own and others' interventions, while modifications in own or others' interventions are more evenly distributed. Both body orientation and lip reading often have a greater impact on own interventions. Although, in the case of orientation, the most common change is the modification of one's own intervention, and in the case of lip reading, it leads to a new own intervention. The total number of modified own and others' interventions is nearly equal, thanks to the large number of other's interventions modified by independent gestures. The rest of the causes (except actions) tend to affect own modified interventions more than others'. In general, independent gestures and orientation are the main causes of change, while lip reading, coverbal gestures, and actions provide support.

As for the causes of social changes (new interventions or turns, modifications from turn to intervention and vice versa) (Graph 9), the hypothesis that the majority of new interventions caused by independent gestures are non-turn interventions holds. This trend is also present in new interventions discovered through lip reading, albeit not as pronounced. This may be because reactive interventions tend to be spoken at a lower volume and are not captured in the recording. However, in the rest of the causes (coverbal gestures, orientation, and actions), although they have a smaller impact on new interventions, the results are reversed: the number of new turns is higher than interventions. Regarding modifications to existing interventions, generally, all causes result in a higher proportion of changes from intervention to turn ($I > T$) than from turn to intervention ($T > I$).

In conclusion, the changes between oral and visual segmentation are significant. Out of these changes, 67% are due to completely new interventions that did not exist in the oral segmentation. Of these new interventions, 65% are gestures in an independent intervention position. Among these, 83% are non-turn interventions. Therefore, 311 independent gestures in non-turn interventions account for 36% of all causes of changes in segmentation (311 out of 864). This effect due to this cause groups the highest percentage of changes in segmentation, but the combinations and other causes together sum up to 100%. The sum of other causes slightly outweighs all changes generated by independent gestures (51% vs. 49%). The least impactful cause is extralinguistic action (5% of changes), so possibly it is not a determining factor in conversation segmentation.

Lip reading is the second cause of new interventions in visual segmentation. Among the new interventions, most are not validated. The rest of the causes (especially body orientation) have a greater impact on the modification of existing interventions in oral segmentation than on the creation of new interventions. Furthermore, there are more modifications of interventions that also result in a social change, not just a change in the structural type of intervention (but orientation, lip reading, and actions maintain some symmetry). Of the changes in the validation of existing interventions, there are more changes from intervention to turn than vice versa. Finally, distinguishing between the effects on one's own intervention and another's intervention provides interesting data, such as the fact that independent gestures, although they generate many new interventions, tend not to modify the segmentation of other's interventions. In general,

creating a new intervention for someone other than the person performing the visual element is a minority compared to other types of changes.

CHAPTER 6.2.

In this second analysis, the functions that independent and co-verbal gestures fulfill in conversation are observed, as well as their forms. To classify the gestures in the corpus into pre-established functions (referential, performative pragmatic, interactive pragmatic, modal pragmatic, discursive pragmatic) (§5.5.4.), their position and their use in the sequence in relation to speech and other gestures is considered, and their form across all cases, looking for patterns to form classification criteria, as each factor on its own (form, position, sequential context) is not sufficient to establish the function.

First (§6.2.1.), the most frequent general functions annotated in the corpus are shown, as well as their distribution between independent and coverbal positions. Then, these functions are divided into minor functions found throughout the corpus and each one is explained to understand the reasons why gestures have been classified into one or another function.

Most of the independent and coverbal gestures that affect dialogic segmentation structurally and socially are pragmatic. The hypothesis that the interactive function is the most frequent (68%) is confirmed, with a significant margin over the other frequent functions: performative (19%) and modal (8%) (Graph 13). Due to the number of interactive gestures (314), it is the only function for which a typical behavior can be proposed. The other functions are anecdotal in a corpus of 131 minutes, so it can be affirmed that they do not constitute the primary changes in dialogic unit segmentation.

The classification of gestures in the corpus into these functions came from the formulation of minor functions that helped determine the general function to which they belonged. This process is part of phases 3 and 4 of the MGA (§5.5): first, gestures are analyzed in their local context, and then similar functions are compared in different usage contexts to distinguish a main function independent of contextual variants. Therefore, a series of functions that fit into the pre-established major functions are identified (Graph 14).

The performative pragmatic functions are question, assertion, negation, epistemic functions (their independent position is the criterion for classifying them as performatives rather than modals), and evaluation.

The interactive pragmatic functions include following, agreement and disagreement, closing third, acceptance of contribution, pre-dialogue contact, and turn management. Following, closing third, agreement, and acceptance of contribution are the most frequent, especially the first one, and their distinction depends primarily on the sequential context, as they often coincide in form: head nodding is their predominant form. The decision to include agreement and disagreement within interactive functions is based on the subact segmentation of the Val.Es.Co. model (Pons Bordería 2022), which includes these functions as interactive adjacent subacts because they confirm the recipient's stance as favorable or unfavorable to connect their intervention with the previous intervention. This procedural view of agreement is linked to the conversation as negotiation approach (§2.2.2), which requires agreement between interlocutors for it to progress.

Modal pragmatic functions include mitigation, intensification, attitude, lack of knowledge, understanding, and affiliation. The only discursive pragmatic function found is formulation. Pure referential gestures, which convey semantic content, have been differentiated from referential-pragmatic gestures, which combine other interactive or performative functions.

Following is the most frequent function (152 gestures), making the interactive function stand out, along with closing third (73 gestures) and agreement (36 gestures). Also, as performative and modal gestures, evaluations (52 gestures) and affiliation displays (31 gestures) are frequent since they are typical of the reactive listener's intervention. In contrast, there are fewer than 15 cases of each of the other functions (performative, modal, referential, or discursive), so it is not possible to validate a typical position, type of intervention, or form for these gestures, but hypotheses for future multimodal conversation studies can be proposed.

Secondly (§6.2.2), although the positions occupied by the gestures in the intervention and sequence have been shown when explaining the gesture functions, the function-position relationship is examined to find any patterns. There are more

independent gestures than coverbal gestures in total, and regarding the distribution of these in functions, 93% of the gestures occupy functions that have more independent gestures than coverbal gestures. Of the interactive and performative functions, there are a higher number (314 and 87), and over 80% of them are independent (Graph 16). These are the most solid data, while there are few cases for the others to ensure a trend. The micro-functions that always occupy the independent position are the epistemic performative function, the referential-pragmatic function, affiliation, and disagreement, although there are few cases (Graph 17). Precisely because these epistemic functions occupy that independent position, they become the sole act of the intervention and can respond to questions or doubts, while in the coverbal position, they have been classified as modal gestures. So, in these functions, the position-macro-function relationship is direct. Other functions whose independent gestures are in the 83%-70% frequency range are evaluation and assertion as performatives and pre-dialogue contact, agreement, and closing third as interactive functions (Graph 18). Functions with an equal number of independent and coverbal gestures are referential, acceptance of contribution, and discursive formulation, but there are few cases as well (Graph 19). Functions with a predominant coverbal position (Graph 20) are questioning and negation (performatives), turn management (interactive), and modal functions. Other interactive functions have more effects in the independent position, but precisely these manage verbal turns, so they are usually coverbal to them. Finally, except for affiliation, the rest of the modal functions, which are only 8 in the corpus, are always coverbal, as they affect the modal interpretation of the verbal intervention.

As for the segmentation of independent gestures as interventions, the vast majority, 318, are unvalidated reactive interventions; the other types of interventions are only occupied by 37 gestures. Although in the case of evaluation, the proportion is not as significant (21% validated - 79% unvalidated). The only functions, which are used to a lesser extent in the analysis, that do not follow this trend are assertion, negation, pre-dialogue contact, and semantic content, which in the independent position usually occupy validated interventions (Graph 21). The cause may be that these performative functions are part of adjacent pairs, as well as the referential function when it is the sole act of the intervention. In this corpus, pre-dialogue contact usually achieves its goal of establishing the start of the dialogue, so it is validated and not ignored.

It has also been found that a gestural intervention can not only consist of a single gestural act in an independent intervention position but can have two gestural acts (§5.4.3). This occurs in the corpus in 14 gestural interventions. All of them consist of two gestures, except for one complex case in which 6 gestures are combined. In 6 of these interventions, interactive functions in a reactive position are combined; another 4 combine functions in an unvalidated reactive position; 3 cases are in a reactive-initiative position.

In the third part (§6.2.3), all the data is observed from a formal perspective to determine which functions are typically performed with a specific body part or movement configuration. This helps in understanding the multifunctionality of gestures and finding the form as a criterion to distinguish certain functions or link similarities between them to types of similar functions.

The form of the gestures is closely linked to the particular meaning of each function, and for many of them, there are not enough samples to formulate prototypical forms. But there is a higher incidence of head and facial movements for interactive and performative gestures, which are the ones that most affect dialogic segmentation. The higher frequency of manual gestures in referential functions is also confirmed. There are more samples for interactive gestures, especially following gestures, but the analysis shows that the form is not a determining factor in distinguishing between interactive functions. Therefore, in the gestures of the corpus, form is a criterion that necessarily combines with sequential context to identify the function. It is not sufficient on its own because there are not enough samples or, precisely, interactive gestures depend more on sequential criteria.

Since there is a very broad functional typology, there is also a lot of formal variety. Some functions have very few cases (discursive pragmatic, modal pragmatic, referential, referential-pragmatic), so a trend in the configuration of the gestures cannot be asserted. The most used body parts (head, hands, face, etc.) are observed (Graph 22), but due to the variety of functions and the small number of cases, an analysis of specific forms would not be useful. The clearest result is that performative and interactive gestures are usually performed with the head, especially interactive ones; less frequently, the face, shoulders, or hands are used, in that order. Since these are the gestures with the most samples, these results are the most reliable.

Referential gestures, whether used to express semantic or pragmatic content, are typically performed about 60% of the time with the hands. This result is consistent with

the general description of referential gestures as deictic manual gestures or gestures that represent concrete or abstract objects or actions. There are 87 performative gestures, but their forms depend on the type of act they perform.

In Graph 23, the most frequent interactive functions (following, agreement, closing third, and acceptance of contribution) and the body parts with which they are performed are shown. Head movement alone is the most common in all functions, but especially in following. Gestures that combine head and face are 30, but there is an increase in the agreement function. There are only 19 facial gestures, almost 50% following and 50% closing third; agreement is not expressed with the face alone. In the rest of the combinations, there are only 1 to 4 cases, so generalizing is not possible. However, combinations that include hand movement tend to have an agreement function, and when shoulders or hands are involved, it is usually not for following. Simple nodding (a single nod of the head) and repeated nodding are the most commonly used movements in general. Usually, the form of these interactive gestures is not determinant in differentiating which specific interactive function the gesture serves but depends on the sequential context.

Finally (§6.2.4), with respect to the orientation of gestural interventions (Graph 25), in general, the orientation tends to be directed towards the interlocutor in all types of interventions. Although Kendon (1967) states that head orientation toward the interlocutor is more frequent during the role of the listener, the results cannot confirm that there is a significant difference between interventions that are turns and those that are not validated as turns.

In conclusion, performative functions (evaluation, question, affirmation, negation, epistemic function) generally appear in an independent position, although questions and negations tend more towards the coverbal position. Furthermore, gestural interventions formed by a question, an affirmation, or a negation tend to be turns (in contrast to the rest of the gestures in the corpus). Although more than half of the cases are performed with the head, this is the macrofunction in which the four body parts (head, face, shoulders, hands) are most evenly distributed. In modal functions, there is a significant difference between affiliation, which is always found in an independent position, and the rest (mitigation, intensification, attitude, comprehension, lack of knowledge), which are all coverbal. Regarding form, most are performed with the face and are affiliative; among the rest of the modal functions, the head and hands are used to the same extent. Referential

functions are evenly split between independent and coverbal gestures when they simply convey semantic content, but the balance tilts towards independent gestures when including referential gestures that adopt a pragmatic function in that position. This is where more hand gestures are found, as expected. There are barely two cases of discursive formulation functions, one independent and one coverbal, with the mouth and another with the hands. But it is not possible to make any hypotheses.

CHAPTER 7

In conclusion, this study explores multimodal interaction in conversations from a pragmatic structural perspective, a novel approach in this field. The Val.Es.Co. model, which categorizes conversational units hierarchically, is used to systematically analyze linguistic, pragmatic, and social aspects within conversations. This study incorporates gestures, specifically kinesic movements with intention and interactional functions, into the dialogical level of the Val.Es.Co. structure, which includes the concept of interventions and their pragmatic relationships.

By applying position theory to gestures, the study distinguishes between independent gestures (those not accompanying verbal interventions) and coverbal gestures (those integrated into verbal interventions). Gestures are related to the validation of interventions as turn units, a crucial aspect of the social dimension in dialogical structures.

In addition to its analytical findings, this study contributes methodologically by providing guidelines for collecting and annotating a corpus of real video-recorded conversations, which can benefit future multimodal studies in this domain. The corpus comprises 131 minutes of Spanish video conversations, preserving the characteristics of informal spoken discourse. The study employed a secret recording methodology, ensuring authentic data and the authorization of participants.

The research aims to investigate whether multimodality significantly influences the segmentation of conversation according to Val.Es.Co.'s dialogical units. The comparison between audio and video recordings of conversations segmented using dialogical units reveals a 22% difference, a statistically significant result. These changes consist of new interventions (67%) and modifications of existing ones (33%).

Notably, gestural interventions have the most significant impact on segmentation, accounting for the majority of changes. These interventions often serve reactive functions, indicating listener attentiveness to the speaker. However, their overall contribution to segmentation is 145 out of 459 gestures, which is a subset of the 803 changes driven by various visual factors (including gestures, lip reading, body orientation, and extralinguistic actions). Therefore, while independent gestures are crucial, other visual elements, positions, and functions also affect dialogic segmentation.

The study classifies gestures into five functional categories: referential, performative pragmatic, interactive pragmatic, modal pragmatic, and discursive pragmatic. The majority of gestures are pragmatic, with interactive and performative functions being the most prevalent (68% and 19%, respectively). These results align with expectations, as referential, modal, and discursive gestures typically act as coverbal elements within turns. However, there is a small percentage (8% modal and 4% referential) that impacts dialogic segmentation.

The study shows that gesture form is closely tied to function, but many functions lack sufficient samples to establish prototypical forms. Head and facial movements are prevalent in interactive and performative gestures, which have the most significant impact on dialogic segmentation. Manual gestures are more common in referential gestures. While there are more samples for interactive gestures, especially "following", form alone is not determinative in distinguishing between interactive functions. Contextual factors also play a crucial role.

In summary, this study advances the understanding of gestures in conversation by introducing a new theoretical and methodological framework. It offers insights into how multimodal information affects the segmentation of conversations and provides a valuable corpus for further research in this area. The study's broad approach encompasses various gesture functions and positions, shedding light on patterns between gestures with different characteristics. This complements narrower studies focused on specific functions or forms. Ultimately, this research enhances our understanding of multimodal interaction in conversations and opens avenues for future investigations.

CAPÍTULO 1. INTRODUCCIÓN

1.1. MOTIVACIÓN DEL ESTUDIO.....	59
1.2. OBJETIVOS DE LA INVESTIGACIÓN	60
1.3. ESTRUCTURA DEL TRABAJO	61

Go with the flow. Like in all empirical science, serendipity is a source of new, unexpected insights. In the field, you are in a world that belongs to other people. The dynamics of daily life can be like the changing currents of the surf: If a rip tide takes you, don't fight it. There's no point. (Enfield 2013: 977)

1.1. Motivación del estudio

Este trabajo asume que la conversación es un género que prototípicamente ocurre ante la presencia física de sus interlocutores (Briz 1996) y sus elementos no pueden estudiarse sin tener en cuenta esa realidad. Por tanto, la conversación no solo debe analizarse a través de su contenido verbal, ni siquiera por otros elementos paralingüísticos audibles, sino también a través de otros modos de comunicación, los recursos relevantes corporales y materiales captados visualmente, que forman parte del orden secuencial de la interacción social (Mondada 2014: 98).

El análisis de la conversación, que enfoca el habla como acción social, si bien en sus inicios se centró en el lenguaje verbal (Sacks et al. 1974), pronto abrió el camino a otros modos visuales que participaban en la interacción, como la mirada (Goodwin 1981); teniendo que hablar no de *talk-in-interaction*, sino de *talk-and-other-conduct-in-interaction* (Schegloff 2005). Selting (2013: 597) afirma la necesidad de integrar los estudios sobre la multimodalidad y la lingüística interaccional. En las últimas décadas, ha habido un desarrollo de estas investigaciones que contemplan los gestos y otras acciones que utilizan el cuerpo en la organización social secuencialmente estructurada de la conversación (Mondada 2013a, Goodwin 2018, Streeck 2009a, 2017).

Los gestos también se han analizado desde un punto de vista lingüístico, atendiendo a su semántica, a la relación entre forma y significado y su interpretación en el contexto comunicativo (McNeill 1992, Kendon 2004, Müller 2004, Cienki y Müller 2008). La pragmática, que considera los enunciados como acciones que tienen un efecto en su contexto de habla, también ha ampliado el concepto de acto a movimientos corporales que se combinan o no con el habla. Estos han recibido múltiples denominaciones, entre ellas, *visible actions* (Kendon 2004) o *composite utterances* (Enfield 2009).

Tomando en cuenta el contenido oral, alrededor de la unidad acto se han propuesto modelos de segmentación del discurso en una serie de unidades jerárquicas que explican

la estructura de la conversación. Entre ellas, el Modelo de la Co-enunciación ya incluye la mirada y los gestos, en conjunción con la prosodia, como indicadores para la segmentación del turno oral. Sin embargo, otros modelos de segmentación jerárquicos que incluyen unidades superiores al turno —que consideran la relación que se establece entre turnos, formando intercambios y unidades mayores— no han considerado la información visual en su segmentación. Nos referimos al Análisis del Discurso de Birmingham (Sinclair y Coulthard 1992), a la Escuela de Ginebra (Roulet et al. 1985, 2001) y al grupo Val.Es.Co. (Briz *et al.* 2003, Pons Bordería 2022).

Las unidades jerárquicas de estos modelos de segmentación de la conversación y su relación posicional han permitido describir el funcionamiento de elementos lingüísticos en el nivel dialógico, pero no se han aplicado para describir el comportamiento multimodal. Este enfoque estructural podría enriquecer el conocimiento sobre la multimodalidad y, a su vez, la multimodalidad completaría los estudios estructurales de la conversación que, sin contar con todos los elementos verbales y no verbales que forman parte de la comunicación, están sesgados. Este trabajo pretende contribuir en llenar el vacío que existe en estos dos campos.

Puesto que la cantidad y la variedad de gestos que contribuyen a la construcción de la conversación en todos los niveles de la jerarquía no es abarcable para este trabajo, en este estudio exploratorio nos centraremos en los elementos multimodales que afectan a la segmentación dialógica. El modelo de segmentación elegido es el de Val.Es.Co., que se distingue de los demás por la inclusión de dos dimensiones que explican el desarrollo de la conversación: la dimensión estructural y la dimensión social. En el nivel dialógico esto crea dos unidades: la *intervención* y el *turno*. Mientras la intervención es una emisión, generalmente asociada al cambio de emisor, que es o provoca una reacción, protótipicamente, lingüística; el turno es un tipo de intervención que es reconocida o aceptada por otro interlocutor o interlocutores.

1.2. Objetivos de la investigación

El objetivo general es describir la manera en la que la multimodalidad afecta a la estructura dialógica y a la dimensión social de conversaciones coloquiales en español. De este surgen una serie de objetivos específicos que enumeramos a continuación:

1. Describir los elementos multimodales que afectan a la segmentación dialógica estructural y social. Esto incluye la descripción formal, funcional y de posición de

- los gestos, así como la identificación de la información contextual visible que pueda cambiar la interpretación de las intervenciones estructural o socialmente.
2. Incorporar la anotación de elementos multimodales a la transcripción de Val.Es.Co. (Pons Bordería 2022) —cuya disposición refleja la segmentación dialógica—, tanto en papel como en programas informáticos de anotación de vídeo.
 3. Describir los efectos de los elementos multimodales en la segmentación estructural y social.
 4. Evaluar qué elementos multimodales tienen mayor efecto en la segmentación dialógica estructural y social.
 5. Demostrar que la segmentación a partir de los elementos verbales de la conversación refleja parcialmente la realidad de este tipo de discurso y que es necesaria la incorporación de los elementos multimodales para que la descripción a nivel dialógico sea completa.

1.3. Estructura del trabajo

El bloque I, que incluye los capítulos 2, 3 y 4, establece el marco teórico del estudio, atendiendo a las características del tipo de discurso y el tipo de unidades en los que pretendemos incorporar un enfoque multimodal, a las características de la propia multimodalidad y la manera en la que podemos obtener datos reales para el análisis.

En el capítulo 2, se define el género discursivo en el que se sitúa el objeto de estudio, la conversación coloquial, y el turno, como unidad dialógica que forma parte de una estructura de niveles jerárquica y que depende de un sistema social para su construcción conjunta entre hablante y oyente. Se justifica el modelo de segmentación conversacional del Grupo Val.Es.Co. como base para este trabajo, así como las ventajas de aplicarlo al estudio de la multimodalidad.

En el capítulo 3, se abordan los componentes multimodales que deben tenerse en cuenta en la interacción multimodal y se definen en virtud de sus características semióticas, temporales y funcionales.

En el capítulo 4, se revisa brevemente la metodología de otras investigaciones para la recolección de corpus y la anotación de la multimodalidad, con el objetivo de que sirvan de base para nuestra propia metodología.

El bloque II, formado por los capítulos 5, 6 y 7, muestra el estudio realizado: su metodología, sus resultados y sus conclusiones. La comparativa entre conversaciones grabadas en audio y grabadas en vídeo da lugar a dos análisis: uno sobre los cambios producidos en el nivel dialógico de la conversación al añadir la información multimodal y otro sobre las características de los gestos que producen esos cambios.

En el capítulo 5 se desarrolla la metodología empleada para lograr nuestros objetivos, se justifica las decisiones sobre la elaboración del corpus de conversaciones coloquiales reales en vídeo y sobre las variables de los dos análisis, y se explica la anotación del material para el análisis y las fases de trabajo realizadas.

El capítulo 6 se divide en los resultados de los dos análisis. En el primero, se analizan cuantitativamente los cambios estructurales y sociales producidos en la conversación entre la grabación de audio y la de vídeo para evaluar la incidencia de los elementos visuales en el análisis, qué tipo de cambios son más frecuentes y qué tipo de elementos los fomentan. En el segundo análisis, se presenta cualitativamente las características de los gestos que han producido los cambios en la segmentación, tomando en cuenta su función, la posición que ocupan en la interacción y su forma.

En el capítulo 7 se exponen las conclusiones generales extraídas de este estudio y se sugieren posibles líneas de investigación futuras.

BLOQUE I

CAPÍTULO 2. UN ENFOQUE ESTRUCTURAL Y SOCIAL DEL TURNO

2.1. INTRODUCCIÓN	67
2.2. LA CONVERSACIÓN: ENFOQUES PARA SU ESTUDIO	67
2.2.1. CONVERSACIÓN COMO CONDUCTA SOCIAL	68
2.2.2. CONVERSACIÓN COMO NEGOCIACIÓN	70
2.2.3. CONVERSACIÓN COMO ESTRUCTURA	72
2.2.4. ESTRUCTURAS JERÁRQUICAS DE LA CONVERSACIÓN	74
2.3. EL TURNO	79
2.3.1. EL TURNO Y SU CONSTRUCCIÓN SOCIAL.....	80
2.3.2. LA RELACIÓN PRAGMÁTICA ENTRE TURNOS.....	84
2.3.3. TURNOS DE HABLANTE Y TURNOS DE OYENTE	91
2.3.4. RATIFICACIÓN DEL TURNO	100
2.4. RECAPITULACIÓN: LA MULTIMODALIDAD EN UN MODELO ESTRUCTURAL DE LA CONVERSACIÓN	106

2.1. Introducción

Con el objetivo de incorporar los elementos multimodales en el estudio de la estructura de la conversación, en este capítulo introducimos, primero, lo que se entiende como el género discursivo *conversación* (§2.2), las perspectivas desde las que se estudia, más o menos centradas en su construcción social o en su jerarquía estructural; después, nos centramos en el *turno* (§2.3) como unidad que construye la conversación, a través de las relaciones pragmáticas que los interlocutores establecen entre los turnos, creando roles interactivos en cada momento, que pueden identificarse en el análisis de la conversación mediante el concepto de ratificación del turno de Val.Es.Co. Así, esta revisión nos ayuda a identificar en nuestro análisis cómo los elementos multimodales pueden participar en el proceso dialógico de construcción de la conversación (§2.4).

2.2. La conversación: enfoques para su estudio

Nuestro interés en el género de la conversación parte de la premisa de que el prototipo de comunicación humana es la conversación cotidiana (Briz 1998: 32). Otros géneros orales, sujetos a objetivos o contextos específicos, pueden compartir características con ella o diferenciarse, pero tomamos la conversación coloquial como punto de partida para esa comparación. Asumimos, entonces, que los gestos analizados en este estudio serán diferentes a los que se puedan encontrar en otros géneros interactivos orales, ya sea por su función, su forma o su frecuencia.

La definición de conversación de Briz (Briz 1996, 1998, 2010a, 2010b, 2012, Grupo Val.Es.Co. y Briz 1995) es exhaustiva en las características que la diferencian de otros géneros y porque esos criterios son los que ha seguido Val.Es.Co. (Briz et al. 2002, Pons Bordería 2022) para la recopilación de conversaciones coloquiales de la ciudad de Valencia (España); de manera que se facilita una futura comparación de nuestros resultados con este corpus. La conversación es un tipo de discurso que se caracteriza por una interlocución en presencia (cara a cara); inmediata (aquí y ahora); con toma de turno no predeterminada; dinámica, en la que la alternancia de turnos favorece la tensión dialógica; retroalimentadora, en la que se coopera con el tema de conversación y la intervención del otro (Grupo Val.Es.Co. y Briz 1995: 28-29). La falta de alguno de estos rasgos identifica a otros tipos de discurso, como la conversación telefónica (una tipología de conversación sin presencia), los mensajes de texto o audio (no inmediato), el debate

(el orden planificado de los turnos), la entrevista (a veces sin cooperación del entrevistador a lo dicho por el entrevistado), etc.

La conversación prototípica es coloquial, pero “conversacional” no es sinónimo de “coloquial” (Briz 1996: 25): son un tipo de discurso y un tipo de registro, respectivamente, que suelen coincidir; pero también hay conversaciones formales y textos escritos coloquiales. Para utilizar el registro coloquial se debe cumplir una serie de condiciones dadas por el contexto comunicativo. En primer lugar, la ausencia de planificación (o planificación sobre la marcha), la finalidad interpersonal o socializadora y el tono informal como rasgos primarios (Grupo Val.Es.Co. y Briz 1995: 29-30). En segundo lugar, los rasgos coloquializadores caracterizan a la conversación coloquial prototípica y son capaces de acercar el registro coloquial a una conversación que no lo es: la relación de igualdad entre los interlocutores (por sus papeles sociales o por la propia situación), la relación vivencial de proximidad (saber y experiencia compartida), el marco de interacción familiar (espacio físico cotidiano y relación concreta de los participantes), la temática no especializada (Grupo Val.Es.Co. y Briz 1995: 30-32).

Una vez definida la conversación, se puede estudiar desde distintos enfoques, que pueden ser complementarios: la conversación 1) como conducta social, 2) como negociación o argumentación y 3) como estructura (Briz 1998: 43). Aunque Cortés Rodríguez (2002: 172) argumenta que la distinción entre modelos estructuralistas y sociológicos de análisis del discurso ha perdido la utilidad porque, si bien originalmente las diferencias metodológicas eran mayores entre la Escuela de Birmingham y los etnometodólogos americanos, es clara la retroalimentación de ambas líneas en estudios como los de la Pragmática Dialógica (Escuela de Ginebra, Escuela de Lyon) o los de tipo sociológico (Pomeranz, Atkinson, Heritage o Wootton).

A continuación, desarrollamos las diferencias entre estos enfoques, pero advertimos que se pretende una revisión de la descripción de la conversación en general y que las cuestiones relacionadas específicamente con el turno se muestran con más detalle en §2.3.

2.2.1. Conversación como conducta social

Los estudios de la conversación en tanto que conducta social han sido capitaneados por los etnometodólogos del Análisis de la Conversación (AC). Presentan una metodología de trabajo y una teoría en la que convergen la sociología y la lingüística,

pero también la antropología y la psicología, por su acercamiento al habla como acción social (Schegloff 1991:46). Reciben influencia de la etnometodología de Harold Garfinkel (1967) y la teoría del orden de la interacción en la microsociología desarrollada por Erving Goffman (1967, 1983). Goffman se interesa por la interacción social como institución social, en tanto que existen unas prácticas y estructuras sociales en la interacción que la definen como un sistema organizado. Define la conversación como: “una actividad que se produce durante el tiempo en que determinado grupo de participantes se han acreditado entre sí para la conversación y mantienen un solo foco móvil de atención” (*trad.* Goffman 1971: 38-39).

El AC busca patrones interactivos en la organización de la conversación, que se explican mediante expectativas secuenciales orientadas por los participantes (Levinson 1983: 323). Sacks, Schegloff y Jefferson (1974: 700-701) observaron conversaciones reales para extraer 14 normas que describen este tipo de interacción en el que los participantes se coordinan para construir unos turnos cuyo contenido, tamaño y orden no está fijado previamente. Una noción de gran impacto en los estudios que tratan la interacción oral es el sistema de toma de turno o *local management system*: el proceso social por el que los participantes de la conversación gestionan la alternancia de turnos momento a momento —véase el sistema en detalle en §2.3.1—. Gracias a este sistema, se explican fenómenos propios de la conversación que no suelen aparecer en otros géneros dialógicos más planificados: transiciones de turno solapadas, colaboraciones solapadas, cruces y escisiones conversacionales, reinicios, autointerrupciones, reparaciones, solapamientos competitivos, etc.

A la propuesta inicial se le criticó la falta de diferenciación entre la conversación cotidiana y otros tipos de intercambio, así como la omisión de la retroalimentación del oyente para la regulación, de los elementos no verbales o de las variaciones producidas en diferentes situaciones comunicativas o grupos sociales (Philips 1976: 83). Sin embargo, estas son cuestiones que se han ido incorporando a este enfoque posteriormente. Por ejemplo, el término *habla institucional* (*institutional talk*) se refiere a interacciones en las que hay normas, roles y actividades de instituciones sociales particulares, frente a la conversación ordinaria (Raymond y Olguín 2022: 14-15). Hay que dirigirse a estudios de otras lenguas para encontrar las particularidades en la alternancia de turnos distintas a las del inglés; por ejemplo, en español, las alternancias de turno solapadas son más comunes que en inglés (Cestero 2000a, 2005).

Los estudios del AC centrados especialmente en las formas lingüísticas recurrentes en la interacción social, sirviéndose de la investigación y las terminologías lingüísticas, han derivado en la subdisciplina de la Lingüística Interaccional (Couper-Kulen & Selting 2001, 2018, Fox et al. 2013), aunque a menudo la frontera entre ambas disciplinas se diluye. Se centran en la forma lingüística empleada, pero en tanto que está afectada por las acciones.

El AC se distingue de otros paradigmas teóricos en que el énfasis no está en los actos de habla, sino en cómo estos se enmarcan en secuencias de acción social (Raymond y Olgúin 2022: 83) —véase §2.3.2 sobre la proyectabilidad de los turnos y las secuencias—. La pretensión de no influir en el análisis con categorías perjudica a esta disciplina en la falta de un metalenguaje que sistematice su teoría. Los diferentes autores del AC parecen manejar de manera diferente las nociones básicas (*preferencia, secuencia, turno*), por lo que “el cuerpo teórico resulta a veces poco sólido” (Gallardo 1996: 40). A pesar de esta confusión de categorías conceptuales¹, no se puede negar su influencia en estudios lingüísticos posteriores. Las convenciones sociales del turno, la selección del hablante o los efectos en la acción siguiente “eran ideas bastante originales y útiles a partir de las cuales el lingüista iba a incorporar a su concepción del enunciado verbal como actividad o acto, su concepción interlocutiva” (Cortés Rodríguez 2002: 193). Además, Levinson (1983: 319) defiende la metodología del AC por evitar la especulación y categorización inverificable e indefinidamente extensible de los analistas del estilo del Análisis del Discurso (§2.3.2).

2.2.2. Conversación como negociación

Esta construcción dialógica del discurso por parte de los interlocutores se ha considerado también una negociación que requiere cooperación y estrategia para conseguir la aceptación del otro.

Conversar es comunicar cooperativamente y negociar estratégicamente un propósito con el otro. De ahí que la conversación sea una especie de discurso retórico que persuade y busca la aceptación del otro de forma continuada (Briz 1998: 44)

¹ Coulthard y Brazil (1992: 54-55) también ejemplifican la duplicación de categorías que genera la falta de un marco descriptivo general al comparar la *insertion sequence* (Schegloff 1972) y la *misapprehension sequence* (Jefferson 1972): ambas son lo mismo, pero se han etiquetado desde diferentes perspectivas, una estructural y otra semántica, respectivamente. Para el Análisis del Discurso, el *par adyacente* es el único concepto que el AC ha descrito con detalle.

Esta visión tiene su base en nociones pragmáticas. La conversación sigue el *principio de cooperación* (H.P. Grice 1975, 1978), ya que la alternancia de turnos está regulada por los propios interlocutores para que el discurso avance. No obstante, las *máximas de cantidad, cualidad, relación o manera* no siempre se cumplen en la conversación coloquial (Briz 1998: 45) por que las estrategias de los participantes siguen otros fines como la *cortesía* (Lakoff 1973; Brown y Levinson 1978, 1987, Leech 1983, Haverkate 1994); de manera que el participante trata de reforzar lazos con el receptor, no imponerse a él, realzar su imagen, etc. Asimismo, las máximas de la cooperación o de la cortesía se pueden violar para favorecer una interpretación del enunciado en un contexto dado: principio de relevancia (Sperber y Wilson 1986).

Un análisis pragmático siempre se puede dar en el contenido lingüístico que forme parte de un discurso real, en el que siempre hay un emisor que codifica la información y un receptor que la descodifica. En una conversación este proceso es simultáneo y requiere de estrategia por ambos interlocutores, pues el hablante se adelanta a los efectos que su emisión provoca en el oyente y modifica su discurso según la reacción que recibe, y el oyente interpreta la intención del hablante y se anticipa para reaccionar adecuadamente. Una conversación no es un par mínimo de pregunta y respuesta, sino que se construye mediante turnos sucesivos. Según la teoría de la Argumentación (Anscombe y Ducrot 1983), las expresiones lingüísticas orientan el discurso hacia unas conclusiones y delimitan de este modo su continuación.

La Escuela de Ginebra (Roulet, Trognon, Auchlin, Rémi-Giraud, Moeschler, Zenone) se funda sobre una concepción del discurso como negociación entre los interlocutores, que lo construyen cooperativamente y expresan argumentos para llegar a un acuerdo (Roulet *et al.* 1985; Moeschler 1985). Desde su modelo, cada intercambio pretende conseguir la *complétude interactionnelle* (Roulet *et al.* 1985) o *complétude dialogique* (Roulet *et al.* 2001), es decir, que la intervención iniciativa reciba la reacción favorable o preferida que el hablante pretendía y se produzca un doble acuerdo que cierre la negociación². Además, la negociación también trata de alcanzar la *complétude interactive* (Roulet *et al.* 1985) o *complétude monologique* (Roulet *et al.* 2001), o sea, que durante el intercambio se asegure una buena transmisión del mensaje y se respete la

² En la jerarquía de Val.Es.Co., esta interacción produce el cierre de la unidad *diálogo* (Grupo Val.Es.Co. 2014: 31), mediante una última intervención que es una *reacción tópico-discursiva* (Padilla García 2020).

imagen de los interlocutores³. Si la negociación no cumple estas restricciones, se genera una negociación secundaria para conseguir la reacción deseada o clarificar la intervención, creando una organización jerárquica recursiva.

Esta negociación sirve de base para la creación de un modelo estructural de la conversación, por lo que la Escuela de Ginebra también se incluye en los enfoques estructurales (§2.2.3.) y su modelo se explica y compara con más detalle en §2.2.4.

2.2.3. Conversación como estructura

En cuanto a la estructura de la conversación, tanto desde el enfoque social como desde el negociador/pragmático se han formulado descripciones estructurales más o menos complejas. El AC contempla una relación a nivel local entre cada turno, pero también una relación secuencial que agrupa varios turnos y tipos de secuencias que pueden enmarcar algunos tipos de conversaciones. Aunque estas estructuras crean una descripción sistemática del funcionamiento de la conversación, debido a la asunción de la contingencia de la producción del AC, estas no permiten explicar la relación posición-función de todos sus elementos (Pons Bordería y Fischer 2021). Es decir, que en la conversación hay ciertas construcciones cuya posición y relevancia no se puede explicar mediante un análisis en el que las emisiones se encadenan unas tras otras de forma puramente lineal, por lo que las propuestas de enfoque pragmático han desarrollado estructuras jerárquicas del discurso (Stubbs 1987:182).

Estas jerarquías se asimilan a las de las unidades lingüísticas, en las que unidades inferiores conforman unidades superiores; como fonemas o morfemas que forman palabras, sintagmas que forman una oración o elementos extraoracionales sujetos a las condiciones de la enunciación y que han sido objeto de la pragmática. Sin embargo, la unidad superior que estudia la gramática es la oración, porque no hay restricciones en la combinación de oraciones que puedan definir de la misma manera la estructura del párrafo, aunque haya enlaces cohesivos entre oraciones⁴ (Coulthard y Brazil 1992: 59). En concreto, los fenómenos observables en la conversación no se pueden explicar desde la sintaxis oracional, la cual solo puede calificarlos como pobres, descuidados o

³ La *complétude interactive* se basa en las restricciones de sistema y en las restricciones rituales de Goffman (1981).

⁴ No nos vamos a referir a la otra rama del AD, la de los gramáticos del texto, que analiza el texto como una concatenación de oraciones unidas por conectores ampliando los conceptos de la sintaxis (Levison 1983); ya que no es factible para la conversación.

incompletos (Narbona 1988). Por tanto, es necesario un sistema con otra perspectiva para las unidades conversacionales, de la misma manera que no se clasifican las unidades fonológicas desde una perspectiva léxica o sintáctica (Coulthard y Brazil 1992: 60-61).

Desde un enfoque pragmático, los modelos de segmentación del discurso aplican sistemas de unidades jerárquicos para segmentar la conversación en unidades de rango mayor y menor que superan los límites de la sintaxis. Generalmente, utilizan el acto como unidad mínima, “asimilándolo a la unidad básica de la pragmática enunciativa que se desarrollaba paralelamente desde Austin o Ducrot” (Gallardo 1996: 79). Aunque sus unidades de análisis son pragmáticas, utilizan la metodología y los principios teóricos de la lingüística para describirlas: el aislamiento de un grupo de categorías básicas o unidades del discurso, y la formulación de una concatenación de reglas sobre esas categorías (Levison 1983⁵). Frente a la perspectiva inductiva del AC, esta segmentación jerárquica es deductiva, es decir, que una vez creadas unas categorías previas, los fenómenos de la conversación se encajan en ellas (Cortés Rodríguez 2002: 175); su punto débil es que no siempre cumplen los rasgos fundamentales del aparato teórico (Cortés Rodríguez 2002: 179).

En la lingüística románica, la segmentación del discurso ha sido de gran interés desde 1980 (Pons Bordería 2014). Pons Bordería cataloga los modelos de segmentación, primero, por su enfoque basado en la sintaxis o en el uso pragmático y, después, por su enfoque prosódico. Los modelos sintácticos aplican la sintaxis al análisis del discurso oral añadiendo nuevas funciones y etiquetas que expanden las unidades sintácticas; bien añadiendo una capa superior, o sea, una macrosintaxis —como el Grupo de GARS (Blanche-Benveniste y Jeanjean 1987)—, bien añadiendo nuevas funciones en las periferias a derecha e izquierda de la oración. Los enfoques basados en el uso consideran que la gramática descriptiva no puede ser adecuada como punto de partida, así que toman una base pragmática. Dentro de este enfoque se encuentran el modelo de Friburgo (Berrendonner 1990), el de la Coenunciación (Morel & Danon-Boileau 1998), el de las Unidades Básicas del Discurso Basic (*BDU*) (Degand y Simon 2009), el de Ginebra (Roulet et al. 1985, 2001) y el de Val.Es.Co. (Briz *et al.* 2003, Grupo Val.Es.Co. 2014, Pons Bordería 2022). El de Friburgo, el de las *BDU* y el de la Coenunciación basan la segmentación en el criterio prosódico, utilizando la entonación para definir unidades

⁵ Aunque esta descripción se refiera al AD en contraste con la metodología del AC, se puede aplicar a otros modelos de segmentación.

menores de forma precisa, sin embargo, no sirven para describir unidades mayores al turno (Pons Bordería y Fischer 2021).

El modelo de Ginebra y el de Val.Es.Co. desarrollan también la jerarquía de unidades en el nivel dialógico⁶, es decir, en unidades que explican las relaciones entre emisiones de distintos hablantes. Desde Birmingham, el Análisis del Discurso (Sinclair y Coulthard 1975, 1992, Coulthard y Montgomery 1981; Stubbs 1983) también incluye una estructura dialógica. Puesto que nuestro estudio de los gestos se ubica dentro de este nivel dialógico, vamos a desarrollar estas jerarquías en §2.2.4 antes de centrarnos en la descripción del turno en el apartado §2.3.

2.2.4. Estructuras jerárquicas de la conversación

La jerarquía del AD sirve como inspiración para el primer modelo de Ginebra (Roulet et al. 1985), junto a otros aportes como la estructura jerárquica recursiva de Pike (Pike y Pike 1983) y la relación entre los constituyentes de la intervención de Ducrot (Ducrot et al. 1980) y de Frank (1980) (Roulet et al. 2001:25). El modelo del AD y el de Ginebra inspiran al modelo Val.Es.Co. –junto al AC norteamericano, el grupo de la Sorbona (Morel y Riolland 1992) y grupo de Aix-en-Provence (Blanche-Benveniste y Jeanjean 1987)–. En los tres sistemas existe una organización escalar entre las unidades que forman la estructura de la conversación (Tabla 4). A simple vista, pudiera parecer que hay una correspondencia alta entre las unidades, pero los tres sistemas tienen características bien distintas.

En el rango 1 se encuentran las unidades superiores. La *lección* (si se analiza el discurso educativo) o la *interacción* (en otros contextos) del AD se produce prototípicamente entre transacciones de saludo y de despedida (Francis y Hunston 1992); mientras que el *discurso* de Val.Es.Co. depende del contexto interactivo del encuentro: si cambia el número de interlocutores, su papel funcional, su grado de igualdad jerárquica,

⁶ Algunos trabajos del Grupo Val.Es.Co. que se centran de forma más específica en las unidades dialógicas son Briz (2006, 2007a, 2007b), Padilla (2004, 2020). Su validez se demuestra en los estudios sobre la conversación en los que se han empleado las unidades monológicas (Pons Bordería 2016, Pascual Aliaga 2018) y las dialógicas, que demuestran las diferencias estructurales entre la conversación y otros géneros dialógicos: la conversación durante una partida de un videojuego MOBA (Badia Climent 2018, 2023), la entrevista periodística y la entrevista sociolingüística (García Ramón 2020) y el diálogo de una serie (Hernández Ruiz et al. 2022). El Corpus Val.Es.Co. 3.0 incluye una selección de 16 conversaciones que han sido segmentadas mediante estas categorías conversacionales por el consenso de al menos dos analistas (Pons Bordería 2022).

	Rango 1	Rango 2	Rango 3	Rango 4	Rango 5
Análisis del Discurso (Sinclair y Coulthard 1992)	lección	transacción	intercambio	movimiento	acto
Módulo jerárquico (Roulet et al. 2001)			intercambio	intervención	acto
Orden estructural del modelo Val.Es.Co. (Briz y Grupo Val.Es.Co. 2003)	discurso	diálogo	intercambio	intervención	acto

Tabla 1. Unidades del AD, del módulo jerárquico y de Val.Es.Co.

el registro (planificado/no planificado) o la dinámica de toma de turno (predeterminada/no predeterminada), cambia el discurso (Estellés y Pons 2014). En el segundo modelo de Ginebra (Roulet et al. 2001: 58), se suprimen de la estructura jerárquica las unidades superiores del primero (Roulet *et al.*1985), la *incursión* y la *transacción*⁷, que se resisten a una descripción sistemática, para incluirse en la estructura praxeológica (módulo referencial).

En el rango 2, la *transacción* del AD es una unidad temática, al igual que la *secuencia* del AC. En cambio, el *diálogo* de Val.Es.Co. es una unidad estructural basada en la combinación de intercambios sucesivos limitados por una primera intervención iniciativa y una intervención reactiva al final⁸. Al contrario que la secuencia del AC, el diálogo es una unidad estructural, no temática. Por tanto, el cambio de tópico podría ser un indicio, pero el criterio esencial para diferenciar diálogos es la falta de vinculación pragmática entre las dos partes de la emisión. Por sí mismo, el cambio de tema no inicia un nuevo diálogo: “La introducción de un tema nuevo relacionado con algo anteriormente dicho, sentido o interpretado, así pues, sin ruptura aparente, es también un acto reactivo” (Briz 1998: 58).

⁷ Roulet (1985: 23-24) ya dice que la *incursión*, delimitada por el encuentro y separación de dos interlocutores, no es una unidad muy desarrollada porque no se puede explicar lingüísticamente. Aunque señala que suele presentar tres constituyentes: un intercambio subordinado con función de apertura, un intercambio principal con función de transacción y un intercambio subordinado con función de clausura.

⁸ En Grupo Val.Es.Co (2014) se señala que prototípicamente la intervención reactiva al final es turno, pero no siempre: la intervención iniciativa turno es el único límite seguro.

El *intercambio* de Val.Es.Co., al igual que el de Roulet, está formado por dos intervenciones de distintos hablantes; aunque el segundo permite que el intercambio se extienda con más intervenciones, por lo que puede llegar a ser un diálogo para Val.Es.Co. Inicialmente, el AD solo se planteó dos rangos: 1) el *enunciado* como todo lo dicho por un hablante antes de que otro empiece a hablar; y 2) el *intercambio* como dos o más enunciados de diferentes hablantes (Sinclair y Coulthard 1992: 2). No obstante, detectaron que dos partes de un enunciado podían pertenecer a intercambios diferentes; por lo que se creó la unidad *movimiento*, que puede ser de iniciación, de respuesta o de retroalimentación.

También en Val.Es.Co. se contemplan ciertos casos en los que una misma emisión de un interlocutor incluye dos intervenciones diferentes. En esta intervención compuesta⁹, una parte puede ser una reacción a lo anterior y otra el inicio de una nueva secuencia dialógica, desvinculada intencional o temáticamente del diálogo previo (Grupo Val.Es.Co. 2014: 22). El hecho de que dos partes de la emisión pertenezcan a dos intercambios diferentes podría interpretarse como la unidad movimiento del AD; sin embargo, parece que esta división solo se produce para dividir dos diálogos, no dos intercambios dentro del mismo diálogo.

En el segundo modelo de Ginebra (Roulet et al. 2001), se incluye la unidad *intervención* entre el intercambio y el acto. La *intervención* de Val.Es.Co. (2014) tiene en cuenta el valor ilocutivo en relación con las intervenciones de otros interlocutores y el cambio de voz es un indicio de este hecho. Se describirá con más detalle esta unidad y su lugar en el intercambio en el próximo apartado (§2.3.).

En los tres sistemas, en el rango inferior se encuentra el *acto*. Esta unidad coincide a menudo con la unidad gramatical *cláusula*; sin embargo “Grammar is concerned with the formal properties of an item, discourse with the functional properties, with what the speaker is using the item for” (Sinclair y Coulthard 1992:8). La descripción de *acto* del AD no es la misma que la de filósofos de la Teoría de los Actos de Habla, que estudian las funciones psicológicas y sociales del acto externas al discurso que se está produciendo; por el contrario, el AD describe los actos según las funciones internas que desempeñen

⁹ Gallardo (1996: 91) también diferencia la respuesta/inicio (R/I) de la intervención mixta (R+I), ya que esta última consta de dos movimientos con distinta orientación interaccional. El primer movimiento puede ser una respuesta o una reacción a una intervención previa y el segundo un inicio o un informe. En cambio, en el enunciado de la R/I no se pueden descomponer estos dos valores.

dentro del propio discurso, incluyendo actos que funcionan como señales manifiestas de organización del discurso (Stubbs 1987: 153). Son, precisamente, los actos que permiten la estructuración de la conversación mediante repeticiones, énfasis, atenuación y operaciones “de relleno” en general los que constituyen el foco del AD y no los actos que expresan contenidos proposicionales, porque permiten conocer la estructura subyacente del discurso o las funciones subyacentes de las emisiones individuales (Stubbs 1987:177).

El modelo original de Ginebra (Roulet et al. 1985) sí utilizaba la concepción del acto de habla de Searle (1972), como combinación de fuerza ilocutiva y contenido proposicional. Pero en el segundo modelo (Roulet et al. 2001: 59-67) lo abandonan a favor del acto textual, basado en el concepto de enunciación de Berredonner (1990), como registro en la memoria discursiva, porque es un tratamiento cognitivo de la información que se amolda bien a la concepción de Roulet de la interacción como negociación. Esta unidad textual mínima depende más de la organización de la actividad discursiva que de la estructura de la lengua.

Como hemos mencionado, otros modelos de segmentación del discurso describen el acto utilizando también criterios prosódicos. Val.Es.Co. define el acto como “la mínima unidad de acción e intención, que posee las propiedades de aislabilidad e identificabilidad en un contexto dado” (Grupo Val.Es.Co. 2014: 37). Es decir, que posee fuerza ilocutiva propia y presenta ciertas marcas lingüísticas, prosódicas y semánticas que lo limitan.

Los modelos del AD y de Ginebra clasifican los actos en directores o *heads* (con un valor pragmático dominante) y en subordinados o *pre-head/post-head*. Por ejemplo, en el AD, los *heads* pueden estar precedidos por un marcador (*ok, well, look*) y/o por un *pre-head*, que suele ser un *starter*, que da información necesaria para que aparezca el acto director, o un marcador para respuestas (*receive*), y pueden estar seguidos por un comentario u otro acto vinculado a la función del director (Francis y Hunston 1992: 134). En la Figura 1 se observa un ejemplo de la estructura de una intervención formada por actos según el modelo de Ginebra: en un intercambio en el que se trata un problema con el precio de unos billetes de avión, en la intervención *ah eso eso ya me acuerdo yo había pagado ese precio allá*, el acto principal es *yo había pagado ese precio allá*, es la idea

principal que quiere transmitir el hablante; el resto de los constituyentes son actos subordinados¹⁰.

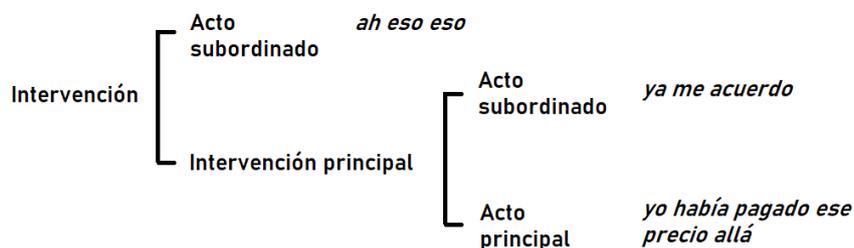


Fig. 1. Ejemplo traducido de segmentación de una intervención (Roulet, 2001:85)

En cambio, el modelo Val.Es.Co. realiza esta vinculación subordinante en unidades internas del acto: el *subacto*. De esta forma, en Val.Es.Co. (2014), el acto es la mínima unidad de acción e intención, pero se construye mediante segmentos informativos identificables, habitualmente, mediante marcas semánticas y prosódicas, que ya no forman parte de la dimensión estructural de la conversación, sino de la informativa.

Esta jerarquía de unidades se encuentra en la dimensión estructural (Val.Es.Co.) o en el módulo jerárquico (Roulet), pero tanto una como otra por sí mismas no explican todo lo que sucede en la conversación, sino que se combinan con otras dimensiones o módulos. El segundo modelo de Ginebra aspira a describir la organización del discurso basándose en tres tipos de restricciones: las situacionales, ligadas al universo de referencia y a la situación de interacción; las lingüísticas, ligadas a la sintaxis y el léxico; y las textuales, ligadas a la estructura jerárquica del texto (Roulet *et al.* 2001: 44). A partir de estas restricciones se crean cinco módulos –referencial, interaccional, sintáctico, léxico, jerárquico– en los que se descompone el discurso. Cada uno presenta un sistema de informaciones simples con una forma de organización específica y nocionalmente independiente, es decir, que pueden ser descritos sin referencia a los otros sistemas (Roulet *et al.* 2001: 42); sin embargo, estos módulos se combinan dentro de la producción y la interpretación del discurso. Esta compleja organización se denomina *heterárquica*. Desde esta perspectiva, los debates sobre la definición de unidades discursivas no deben centrarse exclusivamente en categorías textuales y sus fronteras con las unidades léxicas y sintácticas, sino que deben adoptar una perspectiva transversal (Roulet 2001:217-218).

¹⁰ Por la recursividad característica de este modelo, los constituyentes facultativos (marcadores discursivos) pueden relacionarse con otras unidades (superiores) que no son el acto (Roulet 1985:34-40; Roulet 2001:45, 90).

Aunque el modelo Val.Es.Co. se inspira en el ambicioso modelo de Roulet et al. (2001), no combina tantos módulos. Diferencia tres dimensiones u órdenes: estructural-interno, social-externo e informativo (Tabla 5). Las unidades estructurales son las que se acaban de comparar con los modelos previos; las unidades sociales, *turno* y *alternancia de turno*, permiten estudiar el reparto de papeles comunicativos durante la conversación; el *subacto* se vincula con la continuidad semántico-informativa y, además, “permite enlazar un sistema de unidades de base funcional con una sintaxis oracional, aspecto este de vital importancia, puesto que, aunque ambos constructos son de naturaleza diferente, no deberían considerarse carentes de relación” (Grupo Val.Es.Co. 2014: 13).

NIVEL	DIMENSIONES		
	ESTRUCTURAL	SOCIAL	INFORMATIVA
Dialógico	discurso diálogo intercambio	alternancia de turnos	
Monológico	intervención acto	turno	subacto

Tabla 2. Sistema de unidades (Briz y Grupo Val.Es.Co. 2003)

Puesto que nuestro interés está en encontrar los efectos de la multimodalidad en el nivel dialógico de la conversación, una vez visto qué es la conversación y los diferentes enfoques para su estudio, en el siguiente apartado de este capítulo nos centramos en el *turno*. Utilizamos los enfoques ya desarrollados de manera combinada para describir tanto la posición del turno en la jerarquía como las relaciones pragmáticas y sociales que lo construyen, ya que el modelo de segmentación que seguimos en este estudio, el de Val.Es.Co., se inspira en los modelos anteriores y los combina en una descripción estructural y social del turno.

2.3. El turno

Se ha podido observar que cada enfoque del estudio de la conversación describe sus unidades desde distintos parámetros. El objetivo central de este capítulo es explicar la unidad intervención del modelo Val.Es.Co., su lugar en la estructura jerárquica de la conversación, la manera en la que los turnos se relacionan entre sí en la estructura y sus efectos sociales. Es pertinente no solo restringirnos a las publicaciones del Grupo Val.Es.Co. sobre esta unidad, sino también revisar los otros modelos de los que parte su concepción de la intervención. Pretendemos así una visión completa de la unidad que nos

permita justificar la idoneidad del modelo Val.Es.Co. para evaluar los efectos de los gestos y otros elementos visuales en el análisis de la conversación. Asimismo, podremos recurrir a las teorías vistas en este capítulo para explicar algún fenómeno multimodal que así lo requiera.

Por tanto, la descripción del turno de este apartado no se divide por enfoques, escuelas o modelos, sino que los rasgos en los que nos centramos se explican mediante el contraste o las semejanzas de los modelos necesarios, ya conocidos en §2.2. Los rasgos en los que se divide el apartado son: 1) el turno y su construcción social (§2.3.1), en el que nos centramos en su segmentación temporal, que depende de su construcción social; 2) la relación pragmática entre turnos (§2.3.2), marcadas por las relaciones de proyectabilidad y reacción que ejercen entre sí; 3) los turnos de hablante y los turnos de oyente (§2.3.3), pues los interlocutores alternan estos roles para sostener la tensión dialógica de la conversación, en consecuencia, nos centramos especialmente los turnos que emite el oyente para sostener las estructuras de las secuencias; y 4) la ratificación del turno (§2.3.4), como herramienta que utiliza el Grupo Val.Es.Co. para explicar la vinculación entre la estructura de los intercambios y los roles sociales de los interlocutores.

2.3.1. El turno y su construcción social

Los investigadores de la comunicación, ya sean lingüistas, antropólogos o psicólogos, se han centrado en distintas unidades en las que se puede dividir la emisión de los hablantes durante la conversación: actos, enunciados, turnos, intercambios y otros; pero estas unidades no se rigen por una misma definición para todos.

Una unidad utilizada de forma generalizada para describir la conversación es el *enunciado (utterance)*, cuyos límites son los silencios y el cambio de hablante (Harris 1951; Berstein 1962; Frake 1972). La primera definición de turno como periodo de tiempo que comienza cuando una persona empieza a hablar en solitario y finaliza cuando otra persona comienza a hablar en solitario se les atribuye a J. Jaffe y S. Feldstein (Edelsky 1981, Goodwin 1981).

The speaker who utters the first unilateral sound both initiates the conversation and gains the possession of the floor. Having gained possession, a speaker maintains it until the first unilateral sound by another speaker, at which time the latter gains possession of the floor. The conversation terminates at its last sound (Jaffe y Feldstein 1970:19).

Esta definición se basa en el cambio de emisor; en cambio, los silencios del hablante no influyen como sí lo hacían en el enunciado. Jaffe y Feldstein son conscientes del problema que el habla simultánea conlleva para esta definición y deciden excluir esos casos de su estudio. Parece una decisión en línea algunos autores que menciona con Edelsky (1981: 397), para los que el habla simultánea es “degenerate” (Goffman 1967: 40), “a breackdown” (Duncan 1973: 33) o “something requiring repair” (Mchoul 1978: 199). No obstante, otros han demostrado que los solapamientos no son un fenómeno caótico, sino ordenado con precisión (Jefferson 1973; en español, Cestero 2000a).

Aunque el Grupo Val.Es.Co. (2014) mencione el cambio de emisor para describir la unidad intervención, dejan claro que es una característica prototípica. En géneros dialógicos más planificados como la entrevista la mayoría de las intervenciones son continuas, es decir, la intervención se corresponde con la emisión de habla delimitada por el cambio de hablante, pero en la conversación coloquial hay más intervenciones discontinuas (García Ramón 2020). Estas intervenciones han sufrido algún tipo de interrupción o solapamiento de otro interlocutor, pero “la continuidad y la progresión de la construcción fónico-sintáctica y temático-intencional (informativa, intencional) indican que se trata de una única intervención discontinua” (Grupo Val.Es.Co. 2014). También las marcas paralingüísticas o no verbales pueden ayudar a identificar la continuidad o el cambio de intervención. Por tanto, mientras otros analistas basan la intervención en la frontera física que es el cambio de voz, la característica principal y definitoria de la unidad en Val.Es.Co. es “el mantenimiento de una misma fuerza ilocutiva vinculada a una estrategia única” (Briz y Grupo Val.Es.Co. 2003). Esta unidad estructural se forma a través del valor pragmático de los constituyentes de la emisión, tal como se indica en Briz (2002-2004): “Los conceptos de inicio y, sobre todo, el de reacción son claves para la definición y reconocimiento de esta unidad”. Por ejemplo, si una emisión de habla (A) se solapa con la de otro hablante (B), (A) será una intervención o dos dependiendo de si reacciona o no a (B); si reacciona, son dos porque responden a distintas estrategias comunicativas; si no reacciona, es una sola intervención porque la estrategia original continúa. Padilla (2020) ve una relación entre la estructura fragmentada y el dinamismo conversacional: “La distinción entre intervenciones continuas y discontinuas obedece, de nuevo, a una dinámica de acción-reacción (...) un número mayor o menor de interrupciones afecta al desarrollo del dinamismo discursivo y a la progresión natural de la conversación” (245).

En lo que respecta a la estructura del turno, el AC describe su construcción como un proceso (Sacks et al. 1974): cuando el hablante empieza a hablar tiene planeado completar al menos una *turn-constructive unit* (en adelante, TCU), que puede ser una oración, una cláusula, una frase o una construcción léxica; sin embargo, no sabe cómo va a terminar el turno, porque depende del otro participante. En esta primera propuesta, un criterio sintáctico identifica las TCU, pero “no se arriesgan a proponer ejemplos de esta segmentación” (Gallardo 1996:79). La Lingüística Interaccional trabaja posteriormente en los elementos que segmentan las TCU (véase en Fox et al. 2013: 732) y Schegloff (2007: 3-4) añade dos criterios más para construir o identificar las TCU: la entonación que acompaña a estos enunciados y la acción que cumplen en el contexto. Es decir, en la TCU se integran recursos sintácticos, prosódicos y pragmáticos.

Cuando el final proyectado se cumple, se produce un *transition-relevance place*, o sea, un *lugar de transición pertinente* (en adelante, LTP) (Sacks et al. 1974). En ese momento el interlocutor puede, si lo desea, tomar el turno sin interrumpir el mensaje del primer hablante. Generalmente, si el hablante selecciona al siguiente hablante¹¹ (fase 1), habrá cambio de turno. Un oyente puede auto-seleccionarse (fase 2) o no. Si no lo hace (fase 3), el hablante puede añadir elementos hasta que otro tome el turno.

They avoid the problems of approaches that conceptualize the turn as a static structure by explicitly providing (...) discrete but mutable boundaries (Goodwin 1981:21).

Esta construcción del turno en tres fases aporta una característica esencial del turno: la longitud del turno se decide en la interacción. Los límites del turno no son una cuestión teórica que solo interesa al analista, sino que los interlocutores deben dominar para asegurar una gestión adecuada de los turnos (Sacks et al. 1974: 719): por un lado, el próximo hablante esperará al LTP más cercano para hablar; por otro, el hablante actual se asegurará de no hacer un LTP hasta haber expresado la información que quería dar. Para que este sistema funcione, es esencial el concepto de *proyectabilidad*: el hablante y el oyente anticipan el final de la TCU para evitar fallos en la gestión del turno.

Este mecanismo de toma de turno se denomina *local management system*, opera en cada turno, organiza solo la transición de un hablante al siguiente y es indiferente al grupo de potenciales hablantes siguientes. Algunas técnicas para producir un cambio de

¹¹ Véase Lerner (2003) sobre indicadores explícitos para seleccionar al hablante siguiente.

turno y seleccionar al siguiente hablante son las fórmulas de tratamiento (Lerner 2003), la mirada (Stivers y Rossano 2010) (§3.3.2) y el par adyacente (Schegloff y Sacks 1973).

Paralelamente a la teoría del AC, desde la semiótica, Duncan y sus colaboradores (Duncan y Fiske 1977) desarrollan más señales que activan el mecanismo de distribución de turnos en un LTP. Cada participante en cada momento de la interacción puede desear un intercambio del turno o el mantenimiento del *statu quo* hablante oyente. A este concepto lo denominan *estado de disponibilidad para la transición* (*transition-readiness state*) y es una intención que no se considera de forma discreta, sino como una escala de valores en la que el hablante puede utilizar más o menos indicios para la señal de turno y el oyente decide si son suficientes según su propia voluntad de tomarlo (Duncan y Fiske 1977: 196-198).

El hablante, que tiene la palabra, utiliza dos señales fundamentalmente: la que cede el turno y la que niega la cesión (Duncan y Fiske 1977: 183). La *señal de turno de hablante* (cesión) se activa cuando el hablante realiza uno de estos seis indicios: 1) tonema ascendente o descendente al final de una cláusula; 2) secuencia sociocéntrica (una expresión estereotipada como “you know” u “or something”); 3) conclusión gramatical de una cláusula; 4) alargamiento o arrastre de los sonidos de la última sílaba o de la sílaba tónica; 5) terminación de gestos de manos o relajación de una posición de mano tensa; 6) descenso de la entonación y/o volumen alto de la secuencia sociocéntrica.

A mayor número de indicios, mayor probabilidad de que el oyente tome el turno y de que el intercambio sea suave, es decir, sin solapamientos (Duncan y Fiske 1977: 194-195). Tras la señal de hablante, el oyente toma el turno durante un silencio breve¹² del hablante (Duncan y Fiske 1977: 187). El oyente no está obligado a reclamar el turno, puede dar un *back-channel* o permanecer en silencio; en cambio, si el hablante no da ninguna señal de cesión, pero el oyente reclama el turno, es inapropiado y puede dar lugar a turnos simultáneos (Duncan y Fiske 1977: 186). La señal que niega la cesión del hablante, *gesticulation signal* (Duncan y Fiske 1977: 188-189), inhibe cualquier señal de turno producida y reduce los reclamos de turno del oyente. Está compuesta por un solo indicio: una o ambas manos del hablante gesticulando o en una posición de mano tensa.

¹² Jaffe y Feldstein (1970) establecieron una duración media de la “switching pause” entre 0.733 y 1.555, pero no diferenciaban entre turnos y *back-channels*.

En español, Cestero (2000a) caracteriza la alternancia de turno sobre dos variables: el final de la cláusula y la superposición, basándose en el modelo de la construcción interactiva de turnos de Wilson, Wiemann y Zimmerman (1984). Este modelo mejora al de Duncan y Fiske porque, además, tiene en cuenta el contexto, que modifica las funciones de las señales (Cestero 2000a:60). Por ejemplo, si el canal es fónico, se usarán la entonación y la conclusión gramatical; si se dispone de más recursos, habrá otras señales también. Además, tiene en cuenta los factores sociales que determinan fenómenos en el intercambio de turnos. Las marcas entonativas del hablante que Cestero identifica para indicar que el mensaje del turno está completo son (Cestero 2000a: 90-112): tonema, curva melódica, movimiento tonal, formante fundamental, alargamiento final, velocidad de emisión, marca de distribución de turno (preguntas ratificadoras y otras). Aparte, incluye otros recursos para finalizar el turno (Cestero 2000a: 120-122): elementos paralingüísticos (como la risa), quinésicos y proxémicos (pero no detalla estos dos últimos); proyección de final en la última cláusula (*total que, resumiendo, en pocas palabras, pero vamos, al final*); locuciones y marcadores conclusivos (*un lío de la leche; hombre, por supuesto*), interjecciones (*jolín, macho, tía, hija*) y resúmenes breves de lo dicho.

2.3.2. La relación pragmática entre turnos

Más allá de señales concretas al final de una TCU que permiten un cambio de turno y de que cada turno depende de los turnos anteriores, el AC pone el foco en lo que las propias TCU proyectan como acción y su efecto en la secuencia. Esta relación entre los turnos de los hablantes se basa en la figura del *par adyacente*. El par está formado por enunciados producidos por diferentes hablantes de forma adyacente (no necesariamente contigua) que se ordenan como primera parte y segunda parte gracias a la *relevancia condicional* (Schegloff 1972): dada una primera parte del par adyacente (FPP), una segunda parte es inmediatamente relevante y esperable (SPP). Algunos ejemplos del tipo de relación que se establece en el par son: saludo-saludo, pregunta-respuesta, invitación-aceptación/rechazo, ofrecimiento-aceptación/rechazo, pedido-otorgamiento/rechazo, evaluación-acuerdo/desacuerdo, despedida-despedida. Aun así, Stivers y Rossano (2010a, 2010b) han cuestionado este funcionamiento y proponen un modelo escalar de relevancia condicional en el que distintas características del diseño del turno hacen más o menos relevante una respuesta. Por ejemplo, peticiones, invitaciones y ofrecimientos ejercen claramente alta presión hacia una respuesta, mientras que anuncios, observaciones

y evaluaciones, no ejercen la misma presión; pero una morfosintaxis interrogativa, una entonación interrogativa, un desbalance epistémico o la mirada del hablante hacia el receptor son rasgos que aumentan la presión hacia una segunda parte.

El AC y el AD comparten con la Teoría de los Actos de Habla el uso de categorías como petición, invitación, saludos, etc. (Levinson 1983), pero el AC las ha utilizado para describir la relación entre turnos y el AD, para denominar componentes internos del turno con esa función. Estos últimos reivindican que sus categorías no son simplemente etiquetas semánticas, sino que responden a unidades distintas estructuralmente (Coulthard y Brazil 1992: 65). En el modelo de Ginebra (Roulet et al. 2001: 191), identifican la relación ilocutiva genérica que caracteriza a una intervención mediante un cálculo inferencial fundado sobre la combinación de información de naturaleza lingüística, jerárquica y referencial; o sea, mediante la paráfrasis (una construcción sintáctica declarativa, interrogativa o imperativa o un verbo performativo como *dice* o *pide*), inferencias, la posición de la intervención dentro de la estructura del intercambio o el uso de marcadores lingüísticos.

La relación de proyectabilidad del AC también se utiliza en el AD, con la denominación de *predictibilidad* (Coulthard y Brazil 1992: 71). Por su parte, la relación se establece entre los movimientos de cada hablante, los cuales forman un intercambio, no entre turnos completos. Su intercambio lo constituye, como mínimo, un movimiento en posición de *iniciación* y otro en posición de *respuesta*. La iniciación no está predicha por un movimiento previo, pero es predictivo de uno posterior; en cambio, la respuesta está predicha por la iniciación, pero no es predictiva de ningún movimiento posterior. Estas posiciones son tomadas también para las intervenciones por el módulo jerárquico de Ginebra y por el modelo de Val.Es.Co. Para estos modelos, el intercambio está formado, como mínimo, por dos intervenciones con funciones ilocutivas distintas (Roulet 1985:25): *iniciativas*, que piden una reacción, y *reactivas*, que responden a una intervención anterior. Para el modelo Val.Es.Co., estas posiciones son las que caracterizan a la intervención como unidad: “La intervención es la unidad monológica máxima estructural, asociada al cambio de emisor, que se caracteriza por ser o por provocar una reacción, prototípicamente, lingüística” (Briz y Grupo Val.Es.Co. 2003: 11).

En el primer modelo del AD (Sinclair y Coulthard 1975¹³), el movimiento *opening* podía incluir: actos de *elicitation*, que piden una respuesta lingüística; *directive*, que piden una respuesta no lingüística, como abrir los libros; e *informative*, que dan información. El movimiento de *answering* incluía *reply* o *react*. Estos valores correspondían a actos, no a movimientos; mientras, se mantenía una correspondencia entre la posición iniciativa y la función del movimiento¹⁴. Estas mismas funciones son aplicadas por la Escuela de Ginebra: las intervenciones iniciativas son *pregunta*, *petición* e *información*, y las reactivas son *respuesta* y *ratificación* (Roulet 2001: 165-191).

En Val.Es.Co. especifican que las intervenciones iniciativas (Ii) inician el diálogo provocando una reacción posterior; bien de forma directa porque es la primera parte de un par adyacente y requieren una reacción, bien de forma indirecta con valoraciones, opiniones o evaluaciones que no exigen una reacción. En Pons (2022) se expone que los criterios para identificar la Ii son marcas lingüísticas como las preguntas y ciertas partículas (*por cierto*, *pues resulta que*, *bueno pues volviendo a*, *yo el otro día*, etc.), y el hecho de que no reacciona a una intervención previa sino al tema en general.

Compartimos con Levinson (1983) que no existe la fuerza ilocutiva de responder, sino que la respuesta lo es porque ocurre después de una pregunta. Efectivamente, el AD observó que los actos mostraban más flexibilidad de aparición en la estructura, por lo que se recategorizaron como movimientos (Francis y Hunston 1992). Tanto *elicitación* (*eliciting*) y *orden* (*directing*) como *informe* (*informing*) pueden ser movimientos iniciativos (I), pero el *informe* también puede ser respuesta (R) y el *comportamiento* (*behaving*) solo puede ser R a la orden. Véanse estos cambios en la Tabla 3.

En cualquier tipo de intercambio R es obligatorio, ya que ambos participantes deben dar evidencia de que toman parte en la conversación y se escuchan (Francis y Hunston 1992: 147). Un intercambio de elicitación tendrá necesariamente un movimiento

¹³ En la propuesta original de Sinclair y Coulthard (1975) se etiquetaron semánticamente los movimientos y los actos a partir del estudio del discurso del aula, por lo que no son apropiadas para otros tipos de discurso. Por ello, en Coulthard y Montgomery (1981) se discuten los problemas de aplicación del modelo a otros datos y se proponen ciertas alteraciones. La presentación del nuevo modelo con los cambios aplicados es publicada por Francis y Hunston (1992).

¹⁴ En el ámbito hispánico, Gallardo (1996) toma los elementos del intercambio del AD –Iniciación, Respuesta y Respuesta/Iniciación– pero adjudica estas posiciones a la unidad intervención y no al movimiento. Concretamente, establece intervenciones iniciativas (Ii), intervenciones reactivas (Ir) e intervenciones mixtas (R+I). Hay tres tipos de Ii: inicio, informe y relanzamiento (84), correspondientes a *elicitation*, *informative* y *prompt*; y tres tipos de Ir: respuesta, reacción y respuesta/inicio, de las cuales, los dos primeros corresponden a *reply* y *react*.

Sinclair y Coulthard (1975) <i>Teaching exchange</i>		Francis y Hunston (1992) <i>Conversational exchange</i>	
<i>Move</i>	<i>Head act</i>	<i>Move</i>	<i>Head act</i>
<i>Opening</i> (I)	<i>Elicitation</i> <i>Directive</i> <i>Informative</i> <i>Check</i>	<i>Eliciting</i> (I, R/I)	<i>Inquire</i> <i>Neutral proposal</i> <i>Marked proposal</i> <i>Return</i> <i>Loop</i> <i>Prompt</i>
<i>Answering</i> (R)	<i>Reply</i> <i>React</i>	<i>Informing</i> (I, R/I, R)	<i>Informative</i> <i>Observation</i> <i>Concur</i> <i>Confirm</i> <i>Qualify</i> <i>Reject</i>
<i>Follow-up</i> (F)	<i>Evaluate</i>	<i>Acknowledging</i> (R, F)	<i>Terminate</i> <i>Receive</i> <i>React</i> <i>Reformulate</i> <i>Endorse</i> <i>Protest</i>
		<i>Directing</i> (I, R/I)	<i>Directive</i>
		<i>Behaving</i> (R)	<i>Behave</i>

Tabla 3. Cambios en el intercambio Teaching/Conversational desde Sinclair y Coulthard (1975) a Francis y Hunston (1992).

I de elicitación y un R de informe; si no, estará incompleto. En un intercambio de informe habría un I de informe y un R de reconocimiento. Aun así, se producen circunstancias excepcionales en ciertos intercambios que no se consideran incompletos (Francis y Hunston 1992: 147, 155). Por ejemplo, en el discurso profesor-alumno o en el de guafurista, se establece que un participante es el hablante y los otros, los oyentes, por lo que los oyentes no tienen derecho a hablar y se asume que atienden estando presentes sin necesidad de expresar reconocimiento.

En Val.Es.Co, las intervenciones reactivas (Ir), que son reacción a una intervención previa, se pueden resumir en cuatro tipos (Briz 1998: 57-58, Grupo Val.Es.Co 2014: 19-20) (Tabla 4): la Ir cooperativa es predecible a partir de intervenciones iniciativas directas, es decir, que es la segunda parte del par adyacente (contestar a una pregunta, a un deseo, petición, o invitación, etc., responder a una valoración, etc.); la Ir evaluadora responde a las valoraciones de lo dicho por otro (manifestaciones de acuerdo o desacuerdo, de aceptación, aprobación o rechazo de lo que se está diciendo); la Ir fática confirma los papeles comunicativos con reacciones como *ya ya*, *mm mm* o que manifiestan interés por lo que dice el interlocutor (confirmaciones de

actos representativos, de ratificación del hablante, evaluaciones ilocutivas, demandas de información, de confirmación); la Ir completiva completa el acto de otro interlocutor.

Intervención iniciativa (Ii)	Directa
	Indirecta
Intervención reactiva (Ir)	Cooperativa
	Evaluadora
	Fática
	Completiva

Tabla 4. Tipos de intervenciones de Val.Es.Co.

Sin embargo, el par adyacente puede tener más de dos elementos. Opcionalmente, un tercer elemento denominado *posexpansión mínima* o *tercero terminal* (*sequence-closing thirds*) (Schegloff 2007: 118-148) revela que la secuencia fue “suficiente, adecuada, comprendida o, simplemente, no problemática” (Raymond y Olgún 2022: 92). Suelen ser partículas de cambio de estado (*ah*), acuerdo (*vale*) o apreciaciones (*genial*, *perfecto*, etc.). En el AD, esta tercera posición se denomina *seguimiento* (*following*, *F*) y el movimiento realizado es de *reconocimiento* (*acknowledging*), que, además de *F*, puede ser *R* de una *I* de informe (Francis y Hunston 1992) (Tabla 3). La *F* no es predictiva ni predicha, en consecuencia, se considera opcional (Coulthard y Brazil 1992: 70, Francis y Hunston 1992: 146). Puede no aparecer, por ejemplo, en la situación en la que un hablante elicitaba una pieza de información a partir del contenido del movimiento inmediatamente anterior de otro hablante: se entiende que lo ha escuchado y no hace falta el reconocimiento.

Hay más estructuras posibles que el par adyacente, pues, este se puede expandir con turnos adicionales antes del primer par (*pre-expansion*), entre el primero y el segundo (*inter-expansion*) y después del segundo (*post-expansion*) (Schegloff 2007) (Tabla 5). Las secuencias preliminares o expandidas dependen de la aparición o la expectativa de aparición de *respuestas despreferidas*, es decir, de respuestas que no completan la actividad que la primera parte del par adyacente proyecta (Atkinson y Heritage 1984, Schegloff 2007). Para el modelo de Ginebra, el término clave en este proceso es la *negociación*: las reacciones negativas requieren negociación y esta es la causa de que los intercambios puedan contener más de dos intervenciones (Roulet 2001: 165-191). La *pre-expansión* proyecta la posibilidad de que se formule la FPP del par base (una invitación, una historia, etc.) y anticipa o evita el recibimiento de una respuesta preferida o

despreferida (Schegloff 2007: 58). Si no se produce la SPP del par base, la ausencia es remarcable; si aparece otra primera parte en su lugar, es porque hay una secuencia de expansión (Levinson 1983: 303-307), debida a que se necesita más información para posibilitar que la SPP base sea preferida o despreferida (Schegloff 2007: 100). Tras la SPP, la secuencia se puede expandir, especialmente si la SPP es despreferida (Schegloff 2007: 117). También se puede formular un turno que cierra definitivamente la secuencia (reflexión post-terminación o *postmortem*) con una reflexión ambiguamente privada o semipública que ofrece o un análisis o una valoración de la secuencia (Schegloff 2007: 143).

Pre-pre	Petición de permiso para iniciar la pre-secuencia	
	Aceptación de permiso	
Pre-secuencia	Pre-invitaciones, pre-ofertas, pre-anuncios, pre-historias, pre-secuencia genérica (invocación)	
	Respuesta preferida (posibilita par) o despreferida (requiere más negociación para iniciar el par)	
Primera parte del par adyacente	Saludo, pregunta, ofrecimiento, etc.	
Expansión con secuencia insertada	Post-primera: petición de reparación	Pre-segunda: pregunta que posibilita responder adecuadamente a la segunda parte del par adyacente
	Repetición	Respuesta
Segunda parte del par adyacente	Respuesta preferida o despreferida al tipo de acción de la primera parte del par	
Post-expansión	Post-expansión mínima: tercero terminal para cerrar la secuencia (<i>oh, okay</i> , valoración, reflexión final)	Post-expansión no mínima: reparación, reparación por respuesta despreferida, desafío, topicalización, reelaboración

Tabla 5. Estructura de la secuencia del par adyacente (Schegloff 2007)

Por tanto, el sistema no solo tiene un ámbito local sobre el turno, sino que su radio de acción puede ser más amplio gracias a las secuencias (Schegloff 1972, 2007, Stivers et al. 2009, Roberts et al. 2015), que también son proyectables.

That is, sequences of turns are not haphazard but have a shape or structure, and can be tracked for where they came from, what is being done through them, and where they might be going (Schegloff 2007: 3).

Este funcionamiento de las secuencias conceptualiza la estructura de la conversación del AC “no solo de forma lineal y temporal, sino también de forma jerárquica” (Raymond y Olguín 2022:95). Sin embargo, no todos los turnos están sujetos a esta jerarquía del par adyacente, pueden sucederse unos a otros y estar relacionados entre sí sin esta dependencia basada en la proyección (Schegloff 2007: 15-16).

Para describir la jerarquía de los intercambios, el módulo jerárquico de Ginebra opta por la recursividad de sus unidades. Es decir, podemos encontrar la presencia de constituyentes de rango superior, igual o inferior en el interior de cualquier constituyente. Por tanto, dentro de la intervención principal puede haber intercambios e intervenciones secundarias (Roulet 1985: 34-40, Roulet 2001: 45, 90). Esto sucede en intercambios subordinados de preparación antes de la intervención iniciativa principal o de petición de clarificación (Fig.2), de manera que la unidad intervención puede tener dos interlocutores.

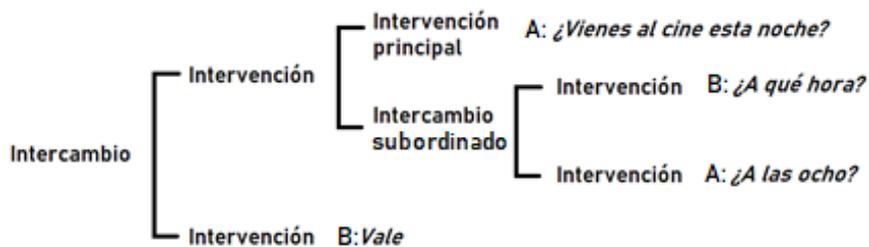


Fig. 2. Ejemplo traducido de segmentación de un intercambio (Roulet, 2001:90)

Algunas críticas recibidas responden a la complejidad de las estructuras, su carácter puramente formal y difícil de motivar, los riesgos de la subordinación y la reificación de las estructuras dialógicas (Roulet 2001: 91).

La recursividad de Val.Es.Co. no contempla intercambios dentro de intervenciones, pero sí intercambios en serie regidos o subordinados a un intercambio principal (Briz 1998: 63-64). El AD también observó que los intercambios eran más largos que la estructura básica de tres movimientos, por lo que fue necesario incorporar un nuevo elemento estructural que cumpliera una función de *Respuesta e Inicio (R/I)* al mismo tiempo (Francis y Hunston 1992: 124), el cual es tanto predictivo como predicho (Coulthard y Brazil 1992: 71, Francis y Hunston 1992: 146). Los movimientos que ocupan la posición R/I son los mismos que podían ser I (elicitación, orden e informe) (Francis y Hunston 1992: 141-143). El modelo de Ginebra y el de Val.Es.Co. toman esta

posición para la intervención reactivo-iniciativa (abreviatura *Ir-i* en Val.Es.Co.), que es reacción a una intervención previa y, a la vez, provoca una reacción posterior reactiva. Como la unidad que describen es la intervención, no el movimiento, concuerdan con el AC en que un turno puede ser al mismo tiempo segunda parte y primera parte de dos pares adyacentes distintos (Schegloff 2007: 14). Val.Es.Co. (2014: 18) destaca que, precisamente, esta *Ir-i* es el tipo de intervención más común en una conversación dinámica en la que se garantiza un cambio fluido de interlocutores.

Sirva la Tabla 6 para observar el paralelismo (no completamente exacto) entre las estructuras del intercambio de los modelos mencionados.

Turno (Schegloff 2007)	1ª parte del par adyacente	2ª parte del par adyacente	Expansión	3º terminal
Movimiento (Francis y Hunston 1992)	Iniciación	Respuesta	Respuesta/Iniciación	Seguimiento
Intervención (Roulet et al. 1985)	Intervención iniciativa	Intervención reactiva	Intervención reactivo-iniciativa	Intervención reactiva
Intervención (Briz y Grupo Val.Es.Co. 2003)	Intervención iniciativa (turno)	Intervención reactiva (turno)	Intervención reactivo- iniciativa (turno)	Intervención reactiva (no turno)

Tabla 6. Estructura del intercambio en el Análisis de la Conversación, el Análisis del Discurso, la Escuela de Ginebra y el Grupo Val.Es.Co.

Levinson hace notar la similitud de la secuencia del AC con los sintagmas de las teorías estructuralistas: “the investigation of how sequential (or *syntagmatic*) considerations restrict the class of items that may expectably follow, and of how items in that class contrast with one another (or stand in *paradigmatic* relations)” (1983: 367; la cursiva es suya). Sin embargo, recalca que no es un modelo sintáctico para la conversación, ya que las reglas describen expectativas no marcadas en vez de secuencias bien formadas. Esto mismo es aplicable también al resto de modelos estructurales de la conversación.

2.3.3. Turnos de hablante y turnos de oyente

La estructura del intercambio y los requerimientos pragmáticos de sus intervenciones se ven afectados por el rol de sus interlocutores, pues, para facilitar la gestión de los turnos, los participantes toman y se intercambian los roles de hablante y oyente. Al final, es una

actividad que se gestiona de manera social entre los participantes. Por tanto, en este apartado explicaremos en qué consisten estos roles, la caracterización de ciertos turnos como propios del rol del oyente y su función estructural.

Tanto Saussure (1916) como Shannon y Weaver (1949) y Jakobson (1950) consideran que la comunicación la establecen las figuras del hablante y el oyente o del emisor y el receptor, donde uno habla/emite, otro escucha/recibe y después invierten los roles. No obstante, es una simplificación de la comunicación humana, ya que los interlocutores no dejan de emitir y recibir información, son emisores y receptores permanentes (Briz y Grupo Val.Es.Co. 2003: 16-17, Padilla 2004).

La construcción del turno no depende únicamente del emisor; sino que requiere la participación activa del receptor para su desarrollo (§2.3.1). Mediante *recipient design*, la acción o no-acción del oyente puede conducir a modificaciones sustanciales en el enunciado que el hablante está produciendo (Sacks et al. 1974, Goodwin y Heritage 1990). El hablante monitorea durante la emisión si su mensaje es adecuado en la reacción de la otra persona, lo que le permite modificarlo si es necesario y que un silencio tras el turno no sea problemático, porque ya se ha ratificado como relevante (Goodwin 1981: 103). El modelo de la Coenunciación (Morel y Danon-Boileau 1998, Bouvet y Morel 2002, Vladimirska 2008), mencionado en §2.2.3., se ha centrado en este proceso para la segmentación formal del discurso. Con el término *coenunciación* se refieren a la representación que el hablante construye en su mente sobre las expectativas del oyente y su grado de conocimiento compartido; por consiguiente, el hablante realiza variaciones prosódicas (tono, intensidad, duración, silencio) y gestos manuales¹⁵ mientras ocurre el intercambio anticipándose a las acciones del oyente (Morel y Vladimirska 2014: 187-192).

En un experimento (Bavelas, Coates y Johnson 2000) se crearon condiciones para distraer al oyente de la narrativa del hablante, de manera que la ratio de respuesta del

¹⁵ Una subida de tono o un gesto hacia el oyente tienen una función apelativa; una caída de tono o un gesto hacia uno mismo expresan una posición personal diferenciada del coenunciador (posición egocéntrica); un mantenimiento o incremento la intensidad del habla o apartar la mirada permite al hablante mantener su derecho a hablar, mientras que una caída de la intensidad muestra la ausencia de demanda del derecho a hablar, dando la oportunidad al colocutor; la pausa corta (medio segundo) da tiempo al oyente para construir el significado que ha escuchado y puede invitarlo a expresar su grado de comprensión o una reacción consensual o diferenciada, mientras que la pausa larga se interpreta como una posición de autoridad por parte del hablante o como una dificultad con la formulación (Morel y Vladimirska 2014: 187-192). Durante un solapamiento de dos hablantes, estos no retiran la mirada (u orientación de la cabeza hacia el interlocutor) para mantener la palabra, es una marca de dominación (Bouvet y Morel 2002: 97-98).

oyente cayó significativamente (especialmente, respuestas específicas), al igual que la calidad de la narrativa del hablante, con un final de historia dubitativo en vez de dramático. Esto evidencia la importancia de la retroalimentación para validar el turno del hablante. La risa compartida tras un chiste (Glen 1989), la finalización coral y cooperativa de los turnos (Lerner 2002), así como otras acciones solapadas descritas por el AC son un logro coordinado que representa un alto grado de afiliación entre sus participantes.

Las actividades fáticas del hablante y las reguladoras del oyente son correlativas (Kerbrat-Orecchioni 1990: 19), ambas sirven para asegurar el desarrollo de la conversación mediante una relación de *cooperación* (Grice 1975), *tensión* (Criado de Val 1980: 19) o, respecto a la coordinación de movimientos corporales entre hablante y oyente, *sincronía interaccional* (Kendon 1970). Sus turnos siguen una estructura lineal debido a la obligación impuesta por la articulación, pero también es circular o retroalimentadora (Gallardo 1992).

Entonces, durante la emisión de un turno, el hablante sigue siendo receptor de estímulos y, al mismo tiempo, el oyente que recibe el turno emite información retroalimentadora que puede modificar el turno o, al menos, confirmar la recepción (Briz y Grupo Val.Es.Co. 2003: 16), ya sea verbalmente o por otros medios (mirada, gestos, sonidos no verbales). Si la emisión de palabras y el silencio no son suficientes para caracterizar el papel de cada interlocutor en un momento dado, se hace necesario distinguir qué tipo de intervenciones son las que demuestran que un interlocutor tiene la palabra y cuáles no. Algunos autores proponen una distinción entre el turno como la emisión realmente temática, de contenido referencial, “dotado de contenido proposicional, que se encarga del avance temático de la conversación” (Gallardo 1998: 34), y otro turno que tiene una función metaconversacional o metadiscursiva. El primero transmite la información que los interlocutores intercambian durante la conversación; el segundo posibilita esa comunicación mediante muestras de seguimiento activas hacia el turno principal. Cada autor utiliza denominaciones distintas, pero Edelsky (1981) renombra a estos dos tipos de turno como *turn* y *floor*, que Gallardo (1996) traduce como *turno* y *palabra*. De esta manera, se puede tener el turno desde una perspectiva estructural sin estar en posesión de la palabra.

Un tipo de emisión que se suele considerar un turno que tan solo tiene una función reguladora de la conversación es el *back-channel*, cuya identificación se le atribuye a Yngve (1970), quien lo contempla como la emisión de una persona que toma el turno

ocasionalmente para mostrar entendimiento o hacer preguntas de clarificación mientras otra persona tiene la palabra. Aun así, hay menciones previas con otras denominaciones (Duncan y Fiske 1977): “conventional signals of attention to continuous discourse” (Fries 1952), “accompaniment signals” (Kendon 1967), “listener response” (Dittmann y Llewellyn 1967, 1968). Otros autores que reproducen la idea de que estos elementos no toman la palabra y ratifican la del hablante son Duncan y Fiske (1977), Edmonson (1981: 157), Owen (1981: 99), André-Larochebouvy (1984), Kerbrat-Orecchioni (1990) o Gallardo (1996). Estos turnos *continuadores* “sólo se explican por referencia a turnos de otro hablante. Su papel es avisar al hablante de que la repartición de papeles participativos se mantiene igual, es decir, de que puede continuar hablando porque se le hace caso” (Gallardo 1998: 33). Mediante esta retroalimentación el oyente evita asumir el rol de hablante, así que no se produce una verdadera transición de turnos. Esto se evidencia en el patrón de miradas: normalmente, el hablante habla mirando tanto al compañero como a otros objetos del campo visual, pero la retroalimentación del oyente viene acompañada de una mirada larga que muestra la atención enfocada en el hablante (Jokinen *et al.* 2013).

Los *back-channels* no constituyen turno, siguiendo a Yngve (1970), es decir, que el hablante mantiene el turno durante la emisión del *back-channel* del oyente y después de este. Los *back-channels* terminan cuando dejan de emitirse, en contraste con los turnos, que se entiende que pueden continuar y no se terminan hasta que el siguiente hablante toma el turno, por lo que el solapamiento del *back-channel* con otro turno no genera un turno simultáneo, sino habla simultánea (Duncan y Fiske 1977: 178). Además, Edelsky (1981) señala que durante turnos simultáneos no es necesario que solo uno de los dos hablantes mantenga la palabra, sino que dos hablantes mantengan el turno paralelamente o los dos tomen parte en la construcción del turno de forma colaborativa, sin que esto se tenga que ver como una degeneración de la estructura que requiera reparación.

En el modelo de Ginebra se incluyen las respuestas fáticas del oyente como intervenciones de un intercambio subordinado dentro de la intervención del hablante. Véase el ejemplo de la Fig. 3 en el que dos respuestas se integran en la jerarquía de la intervención, siendo la emisión lineal del hablante: “quería pedirle una cosa necesito un billete de avión Génova-Barcelona ida y vuelta pero todavía no sé si parto a finales de noviembre o a principios de enero existe la posibilidad de dejarlo pagado/ y telefonar después/ para decirle la fecha una semana antes/ o no” (*trad.* Roulet, 2001:78).

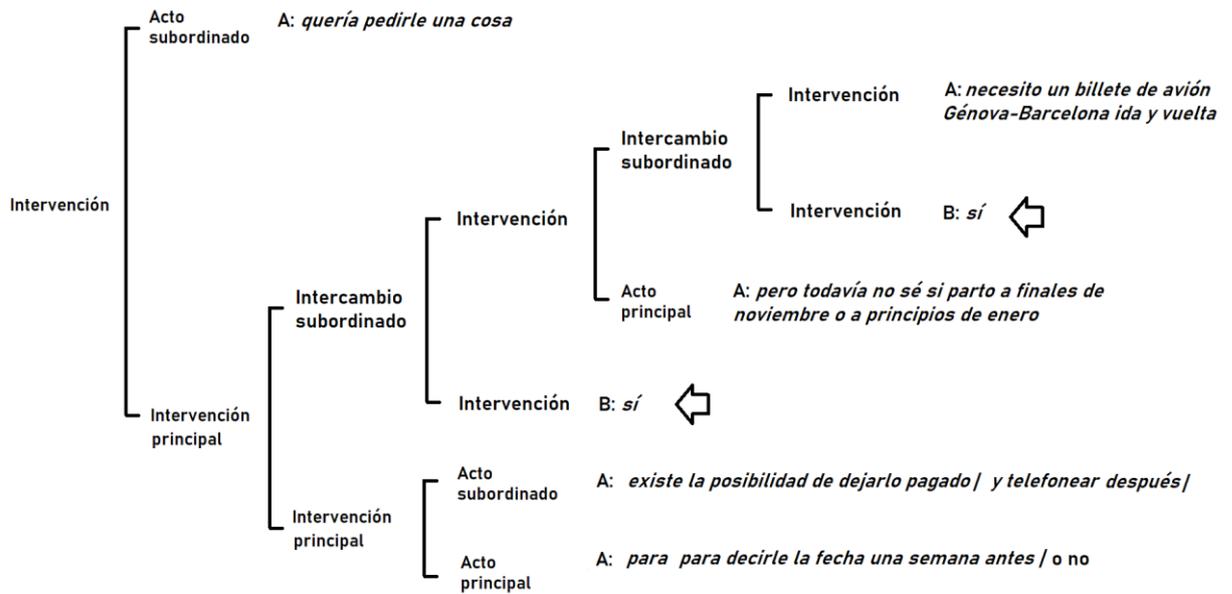


Fig. 3. Ejemplo traducido de segmentación de un intercambio (Roulet, 2001:78)

Duncan y Fiske (1977: 201-202) realizan una descripción detallada de los turnos que pueden considerarse *back-channels* del oyente:

- M-hm: un *back-channel* breve que indica atención; por ejemplo: *m-hm, yeah, right, yes quite, surely, I see, that's true.*
- Fin de enunciado: el oyente completa el enunciado que ha empezado el hablante, pero no continúa más allá; el hablante continúa como si no lo hubieran interrumpido.
- Peticiones de clarificación; suelen realizarse en pocas palabras o una frase.
- Resumen breve: en vez de completar, replantea en pocas palabras lo inmediatamente precedente expresado por el hablante.
- Asentimientos y sacudidas de cabeza; pueden acompañar o no a *back-channels* verbalizados.

La clasificación de otros autores (Oreström 1983, Stubbs 1987, Kerbrat-Orecchioni 1990, Gallardo 1996) no varía demasiado de esta¹⁶. Incluimos a continuación otras funciones añadidas por Cestero (2000b¹⁷), quien realiza un estudio exhaustivo de estos elementos, los cuales nombra como *turnos de apoyo*:

¹⁶ Gallardo (1996: 61-62), por ejemplo, resume estas cinco conductas del oyente en dos: conducta de atención continuada (continuadores, aportaciones) y conducta de colaboración (completar oraciones del hablante, turno colaborativo).

¹⁷ Además de la tipología funcional, Cestero (2000b) presenta otra formal:

- El *back-channel* “m-hm” de Duncan y Fiske lo divide Cestero en tres tipos¹⁸: *apoyos de seguimiento*, que indican que una emisión se ha recibido sin aportar significados añadidos; *apoyos de acuerdo*¹⁹, que respaldan de modo explícito las emisiones anteriores; y *apoyos de entendimiento*, que expresan que la emisión anterior se ha comprendido.
- *Apoyos de conocimiento*, que muestran entendimiento sin que el hablante haya concluido el turno porque se intuye a partir de conocimientos generales o del contexto inmediato.
- *Apoyos de reafirmación*, a través de preguntas de reafirmación (*¿sí?*).
- *Apoyos combinados*, en los que una misma forma combina dos funciones.

Incluimos, además, algunos *back-channels* que no son comunes en todas las tipologías:

- Ciertas expresiones de valor netamente evaluativo o comentativo²⁰ (Oreström 1983: 104-134): *increíble, qué horror, ¡0h! ¡Dios!*
- Turnos de paso: “no se producen en solapamiento; se utilizan para rechazar la posesión de la palabra, especialmente en secuencias de cierre” (Gallardo 1996: 84). Algunos ejemplos son: *y nada, y eso, pues nada, vale*.
- *Anti-régulateurs*: marcan que el participante ya no está presente en la interacción (Kerbrat-Orecchioni 1990).

Parece claro que todos estos elementos cumplen la función principal del continuador: indicar que se está atendiendo de una forma u otra al hablante, participando activamente en la conversación y facilitando la coordinación general de las acciones de ambos participantes en la estructura de la conversación (Duncan y Fiske 1977: 203). En cuanto a la tipología funcional del continuador, Kerbrat-Orecchioni (1990: 188-189) plantea que es problemático distinguir si son turno o son cierto tipo de continuador (que

-
- Simples: Sí/no; elementos cuasi-léxicos; otros adverbios de afirmación; interjecciones; repeticiones de palabras claves del turno de habla; palabras pertinentes respecto al contenido del turno de habla (completivas); llamadas de atención.
 - Complejos: repetición de apoyo simple; repetición de sintagma clave del turno de habla; repetición de oración o parte de oración clave del turno de habla; sintagmas pertinentes (completivas); oraciones o parte de oraciones pertinentes (completivas); locuciones adverbiales con valor enfático y expresiones interjectivas; atribuciones veritativas.
 - Compuestos: combinación de apoyo simple y complejo, combinación de dos tipos de apoyos simples, combinación de dos tipos de apoyos complejos.

¹⁸ Estos tres coinciden con los valores que Stubbs (1987:186-187) atribuye al *back-channel*, pero con otra denominación: reconocimiento, apoyo y aceptación.

¹⁹ Según el estudio de Cestero (2000b), los de acuerdo son los más comunes, un 45% del total.

²⁰ Cestero parte de la tipología de Oreström, pero decide suprimir este tipo de *back-channel*.

denomina *reguladores*) una reformulación, una evaluación o un comentario de la intervención precedente. En estos casos, dice, criterios como el solapamiento, la intensidad vocal débil y la brevedad no son satisfactorios. En cuanto a los dos primeros criterios, es difícil establecer el límite de intensidad²¹ y hay turnos que se hacen en solapamiento y reguladores que aparecen entre pausas²². En cuanto a la longitud de la emisión, por un lado, los reguladores constituyen el 60% de las emisiones breves, pero no todas lo son: las respuestas de verdadero o falso suelen ser cortas y son turno, en cambio, las tomas de turno fallidas pueden ser cortas y no son turno; por otro lado, hay repeticiones y compleciones que son largas y parecen turnos de palabra, pero son reguladores.

Siguiendo un criterio funcional siempre va a haber un problema de ambivalencia intrínseca, ya que es difícil diferenciar la función de seguimiento con otra función que exprese la alineación del emisor²³ (Kerbrat-Orecchioni 1990). Por ejemplo, un *sí* o un *de acuerdo* puede significar *te escucho, comprendo tu punto de vista o lo comparto*²⁴. Para este problema, Kerbrat-Orecchioni propone la asunción de que el carácter de la emisión como turno o como regulador es fluido. Esto es así porque todas las emisiones tienen una función reguladora al contribuir en el mantenimiento de la conversación, pero algunas incluyen otras funciones pragmáticas y otras carecen de ellas y son reguladores “puros”. No hay una delimitación estricta entre los tipos de turno, sino una gradación entre la escucha en voz alta y la producción claramente informativa de la intención del destinatario. Además, a menudo una contribución que se presenta como un regulador se metamorfosea progresivamente en turno y, al contrario, una toma de turno interrumpida se “degrada” a una emisión reguladora. Respecto al proceso de un regulador convertido en turno, Francis y Hunston (1992: 148) también mencionan que los elementos en posición de seguimiento, en forma de comentario u observación, se puede convertir en el

²¹ En el modelo de la Coenunciación (Bouvet y Morel 2002: 128-129), las emisiones del oyente se relacionan con una entonación mimética a la del hablante o con una melodía creciente, que apuntala el discurso del hablante y expresa interés; en cambio, una melodía descendente perturba la continuación del discurso del hablante porque anticipa la intención de recuperar el estatus de hablante.

²² Reflexión que también comparte Edelsky (1981), quien valora sobre todo la intencionalidad al definir la interrupción. De hecho, los modelos de Sacks et al. (1974) y de Duncan & Fiske (1977) son limitados porque se enfocan en el mismo tipo de intercambio, el suave, en el que hay poco o nada de solapamiento.

²³ Gallardo (1996: 90) también destaca la adscripción fronteriza entre la reacción evaluativa y el *back-channel*.

²⁴ Edelsky (1981) ejemplifica este problema con *that's right*: puede ser un turno si se ha dicho con un volumen y patrón entonativo que se escucha como *estoy de acuerdo contigo completamente*; pero se puede categorizar como *encourager* (animador, motivador), un tipo de *back-channel*, si se entiende como una indicación para que el hablante continúe.

movimiento iniciativo de un nuevo intercambio: “Whatever act or move an utterance appears to realize, it is always subject to reclassification in the light of subsequent contributions by other participants” (Francis y Hunston 1992: 156). En cuanto al proceso contrario, Edelsky (1981) defiende el criterio de la finalización del mensaje para considerarlo turno. Es decir, cuando un interlocutor intenta realizar un turno, pero falla y no consigue decir un mensaje referencial, se considera este un *aborted utterance*, no un *aborted turn*, porque no se ha revelado el mensaje. Por tanto, los turnos que se emiten no dependen de uno mismo, sino de si otros lo ratifican y de lo que sucede en negociación de la interacción.

En cuanto la relación entre el rol de oyente y el *back-channel*, Ducan y Fiske (1977) señalan que también el hablante puede emitir *back-channel signals*, aunque con menos frecuencia. En estos casos, los realizados por el hablante bien se producen al inicio de un turno de habla, reconociendo el turno previo del compañero, bien reconocen un *back-channel* largo del oyente. Desde su perspectiva, este *back-channel* del hablante a otro del oyente no se considera una renuncia al turno, así que no se altera el estado de los participantes. En cambio, para Yngve (1970, *apud.* Edelsky 1981), si un *back-channel* recibe la reacción de otro *back-channel*, hay que considerarlo turno.

Tanto en el AC como en el AD, las reacciones son predecibles en los pares adyacentes, ya estén formados por turnos o por movimientos en cada modelo; en cambio, los comentarios que no exigen explícitamente una respuesta, como los movimientos de *informe* del AD, pueden recibir *reconocimiento* o no, creando un intercambio en el caso de que la reciban. Como también explica Kerbrat-Orecchioni (1990: 201-202), hay ciertos actos en los turnos que exigen la reacción del interlocutor más que otros: los actos directivos o las apelaciones directas; sin embargo, apunta que toda aserción solicita reacción implícitamente; especialmente, un ataque, una provocación o un anuncio. Parece que Cestero (2000b) coincide con este planteamiento, ya que remarca que los turnos de apoyo también responden a fuertes requerimientos pragmáticos del enunciado, o sea, de su fuerza ilocutiva, formando pares adyacentes o secuencias del tipo juicio-apoyo²⁵, explicación-entendimiento, narración-conocimiento o narración-petición de más información (Cestero 2000b: 49-50). En las secuencias de narración las emisiones del

²⁵ Este par se puede vincular con el par evaluación/evaluación de Pomerantz (1984).

oyente deben mostrar afiliación y una postura hacia la historia, así como promover la progresión (Stivers 2008).

Además, Cestero destaca la mayor probabilidad de apoyos ante la aparición de marcas lingüísticas o paralingüísticas en el turno de habla (Cestero 2000b: 51-53): tono elevado; locuciones enfatizadoras (*que conste*) o partículas exclamativas como (*¡eh!*); llamadas de atención (*fíjate tú, mira*); marcadores de explicación (*quiero decirte, o sea, digamos que, sabes cómo*); marcadores de duda o incertidumbre (*a lo mejor, quizás, no sé*); verbos epistémicos (*creo que, me parece que, es de lógica*); titubeos, pausas reflexivas, reflexiones que obligan a emitir turnos de apoyo. También en el modelo de la Coenunciación (Bouvet y Morel 2002: 126-127) se observan indicadores previos auditivos y visuales del hablante que autorizan al oyente a manifestarse: subida de intensidad de últimas sílabas (necesaria, pero no suficiente) y orientación de la cabeza y la mirada hacia el oyente (obligatoria). Asimismo, tienden a aparecer tras un punto en el que se completa la estructura sintáctica, prosódica y pragmática, o sea, en una TCU.

Más allá de los continuadores/reguladores/apoyos, hay otros tipos de intervenciones que se han considerado ajenos al objetivo principal de la conversación y, por tanto, no identificables como turnos de palabra de un hablante ratificado. Estas pueden ser las interacciones rituales y las secuencias de apertura, como medios para iniciar la propia conversación (Goffman 1967, 1971, Schegloff y Sacks 1973, Jefferson 1972). También las secuencias paralelas que se alejan del tema o del objetivo principal de la interacción. En este sentido Edelsky (1981) ejemplifica que no toman la palabra los alumnos que hablan entre sí mientras la maestra habla o los participantes que comentan entre sí en una reunión de trabajo. Desde el modelo de Val.Es.Co. estos apartes son diálogos laterales (Briz 2006), paralelos o solapados, o incluso discursos distintos si esa interacción supone un cambio de registro o de las dinámicas entre los hablantes que participan en ella (Estellés y Pons 2014).

Aunque Gallardo (1996: 75) hace notar que no se contempla una distinción entre turnos que toman o no la palabra en el sistema de reglas de Sacks, Schegloff y Jefferson, siempre se habla de la distribución de la posesión de la palabra. En general, para el AC todas las emisiones forman parte de un turno, aunque indican que hay ciertos turnos breves que no pretenden su propia continuación y suelen darse de forma solapada (Sacks *et al.* 1974: 707, Schegloff 1982).

2.3.4. Ratificación del turno

Si bien las dicotomías hablante oyente y emisor-receptor se utilizan indistintamente en este ámbito, Val.Es.Co. usa estos términos para diferenciar el rol social de los interlocutores en cada momento (Briz 2001, Briz *et al.* 2003, Pons Bordería 2022). Los roles de hablante y oyente suponen que el primero es el que posee el turno de palabra, esté emitiendo una intervención en ese momento o no, y el segundo no lo posee, aunque emita alguna muestra de apoyo o entendimiento. En cambio, el emisor es el participante que emite algún tipo de intervención, sea cual sea su rol social, y el receptor es el que lo escucha, sea hablante u oyente en ese momento (Briz y Grupo Val.Es.Co 2003: 17). Esta distinción social y estructural de los interlocutores le sirve al Grupo Val.Es.Co. para analizar los recursos que estos utilizan para mantener o cambiar sus roles.

En esta negociación coloquial existe con frecuencia un desequilibrio en cuanto a las participaciones de los interlocutores, en cuanto a los turnos. Las hay dominadas por uno o algunos interlocutores, que ejercen la autoridad y llevan el peso de la misma, mediante una serie de estrategias (Briz 1998: 52).

La clave en la distinción social del hablante reside en la ratificación de su rol por el oyente. Esta idea no se aleja de los planteamientos etnometodológicos en los que la identidad asumida por un participante es ratificada no por sus propias acciones, sino por las acciones de otro que asume una identidad complementaria hacia él (Goodwin y Heritage 1990: 292). Los participantes se encuentran en lo que Goffman (1971) llama *estado de conversación* cuando se ratifican recíprocamente²⁶ a través de gestos significantes. Clasifica a los oyentes según su grado de ratificación (Goffman 1976, 1981): hay dos tipos de oyentes ratificados, por un lado, el que es en ese momento el destinatario del turno y debe demostrar que está escuchando las palabras que están particularmente dirigidas a él; por otro lado, el oyente ratificado que no es el destinatario de ese turno, pero que puede influir en la voluntad comunicativa del hablante porque también forma parte de la conversación; luego, los *by-standers* actúan como receptor sin que nada los ratifique como tal ni influyan en el hablante, bien porque escuchan por casualidad o a escondidas.

El rol de hablante y oyente es cambiante, así que la prueba de que el hablante efectivamente lo es se basa en la ratificación de su turno por el oyente: “the number of

²⁶ El primer lugar en el que se ratifica a los interlocutores en la situación dialógica son las secuencias de apertura (Gallardo 1992).

times a person speaks will not necessarily be equivalent to the number of times his speaking is ratified or legitimated by others” (Philips 1976: 87). Philips (1976) afirma que la respuesta –definida como un enunciado que solo se entiende mediante el enunciado de un hablante anterior– es crucial en la validación de un turno de habla y en la incorporación de este turno en la secuencia de interacción²⁷. Es decir, que la presencia o ausencia de respuesta indica si el hablante ha sido escuchado o no. Es el oyente de este turno el que ratifica al hablante previo, frente al oyente ratificado por el hablante de Goffman (Padilla 2004).

En otras propuestas de unidades conversacionales jerarquizadas esta distinción social no se contempla. En ellas la intervención o el turno son lo mismo y viene determinado por el cambio de hablante. En cambio, Val.Es.Co. sigue la senda de los estudios vistos en §2.3.3., que sí clasificaban los turnos según su valor proposicional o metaconversacional. Mediante la dimensión social, el sistema de unidades de Val.Es.Co. aporta una nueva perspectiva a la estructura de la conversación y señala cuál es el papel de los interlocutores en cada momento. Las intervenciones que caracterizan al hablante son los turnos, cuya característica principal es que son reconocidos o aceptados por otro interlocutor o interlocutores. “La consecuencia de esta visión comunicativa es que el oyente es el verdadero artífice de los turnos, quien siempre poseerá una función activa en la constitución del turno” (Grupo Val.Es.Co. 2014).

Por tanto, aunque en otros estudios intervención y turno sean sinónimos, en Val.Es.Co. describen categorías diferenciadas. Todas las emisiones delimitadas por el cambio de hablante son intervenciones estructuralmente, pero solo las que reciben validación de algún tipo son turnos socialmente: “todo turno es, al mismo tiempo, una intervención, pero no toda intervención puede constituirse en turno” (Briz y Grupo Val.Es.Co. 2003: 10). Al contrario de las unidades estructurales, que son jerárquicas, el orden social “progresa linealmente, ya que está gobernado por la alternancia de los papeles comunicativos” (Briz y Grupo Val.Es.Co. 2003: 7).

La ratificación depende de la existencia de relación con los turnos posteriores de otros hablantes: “Esta reacción se asocia, por una parte, al contenido del turno previo; y,

²⁷ En este estudio, Philips concluye que este tipo de ratificación funciona en la cultura anglosajona, pero no en la interacción de los hablantes nativo-americanos, ya que la réplica a un enunciado puede aparecer minutos después o, incluso, otro día, sin que esto implique socialmente que el hablante no haya sido escuchado, reconocido e incorporado en la secuencia.

por otra, a la persona que lo posee” (Padilla 2020: 250). Cualquier reacción lingüística es una marca inequívoca de que la intervención anterior es turno (Briz y Grupo Val.Es.Co. 2003: 18-19):

- la segunda parte de un par adyacente;
- las marcas de acuerdo y desacuerdo en primera posición, que pueden ser marcadores discursivos, conjunciones y adverbios (*bien, bueno, vale, pues, pero, sí, no, etc.*);
- las reformulaciones, matizaciones o adiciones de información a ideas o elementos aparecidos en la intervención previa;
- las continuaciones sintácticas de segmentos inacabados en las intervenciones previas;
- las referencias pronominales y las elipsis que hacen alusión al contenido expresado en una intervención previa;
- la repetición léxica, que indica coincidencia en el vocabulario y en la temática, son marcas de continuidad léxica.

En este listado se deberían incluir los continuadores vistos en §2.3.3, que desempeñan principalmente una función metainteractiva de ratificar el turno actual o previo, pero Briz y Grupo Val.Es.Co. (2003) explican estas reacciones fáticas en la clasificación de intervenciones reactivas; téngase en cuenta además que también pueden aparecer como otra marca más de validación que forma parte de una intervención mayor.

Lo que hacen los apoyos es ratificar o legitimar puntos concretos, contribuir a asegurar que los comentarios no pasan desapercibidos. Decir *¡ESO es!* es definir de modo explícito que una contribución es relevante y que merece la pena hablar de ella” (Stubbs 1987:186).

En el AC, el *procedimiento de prueba de turno siguiente* (ing., *next-turn proof procedure*) implica una aceptación inferencial del turno previo a través de la acción que produce en el turno siguiente, sin necesidad de otra confirmación explícita (Heritage y Atkinson 1984:10, *apud.* Raymond y Olgún 2022: 94). El modelo de Ginebra (Roulet 1985: 203) también contempla la legitimación dialogal de los elementos del discurso, la cual es causa de una colaboración asumida que impone derechos y obligaciones a los participantes en toda interacción: un elemento del discurso se legitima si a su enunciación le sigue (de forma prevista o no) un constituyente interpretable como manifestante de una aprobación discursiva, por lo tanto, el elemento es pertinente. El hablante puede solicitar la legitimidad mediante una demanda de ratificación explícita (estrategia interaccional) o

mediante una presuposición (estrategia interactiva). Cuando la vinculación semántica o pragmática de los actos principales no es muy clara, los movimientos de enlace²⁸ con función metadiscursiva la confirman con el turno anterior (enlace retroactivo) o posterior (enlace proyectivo) (Gallardo 1996: 80).

Por otro lado, Briz y Grupo Val.Es.Co. (2003) contemplan una validación que se puede dar de forma previa a su emisión. Este reconocimiento del turno posterior puede venir de actos lingüísticos que ilocutivamente predicen una reacción, es decir, de la primera parte de un par adyacente, pero también de la apelación directa mediante marcadores interactivos como *¿no?* Estas marcas crean exigencia por algún tipo de reacción, mientras que otras no lingüísticas sirven para ceder el turno de una manera opcional. Son los LTP del Análisis de la Conversación. Algunas de estas marcas son: pausas, silencios, tonemas descendentes, estructuras sintácticas acabadas, tonemas suspendidos, estructuras truncadas, alargamientos vocálicos. Si se tomara como turno cada intervención reactiva pospuesta a un LTP, encontraríamos que prácticamente todas las intervenciones de una conversación son turno. Porque pocas veces un emisor va a reaccionar a otro hablante sin una señal mínima que se lo permita; aún menos si abrimos las marcas que elicitan reacción a las modalidades visuales. La conjunción de los criterios de validación previos y posteriores dejaría fuera de los turnos solamente a intervenciones completamente ignoradas o mal introducidas por los emisores, que no encajan en el engranaje de la conversación. De esta manera, la vinculación entre turno-hablante e intervención-oyente, se perdería.

En definitiva, la validación posterior es el criterio principal, pues el hecho de generar una reacción deja claro que esa intervención contribuye al progreso de la conversación; pero la validación previa únicamente ratifica turno en el caso de que sea motivada por un acto iniciativo directo. Este tiene mayor efecto perlocutivo que uno indirecto (Padilla 2020: 277). De esta manera, un turno se valida previamente si es una respuesta exigida por la primera parte de un par adyacente que sea una pregunta o una orden, pues ese hablante ha considerado que la respuesta es necesaria en la conversación. Generalmente, otros actos y otras marcas de LTP no son necesariamente exigentes²⁹, sino

²⁸ No se debe confundir el *movimiento* de Gallardo con el *movimiento* del AD; de hecho, la equivalencia con el AD se daría con ciertos actos subordinados.

²⁹ Como se ha mencionado en §2.3.3., Cestero (2000b), en cambio, considera que los turnos de apoyo pueden ser tan exigidos como las segundas partes de los pares adyacentes.

que favorecen lugares para que otro emisor contribuya a la conversación tomando el turno o no.

Podemos afirmar, entonces, que las *Ii* siempre son turno porque reciben reacción, pero las intervenciones directas también producen que la reacción sea turno, mientras que las indirectas no. En cuanto a las *Ir*, las cooperativas son turno (segundas partes de par adyacente), las fáticas no lo son, pero las evaluadoras y las completivas³⁰ pueden no serlo “en ocasiones” (Grupo Val.Es.Co. 2014).

Las *Ir* fáticas —del tipo *sí, ah, uhm* con función retrocanalizadora o de confirmación de contacto (Briz 1998: 59)— *a priori*, no son turno, incluso si aparecen en un LTP. Al igual que otros *marcadores metadiscursivos* que menciona Briz (1998: 201-203), en forma de preguntas fáticas u organizadores del discurso (*por cierto, bueno, o sea, entonces*), las respuestas fáticas también se vinculan a la organización de la actividad discursiva, en concreto, al control mutuo que ejercen los participantes entre sí (Briz 1998: 203). Estos elementos que al inicio de una intervención son SAI en posición independiente siguen teniendo un uso procedimental que no les permite ser turno. No necesariamente las preguntas fáticas del hablante (*¿eh?, ¿no?, ¿sabes?*) exigen una reacción (Briz 1998: 224-226): pueden cumplir una función expresivo-apelativa que refuerza o justifica razonamientos y que llama la atención para mantener o comprobar el contacto, de manera que implican al oyente solo de forma pasiva; esto no impide que en otros contextos puedan ser fórmulas exhortativas y apelativas (con un tonema ascendente más marcado) que ceden el turno e implican activamente al interlocutor.

Siguiendo este razonamiento, tan solo consideramos turno a la intervención fática si recibe alguna reacción explícita del hablante que perturbe la continuación natural del turno. Aunque la reacción fática sea un elemento necesario para confirmar al hablante que está siendo escuchado, sin la cual podría decidir abandonar su turno (Bavelas, Coates y Johnson 2000), para el sistema de Val.Es.Co., el hecho de que el hablante continúe con su turno tras la *Ir* fática no es una reacción a ese contenido. En este sentido, los turnos también marcan la diferencia entre el contenido proposicional y el procedimental.

³⁰ Las intervenciones completivas también se han denominado colaborativas en Briz (1998). Tomamos el término más nuevo (completivas) que se contrapone a las colaborativas como segunda parte de un par adyacente.

Respecto al criterio para distinguir cuándo las intervenciones evaluadoras y completivas son turno, no se explicita en Grupo Val.Es.Co. (2014), pero es la validación mediante una reacción posterior. Briz (2000) lo explica así:

[las completivas son] (...) intervenciones solapadas que añaden o completan la información del otro, corrigen, ayudan, respaldan, valoran, confirman lo dicho y la argumentación del que está hablando, manifiestan opiniones, en suma, intervenciones que colaboran con lo dicho por otro, sin que lleguen a ser atendidas, ni contestadas por el hablante, por tanto, no turnos.

En cambio, si alguna de estas aportaciones recibe una reacción explícita del hablante a esa valoración o al término propuesto para continuar el turno, la intervención evaluadora o completiva es un turno. Estas ayudas del oyente ratificado “favorecen al dinamismo y tensión comunicativos” (Briz 1998: 61-62).

Respecto a las Ii que inician diálogo —es decir, que no son Ir-i—, en la descripción de Grupo Val.Es.Co (2014) no se explica qué ocurre si no reciben reacción. Para esta situación, Espinosa Guerri (2016: 15) propone la intervención independiente (Iind): “son aquellas que no reaccionan a ninguna intervención previa ni generan reacciones posteriores”, por tanto, no forman parte de ningún intercambio. Esta categoría se ha utilizado en la segmentación del Corpus Val.Es.Co 3.0 (2022). La causa de estas intervenciones puede ser interna³¹, porque “el hablante manifiesta verbalmente una sensación, un estado de ánimo o un sentimiento; es decir, reacciona a un estímulo interno, lo exterioriza pero no genera la reacción posterior de sus interlocutores” (Espinosa Guerri 2016: 15), o externa, porque “algo del espacio donde se da la comunicación llama su atención [del emisor]” (Espinosa Guerri 2016: 16) y abandona la conversación para atender a ese estímulo externo. A estos dos tipos, se le suma otra causa contemplada en la segmentación del Corpus 3.0: los intentos de iniciar diálogo fallidos, en los que “un emisor propone un diálogo nuevo, pero nadie reacciona a él o incluso lo interrumpen” (Pons Bordería 2022). Sea cual sea la intencionalidad de los emisores en desarrollar un diálogo a partir de esa intervención, el hecho de que ningún interlocutor lo ratifique convierte a la intervención en independiente. De la misma forma, una intervención reactiva con un acto iniciativo directo que no es respondido por nadie queda finalmente como Ir, ya que son los oyentes los que reconocen y deciden que se instaure como turno.

³¹ Padilla García (2020: 248) denomina *reacciones aisladas* o *antifáticas* a las emisiones aisladas que verbalizan un pensamiento.

Como Edelsky (1981) y Gallardo (1996: 61-62) defienden, los intentos fallidos de reclamar el turno no son turno.

En definitiva, el tipo de intervención puede dar una pista sobre su ratificación potencial, pero, de hecho, cualquiera de ellas puede ser turno o no, dependiendo de la decisión de los interlocutores en el discurso particular.

2.4. Recapitulación: la multimodalidad en un modelo estructural de la conversación

A través de este capítulo se ha visto que el turno es una unidad que estructura la conversación, a partir de este los hablantes construyen la conversación de manera dialógica, intercambiando sus papeles de emisor y receptor, de forma que ambos tienen que negociar la extensión de sus turnos, los intercambios de turno y el desarrollo general del diálogo. Por tanto, el turno es una unidad estructural que difiere de los enunciados de los discursos monológicos en el intercambio de emisores y la sucesión no tan exacta de una unidad tras otra, pues es lo requerido por una construcción que no depende de la agencia de un solo hablante. Pero el turno también es una unidad social, en el sentido de que está formado por componentes que guían el intercambio de manera negociada entre los interlocutores y caracteriza los papeles sociales que cada interlocutor toma en el desarrollo de la conversación, como hablante o como oyente.

El objeto de este trabajo es averiguar la manera en la que la información multimodal podría afectar a esta estructura de la conversación en el nivel dialógico y a su dimensión social. Es decir, en qué medida la grabación en vídeo de una conversación coloquial aporta nuevas intervenciones que no se perciben en la grabación en audio, modifica la interpretación de los turnos y de las relaciones que establece con otros turnos, o muestra herramientas no orales de validación social de turnos que en la grabación de audio se omiten y cambian la interpretación de la secuencia.

El AC sí ha desarrollado estudios multimodales que contemplan la postura, la mirada o los gestos en la toma de turno y el desarrollo secuencial (§3.2.3); en cambio, los modelos de segmentación estructural, que tienen la ventaja de utilizar una jerarquía de unidades con la que describir de manera sistemática el lugar de los fenómenos que se desarrollan en la conversación, no han contemplado otras modalidades visuales y los efectos que tienen en esta jerarquía. La única excepción es el modelo de la Coenunciación, que ha integrado en su sistema las miradas y los gestos manuales, que contienen funciones

demarcativas y enunciativas complementarias a la prosodia (Bouvet y Morel 2002, Morel 2010a, 2010b, Morel y Vladimirska 2014). No obstante, este modelo se centra en unidades internas de la intervención y no clasifica las intervenciones respecto a su lugar en la interacción o el diálogo, como sí hacen los demás modelos desarrollados en este capítulo.

En §3 revisamos las aportaciones multimodales del AC, así como de otros enfoques lingüísticos y pragmáticos, pero, como el modelo de la Coenunciación es el único estructural que incorpora la multimodalidad, aunque sea en el nivel inferior al turno, a continuación, explicamos brevemente este modelo para ver cómo la multimodalidad ha contribuido a su análisis.

Respecto a sus unidades (Morel y Vladimirska 2014: 194-195), la unidad superior de análisis es el *párrafo oral*, identificado por la caída de tono e intensidad a final de sílaba. Sus unidades constituyentes son el *preámbulo* y el *rema*, cuya cohesión viene del movimiento a final de sílaba de cada grupo constitutivo. El rema transmite el contenido, mientras el preámbulo clarifica la conexión con el contexto secuencial (*ligateur*), el punto de vista modal, los valores epistémicos o apreciativos (*modus*), o el marco referencial. A veces, también hay un posrema de modalidad epistémica o un argumento nominal correferente con un pronombre del rema, que puede cerrar un episodio o discusión (Bouvet y Morel 2002:18). El párrafo oral se hace complejo cuando el tono se eleva al final del rema y la mirada se mantiene, entonces el rema se recategoriza como preámbulo de lo que sigue (Morel y Vladimirska 2014: 200).

La incorporación de las denominadas *marcas posturo-mimo-gestuales* al modelo de segmentación ha permitido la comprensión de cómo índices de distintos planos (palabras, entonación, gestos) se relacionan en la cohesión de unidades (Bouvet y Morel 2002: 111). En general, los gestos anticipan a los indicadores de otros planos (Bouvet y Morel 2002: 130). Al observar la sincronía entre mirada y constituyentes, se aprecia una adecuación perfecta: el hablante no mira al oyente durante el preámbulo (o al inicio del rema si no hay preámbulo), pero sí durante el rema o en su final, muchas veces estos movimientos anticipan la segmentación a la enunciación (Bouvet y Morel 2002: 19, Morel y Vladimirska 2014: 197-198). La cabeza anticipa el movimiento de la mirada (Bouvet y Morel 2002: 19). La delimitación de ambas, mirada y cabeza, ha generado una unidad más grande que agrupa dos o más párrafos: el *hiperpárrafo* (Bouvet y Morel 2002: 112). En cuanto a la vinculación entre los gestos manuales y los constituyentes (Morel y

Vladimirska 2014: 199) (§3.4.3.2), los gestos déicticos suelen aparecer en el preámbulo para apuntar hacia la fuente de habla o hacia el lugar en el que se localiza el referente en el espacio compartido con el oyente o fuera; los gestos ilustrativos también pueden aparecer en el preámbulo para introducir el tema, pero son más comunes en el rema representando una propiedad del referente; así como los gestos de ofrecimiento o de rechazo y gestos metalingüísticos (contar con los dedos, señalar y golpear reiteradamente con los dedos en el mismo lugar, etc.). Respecto a las funciones interactivas que el modelo de la Coenunciación contempla en los elementos de la intervención y que pueden ser de interés para el nivel dialógico, explican que las reacciones sonoras y gestuales del oyente indican que ha percibido las intenciones enunciativas del hablante y que son indispensables para mantener o cambiar su posición social (Bouvet y Morel 2002: 130).

Por tanto, en el modelo de la Coenunciación, la multimodalidad ha permitido la inclusión de nuevas marcas que guían la segmentación de unidades internas de la intervención, ha contribuido a la identificación de unidades por su vinculación de a cierto tipo de gestos e, incluso, ha favorecido la creación de unidades que no se podían intuir con la información auditiva. No obstante, no se ha valorado la multimodalidad como una herramienta más para segmentar intervenciones, identificar la relación pragmática con otros miembros de la estructura o establecer los roles sociales de los interlocutores en cada punto de la conversación.

El sistema de unidades conversacionales de Val.Es.Co contempla el análisis no solo de las intervenciones verbales, sino también de las no verbales: “podemos encontrar intervenciones que sean miradas, gestos o movimientos corporales y expresiones paralingüísticas, como cambios en el tono de la voz, toses o risas” (Grupo Val.Es.Co. 2014: 18); estas pueden ser intervenciones iniciativas, como gestos de sorpresa o insulto, o reactivas, como la risa (Grupo Val.Es.Co. 2014: 18), “a nod or a contingent action” (Pons y Fischer 2021: 104), “los guiños, las miradas, las risas, los bufidos, los suspiros, los carraspeos, etc.” (Padilla 2020: 243) o, incluso, la presencia de un nuevo interlocutor (Padilla 2004). Sin embargo, esta aceptación de los elementos no verbales como parte de la conversación no ha tenido como consecuencia el análisis de conversaciones grabadas en vídeo que contemplen también la información multimodal o que analicen las funciones de estos elementos en la conversación: “Las miradas o gestos también pueden señalar el hablante posterior, si bien el sistema de transcripción utilizado en estas conversaciones no detalla estos aspectos” (Briz y Grupo Val.Es.Co. 2003: 20).

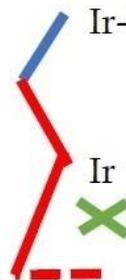
Por ahora, el único acercamiento multimodalidad de Val.Es.Co. al nivel dialógico ha sido la incorporación de los elementos paralingüísticos, “un sistema de signos complementario del lingüístico” (Pons Bordería 2022), en el Corpus Val.Es.Co. 3.0, formado por conversaciones grabadas en audio. Generalmente, acciones fisiológicas como las toses o los bostezos no se han segmentado como partes del discurso porque no han influido comunicativamente. En cambio, ciertos elementos que han aportado un valor modal, textual o interactivo muy evidente, sí se han segmentado; por ejemplo, los chasquidos de lengua como subactos modales o textuales. Algunos poseen “un importante valor interactivo, por lo que su inclusión en las unidades superiores (especialmente en el nivel de la intervención) se hará necesaria cuando supongan o generen reacciones conversacionales” (Pons Bordería 2022). Esto es común con las risas o el sonido nasal (transcrito como *mm*), que son subactos que forman parte de intervenciones o que por sí mismas son intervenciones reactivas (no turno) e, incluso, reactivo-iniciativas (turno) cuando han recibido reacción.

En un estudio preliminar (Cabanes Pérez 2020) al actual, se realizó el análisis de una conversación de 13 minutos contando primero únicamente con la información auditiva y después un segundo análisis incorporando la información visual con el fin de comparar los cambios estructurales y sociales en el nivel dialógico: a las 261 intervenciones del análisis auditivo se le sumaron 32 nuevas intervenciones en el análisis audiovisual; a los 184 turnos se le sumaron 16 nuevos. Estos cambios se produjeron gracias a mejoras en la transcripción y en la identificación de emisores mediante la lectura de labios, a la identificación de receptores mediante la orientación corporal y a la creación de nuevos intercambios mediante intervenciones que estaban formadas únicamente por gestos. En total, hubo 25 intervenciones gestuales, 3 de ellas fueron turno y 6 de ellas ratificaron turnos verbales que no estaban validados en el análisis auditivo.

Sirva como ejemplo el siguiente fragmento de conversación segmentado mediante grabación de audio (1a) y grabación de vídeo (1b) (Cabanes Pérez 2020: 45).

(1)

AUDIO:



Ir-iT	1A1	[¿tú qué tienes] de italiano?¿C1? // hombre en ese nivel ya tienes que entender bastante en todo- las situaciones
Ir	B1	yo me he sacado el [B2 y el C1 de inglés]
	C	[(())]
	B1	y el B2 fue más difícil que el C1 dee- el listening // no sé

(1b)

VÍDEO:



Ir-iT	1A1	[¿tú qué tienes] de italiano?¿C1?
Ir-iT	2C1	{asiente}
Ir-iT	3A2	hombre en ese nivel ya tienes que entender bastante en todo- las situaciones [{encoge hombros}]
Ir	B1	[yo me he] [sacado el B2 y el C1 de inglés]
Ir	C2	{[cuerpo hacia atrás, expone la palma]} ((pero que sea claro))
	B1	y el B2 fue más difícil que el C1 dee- el listening // no sé

El asentimiento funciona como acto con contenido proposicional (afirmación) y como intervención r-i que responde a la pregunta de la intervención 1A1, *¿tú qué tienes de italiano? ¿C1?*; en el análisis auditivo, esta pregunta quedaba sin contestar y se interpretaba como retórica o se presuponía algún tipo de respuesta multimodal porque A continuaba con el marcador *hombre*, que es común en posición inicial de intervención (Briz y Villalba, en línea). El gesto de la intervención C2 y la lectura de labios ayudan a comprender el contenido verbal de la misma, que no se escuchaba bien por estar solapada con la intervención B1.

Estos resultados avalan la investigación actual, pues utilizando el modelo de segmentación de la conversación de Val.Es.Co. –que no solo identifica intervenciones, sino que distingue los turnos ratificados socialmente–, podemos observar cómo los gestos y otra información multimodal descubren nuevas intervenciones e intercambios que

quedan ocultos en el análisis auditivo y cambios en la ratificación social, que pueden afectar de forma global a cómo se interpreta el desarrollo de la conversación y la relación entre los interlocutores.

En el trabajo actual, sobre una muestra mayor, se efectúan dos análisis con más variables que en el estudio previo. En el primer análisis, para observar cómo la multimodalidad afecta a la dimensión estructural y a la social, además de contabilizar el número de intervenciones y turnos globales en el análisis auditivo y en el audiovisual, pretendemos describir en mayor detalle las causas que motivan estos cambios; puesto que en el estudio preliminar se comprobó que las intervenciones gestuales no son la única información que provoca los cambios en la segmentación de intervenciones o su ratificación como turnos. El segundo análisis se centra en la descripción de los gestos que promueven estos cambios. Esta vez no solo se incluye los gestos que por sí mismos forman intervenciones sin ningún contenido verbal, sino también los gestos que acompañan a intervenciones verbales, porque en el estudio piloto observamos que estos también podían afectar al análisis, pero se incluyeron en la misma variable que la lectura de labios o la orientación corporal. Una revisión de la literatura sobre la interacción multimodal y los gestos (§3) aportará criterios funcionales y formales para describir mejor los gestos que afecten al nivel dialógico.

CAPÍTULO 3. LA INTERACCIÓN MULTIMODAL

3.1. INTRODUCCIÓN	115
3.2. LA INTERACCIÓN MULTIMODAL Y SUS ENFOQUES	115
3.2.1. LA MULTIMODALIDAD Y LOS MODOS	115
3.2.2. MODOS PARA UNA INTERACCIÓN MULTIMODAL.....	119
3.2.3. ENFOQUES SOBRE LA MULTIMODALIDAD	122
3.2.4. CONCLUSIÓN	129
3.3. ORIENTACIÓN INTERACTIVA: POSTURA Y MIRADA	131
3.3.1. COMPROMISO CON EL DIÁLOGO	131
3.3.2. GESTIÓN DEL TURNO.....	135
3.4. PARÁMETROS PARA LA CLASIFICACIÓN DE GESTOS	139
3.4.1. GRADO DE CONVENCIONALIZACIÓN.....	139
3.4.2. DEPENDENCIA TEMPORAL CON EL HABLA	145
3.4.2.1. Temporalidad del gesto	145
3.4.3. FUNCIÓN.....	151
3.5. RECAPITULACIÓN	181

3.1. Introducción

Si bien la interacción conversacional se puede explicar a partir de la información oral, es decir, de las intervenciones verbales y sus características prosódicas, el uso de grabaciones en vídeo permite una descripción completa de la interacción y de los elementos visuales que forman parte de ella. En el final del capítulo previo se ha visto un ejemplo del estudio preliminar (Cabanés Pérez 2020), en el que se muestra que la exclusión de la información visual en el análisis a menudo deja interpretaciones erróneas o ambiguas del contenido verbal y de la estructura de la conversación.

El objetivo de este capítulo es revisar los estudios sobre la interacción multimodal y los elementos multimodales que puedan formar parte de la interacción: gestos manuales, expresiones faciales, miradas, acciones del contexto físico, etc. Esta información pretende complementar la revisión de §2 sobre las características sociales y estructurales de la conversación y el turno, de manera que el multimodal se muestre como un enfoque que contribuye a una descripción más completa de lo que ocurre en la conversación. Además, esta revisión también nos ayuda extraer los elementos que contribuyen a la segmentación dialógica de la conversación a nivel estructural y social y que forman parte de nuestro análisis (§5).

En este capítulo, en primer lugar, se explica qué es la multimodalidad, qué modos deben tenerse en cuenta en una interacción multimodal y los distintos enfoques desde los que se han estudiado los elementos visuales de la interacción (§3.2). A partir de esta información, se explica, por un lado, las funciones interactivas de la postura y la mirada (§3.3) y, por otro lado, las características semióticas, temporales y funcionales de otros elementos kinésicos (§3.4).

3.2. La interacción multimodal y sus enfoques

3.2.1. La multimodalidad y los modos

El término *comunicación no verbal*, utilizado por primera vez por Ruesch y Kees (1956), se utiliza para mencionar los sistemas de comunicación que los seres humanos pueden utilizar aparte del lenguaje, entendiendo la comunicación verbal o lingüística como los elementos que forman parte del sistema gramatical y léxico-semántico de una lengua. Aunque el lenguaje permita una transmisión eficaz de contenido y un alto nivel

de abstracción, lo acompañan en su producción otros modos de expresión que complementan de diversas maneras ese contenido.

Sin embargo, este término se ha visto sustituido en los últimos años por el neologismo *multimodalidad* (Payrató 2009, 2016), debido a que *comunicación no verbal* no describe adecuadamente el objeto de estudio. En primer lugar, es un término negativo, que caracteriza al grupo por lo que no es, no por lo que es. En palabras de Birdwhistell, estudiar la comunicación no verbal es como estudiar la fisiología no cardíaca (Knapp 1980). En segundo lugar, “no-verbal” puede dar la impresión de que es un ámbito alejado del verbal, cuando precisamente no debe ignorar el contexto verbal. En tercer lugar, aunque los gestos no tengan una codificación tan restringida como las palabras, algunos siguen unos códigos o patrones bastante cercanos a los lingüísticos y por ello pueden sustituir a las palabras en ciertos contextos. Por tanto, no se pueden distinguir como dos tipos de comunicación. Se opta por nombrarlos como “modos” o “modalidades” dentro de la comunicación, en el sentido de que cada modo utiliza herramientas distintas.

Aun así, *interacción multimodal* y *multimodalidad* también son términos en disputa en el contexto del AC (Deppermann 2013: 2), ya que se acuñaron desde otras disciplinas como la Socio-semiótica y el Análisis del Discurso (cf. Kress 2009, Jewitt *et al.* 2016, Bateman *et al.* 2017, Forceville 2020). Estas estudian todos los recursos semióticos que participan en la comunicación humana y tienen como objeto de estudio géneros visuales como el cómic, el cine, la revista, la canción, la publicidad, etc. Por lo tanto, estos sentidos de multimodalidad no se acomodan al marco del AC. Además, hablar de “modalidades múltiples” puede dar a entender que cada modalidad se trata como una entidad analítica independiente con sus propias asunciones para después juntar los análisis (Deppermann 2013). Esto no es así, la experiencia conversacional es holística, los elementos multimodales funcionan de manera coordinada y sus características destacan gestálticamente en la formación de la acción (Streeck 2013). Por tanto, en este trabajo se utiliza el término *multimodalidad*, entendiendo que la comunicación se construye conjuntamente utilizando varias modalidades que se complementan.

El *modo verbal* está claramente demarcado por el lenguaje, mientras que hay más debate en los elementos que encajan en el resto de los modos y cómo comunican. Otro

Modo verbal
Paralenguaje (Trager 1958)
<ul style="list-style-type: none"> •Cualidades de la voz •Vocalizaciones
Kinésica (Birdwhistell 1952, 1970)
<ul style="list-style-type: none"> •Gestos manuales •Gestos con la cabeza •Gestos con otras partes del cuerpo •Expresión facial •Mirada
Proxémica (Hall 1963, 1966)
<ul style="list-style-type: none"> •Distancia •Postura (Goffman 1961, Kendon 1990, Poyatos 1994)
Otros
<ul style="list-style-type: none"> •Ropa y apariencia física (Duncan y Fiske 1977) •Realidad material: factores ambientales, objetos (Goodwin 2000, 2006, Mondada 2013a, 2013b) •Maneras (Knapp 1972) •Restos de una acción (Knapp 1972)

Figura 4. Modos de la comunicación interactiva

modo que también se percibe por el oído es el denominado *paralenguaje*³² por Trager (1958). Para este autor, estos no son signos lingüísticos, sino cualidades de la voz y vocalizaciones. Estas cualidades son: rango tonal, control labial, control glotal, control entonativo, control articulatorio, control rítmico, resonancia, tempo. Las vocalizaciones son de tres tipos: los caracterizadores vocales³³, como la risa y el llanto, entre otros; los cualificadores vocales (intensidad, altura tonal, *extent*); y las oralizaciones y *clicks* (*vocal segregates*), constituidos por sonidos similares a los fonológicos pero que no forman palabras³⁴ (*ajá*, *mm*, chasquidos linguales, etc.). El primer tipo y el tercero son segmentables, mientras que el segundo no.

En cuanto a los modos percibidos por la vista, Birdwhistell (1952, 1970) denominó *kinésica* a las acciones realizadas por el cuerpo, entre las que incluía los movimientos de manos, de cabeza y de otras partes del cuerpo, la expresión facial y la mirada. El antropólogo propuso un sistema jerárquico de unidades, basado en el estructuralismo de la Lingüística, en el que unidades de menor nivel (*kines*) se combinan

³² Aunque otros utilizan el término paralenguaje para cualquier elemento que acompañe al lenguaje verbal, sea vocal o corporal (Abercrombie 1968, Lyons 1972)

³³ Poyatos (1994:138) los denomina *diferenciadores* y Argyle (1988), *sonidos fisiológicos*.

³⁴ Poyatos (1994:138) los denomina *alternantes*.

para formar unidades de mayor nivel (*kinemas*). Este último es un grupo de movimientos que los informantes ven como uno solo y que porta un significado.

También los estados del cuerpo comunican, es decir, la postura y la distancia. Las posturas son posiciones más o menos estáticas del cuerpo como de pie, sentado, en cuclillas, con las piernas cruzadas, apoyado en la mesa, etc. (Poyatos 1994: 139); pero nos centraremos en la orientación de estas posturas hacia el interlocutor (Goffman 1961, Kendon 1990) en §3.3.1. La distancia como uso del espacio en tanto que elaboración cultural especializada fue lo que Hall (1963) denominó *proxémica*. Hall describió cuatro distancias interpersonales entre humanos (espacio íntimo, espacio personal, espacio social y espacio público) y destacó su impacto en la comunicación interpersonal (1966).

En los enfoques antropológicos o multidisciplinares (Birdwhistell 1952, Argyle 1988, Poyatos 1994) también cualquier movimiento kinésico posee valor comunicativo, sea realizado o no de forma intencionada y consciente, e influye en el receptor o son interpretados por este. Reacciones de nuestro cuerpo como la piel de gallina, el sonrojo, las lágrimas o el sudor, y objetos como la ropa, el perfume o el espacio arquitectónico forman parte de la comunicación y de la interacción como signos pasivos, “ya que todos ellos actúan como componentes interactivos emisores de signos (y como posibles generadores de subsiguientes emisiones) que determinan las características peculiares y únicas de cada intercambio” (Poyatos 2018: 7). Más allá de la información transmitida por el propio cuerpo, algunos investigadores también incluyen la ropa y la apariencia física (Duncan y Fiske 1977); la realidad material, como los factores ambientales y el uso de objetos (Goodwin 2000, 2006, Mondada 2013a, 2013b), las maneras, es decir, cómo se realizan acciones (comer, andar) o los restos de una acción como las huellas (Knapp 1972).

Por la percepción de las modalidades, algunos clasifican la verbal-lingüística y la paralingüística dentro de una *acción audible* (Kendon 2013) o *modalidad vocal/oral* (Stivers y Sidnell 2005), mientras que las otras se perciben mediante la vista y se pueden aunar en una *acción visible* o *modalidad visoespacial*. Como menciona Allwood (2008), si bien existen cinco o más modalidades sensoriales (tacto, calor, etc.), las que se han utilizado realmente en los corpus multimodales son la visión y el oído, que corresponden a las modalidades de producción del gesto y el habla en la comunicación cara a cara. Poyatos (1994: 77) destaca que la percepción de los elementos kinésicos no solo es visual,

puede ser audiovisual si genera sonido, como un golpe en la mesa o contra una parte del cuerpo, o cinestésica si se toca al interlocutor.

3.2.2. Modos para una interacción multimodal

Vistos los distintos modos que conforman la multimodalidad de la comunicación oral, es importante para nosotros distinguir los elementos que en una conversación pueden formar parte de las intervenciones, igual que las emisiones verbales, así como los elementos que afecten a la interpretación del lugar de las intervenciones en la secuencia. En definitiva, qué diferencia la *información* de la *comunicación* (Ekman y Friesen 1969, Burgoon *et al.* 2013).

Desde una perspectiva neogriceana, cualquier signo está abierto a ser interpretado como un intento de elicitación de reconocimiento de intención comunicativa, por tanto, el significado se lo da el que interpreta (Enfield 2011). De esta manera, de cualquier acción se obtiene información gracias a la capacidad humana de inferir creencias; generalmente, el receptor es más consciente de las conductas no verbales del emisor que el mismo emisor, entonces, el receptor percibe más información de la que el emisor es consciente (Argyle 1988). A esto Ekman y Friesen (1974) lo llaman *fuga de información* (*information leak*) y Levinson (1983: 13), *transmisión incidental de información*. El emisor utiliza con intención comunicativa los gestos manuales, pero no puede controlar el rubor, la sudoración o la dilatación de las pupilas, los cuales transmiten información sobre el propio estado. Argyle (1988: 2) lo llama *respuesta fisiológica o comportamiento no verbal* y Wharton (2009), *comportamiento no verbal natural* (frente al *no verbal lingüístico*), que coinciden con expresiones de emociones y gestos menos convencionalizados. Aun así, no siempre es fácil saber el grado de consciencia; por ejemplo, la expresión de emociones, son en parte espontáneas y en parte controladas por normas sociales (Argyle 1988: 4).

Desde el enfoque cognitivo (§3.2.3) se asume que algunos gestos no se diseñan conscientemente para comunicar, sino que son una herramienta cognitiva para el hablante, no dirigida al oyente (McNeill 1992). Algunas pruebas de ello son que los niños ciegos utilizan gestos en la conversación (Iverson & Goldin-Meadow, 1997, 2001) y para expresar pensamientos complejos (Jelec y Jaworska 2014). Sin embargo, para el enfoque interaccional (§3.2.3), los gestos no son meros subproductos del diseño de los mensajes hablados, sino que ellos mismos son productos del diseño. Algunas pruebas de ello son

que los hablantes adaptan sus gestos a la situación comunicativa (Bangerter 2004, Bavelas et al. 2008) y que los oyentes inspeccionan visualmente los gestos (Gullberg y Kita 2009; Streeck 2009a: 104). Para Cooperrider (2017), la gran diferencia teórica entre el enfoque cognitivo y el interaccional es este, que, a pesar de las distintas categorías de gestos, los cognitivistas se han centrado en los *gestos de fondo*, como acompañamiento involuntario del habla, y los interaccionistas en los *gestos en primer plano*, que forman una “señal compuesta” con el habla, comunican algún aspecto crítico del mensaje y de forma consciente el hablante los diseña con cuidado; el gesto es a veces una señal y a veces un acompañamiento involuntario, pero no es ambas cosas a la vez.

Hay ciertos tipos de movimientos que se han descartado como comunicativos de forma generalizada. Por ejemplo, los movimientos de nerviosismo son autoadaptadores (tocarse la ropa, el pelo), que informan de cuestiones internas, pero son involuntarios, o las acciones como fumar, comer, tejer o manipular un objeto, que quizás pueden regular la interacción, creando pausas o manteniendo el turno, pero no son en sí mismos movimientos o turnos conversacionales (Kendon 2004: 8-9). El microanálisis del diálogo cara a cara (*Microanalysis of face-to-face dialogue, MFD*), la metodología utilizada por Bavelas y su equipo (Bavelas et al. 2016), excluye los movimientos involuntarios o reflexivos (parpadeo, respiración), posturas estáticas, apariencias y actos sin función comunicativa como manipulación de objetos o adaptadores (Bavelas y Chovil 2006). Sin embargo, asumen que cualquier acción física puede cumplir una función en la conversación, por lo que ningún movimiento físico o forma puede excluirse directamente sin tener en cuenta el contexto. Si el movimiento cumple una función lingüística, es un gesto (Bavelas 1994). Asimismo, la diferencia entre una acción realizada con la mano y un gesto manual está en las consideraciones sociales y comunicativas (McNeill 1992: 105). Streeck (2009a: 18) explica que se trata de una hipótesis figura-fondo: el significado no es inherente en la experiencia/objeto, sino resultado de su relación con un fondo dado.

Esta dependencia del contexto también la contempla Poyatos, pues modifica la jerarquía entre sistemas primarios y secundarios de comunicación. A pesar de su visión de la CNV *lato sensu*, sitúa en una posición superior a la que llama *la triple estructura básica de la comunicación*, es decir, la comunicación verbal, paralingüística y kinésica (Poyatos 1994: 134), ya que las tres tienen la capacidad de operar simultáneamente de forma superpuesta, alternar mutuamente (es decir, que un elemento paralingüístico o kinésico sustituya a una palabra dentro de una oración) o sustituirse funcionando con

independencia. El resto de los modos no verbales que menciona (reacciones químicas y térmicas, objetos, ambiente, etc.) se consideran, generalmente, sistemas secundarios, que ofrecen información social o cultural y que modifican o refuerzan el significado de los elementos de los sistemas básicos (Cestero 2006: 59); sin embargo, se trata de una generalización y el contexto puede modificar esta jerarquía, “una subida de temperatura, el sonrojo o un silencio pueden encerrar el mensaje principal cuando el lenguaje, el paralenguaje, la kinésica o la proxémica simplemente lo apoyan como información complementaria o lo duplican como verdadera redundancia” (Poyatos 1994: 83).

Otros rasgos para distinguir un gesto con intención comunicativa, además del contexto, son la orientación del movimiento, el esfuerzo para su realización o su sincronía con el habla (Enfield 2009). Mientras los signos convencionales, como las palabras, deben su propia existencia a su papel en la comunicación, que les da un carácter de direccionalidad, los movimientos corporales son signos potenciales y su direccionalidad debe estar especialmente marcada por un activador para que otra persona los interprete como dirigidos a ella, a través de gestos de señalamiento o déicticos verbales, la coocurrencia espaciotemporal, la orientación corporal o un marcado diacrítico especial (Enfield 2009: 16). Por ejemplo, los movimientos de la cara y la cabeza pueden servir de desencadenantes para que la mirada se interprete como un señalamiento y no solo como una mirada.

Un mayor control y esfuerzo sobre el comportamiento es más probable que tenga una intención comunicativa (Enfield 2009, Cooperrider 2017). Rasgos de la acción como un gran tamaño en anchura y/o altura, la fuerza ejercida, la precisión del movimiento o el uso de partes del cuerpo más grandes sugieren más esfuerzo y un compromiso motor mayor, lo que implica la consciencia e intencionalidad del emisor al realizarlo (Vajrabhaya y Pederson 2018: 83-84, Bavelas et al. 2008).

También gesto y habla se vinculan mediante la sincronía, de manera que se asume que forman parte de un significado unificado, único y dirigido (Kendon 2004). Para Enfield (2009) los gestos manuales coverbales siempre están asociados al habla independientemente de que tengan la intención de comunicar o no; en cambio, para Kendon, la clave para distinguir un gesto como elemento comunicativo es que la acción corporal visible debe tener una función semántica o pragmática, si no la tiene, no es intencional y no es enunciado (Kendon 2013). De manera que el gesto no solo expresa contenido, sino que es una forma de acción social (Kendon 2017). Por otro lado, la

ausencia de habla también puede confirmar la pretensión comunicativa del gesto: “if speech is entirely absent from a communicative turn, it is clear that the gesture is intended to bear the full message” (Cooperrider 2017), así como cuando sustituye a un fragmento del turno hablado (§3.4.2.4.).

Por último, en la forma de los gestos pueden reconocerse significados. Los gestos manuales que suelen considerarse comunicativos encajan fácilmente en el modelo triádico de semiosis de Peirce (Iriskhanova y Cienki 2018:27): la relación de similaridad la realizan los gestos icónicos o metafóricos; la relación de contigüidad, los deícticos; y la relación de convención, los emblemas. Asimismo, gestos faciales, de cabeza o de hombros pueden tener un significado convencionalizado (§3.4.1).

3.2.3. Enfoques sobre la multimodalidad

Actualmente, hay dos marcos importantes en los estudios del gesto, uno de orientación cognitiva y otro de orientación interactiva (Cooperrider 2017, Vajrabhaya y Pederson 2018). Ambos son funcionales porque parten de que la estructura del lenguaje refleja los objetivos de los usuarios de la lengua, pero la mayoría de los enfoques funcionales coinciden en que el lenguaje cumple simultáneamente una función cognitiva, que organiza y representa el pensamiento, y una función social, que permite coordinar el comportamiento propio con el de otras personas (Kok et al. 2016). Se puede decir que el enfoque cognitivo se centra en el hablante y el interaccional en su relación con el oyente (Vajrabhaya y Pederson 2018: 67-70). También hay diferencias metodológicas (Cooperrider 2017): la investigación en la cultura cognitiva suele ser experimental, cuantitativa y llevada a cabo por psicólogos y científicos cognitivos; mientras que la investigación en la cultura interactiva tiende a ser observacional, cualitativa y la realizan por analistas de la conversación y antropólogos.

El padre del *enfoque cognitivo* de los gestos es el psicólogo y lingüista David McNeill (1985, 1992, 2005), quien se vale de los gestos que acompañan al habla como una ventana involuntaria a la mente que permite explicar el proceso cognitivo que gobierna el habla. Desde su hipótesis del *Growth Point*, gesto y habla surgen de un mismo proceso cognitivo de formación del enunciado, que tiene un lado visual (*imagistic*) y un lado lingüístico (*linguistic*) (McNeill 1992: 29); en consecuencia, los gestos tienen una relación co-expresiva con el habla y no deberían denominarse *no verbales*, pues están integrados temporalmente y sus significados están relacionados semántica o

pragmáticamente. Ofrece una clasificación semiótica y semántica del gesto, así como la distinción entre los gestos que expresan contenido proposicional de los que no, o sea, de los pragmáticos. Aunque no se centra en los gestos como componentes del enunciado o su rol en la interacción social, sino en los gestos proposicionales no convencionalizados y en cómo ayudan al hablante cognitivamente en el proceso de producción del habla, estas tipologías y descripciones han sido útiles para otros enfoques del estudio del gesto, como el interaccional y el lingüístico.

En el *enfoque interaccional* el enunciado también está formado por modalidades distintas a la palabra (gestos, prosodia, etc.), pero estos estudios se centran en el gesto como herramienta básica para llevar a cabo la acción social. En la base de este enfoque se encuentra Adam Kendon, que fue uno de sus contribuyentes más prolíficos. Inspirado en Birdwhistell y en Condon, basa su teoría en el Análisis del Contexto³⁵, pero también en Goffman (1981) y en su propia visión del gesto como movimiento en la interacción (Kendon 1990). El Análisis del Contexto y el Análisis de la Conversación se desarrollaron como campos de investigación exclusivos, con dominancia del segundo y su restricción a la interacción verbal (Bohle 2014: 1306). Sin embargo, hubo algunas contribuciones tempranas del AC que expandían el análisis a lo visual, como el análisis de Sacks y Schegloff (2002³⁶) de la *home position* del gesto, el estudio de la torsión corporal de Schegloff (1984) y, especialmente, el estudio de Charles Goodwin (1981) sobre la mirada en la producción del turno. La inclusión de los elementos visuales no solo descubre procesos ocultos o malentendidos en el análisis auditivo, permitiendo comprender mejor prácticas ya conocidas, sino que también replantea las nociones básicas del AC³⁷ (Deppermann 2013: 4): los gestos y gestión del turno (Bavelas et al. 1992, Schmitt 2005, Mondada 2006, 2007, Bohle 2007), la relación entre el carácter temporal del gesto y la

³⁵ El Análisis del Contexto deriva de las ideas formuladas por la Escuela de Palo Alto sobre la organización del comportamiento, tanto visual como verbal. La grabación y estudio de entrevistas psiquiátricas y conversaciones suscitaban propuestas sobre la coordinación de los recursos multimodales, la situación de la interacción, su disposición espacial y el contexto material (Kendon 1990: 17-20). Entre los participantes de este grupo estuvo Birdwhistell, que aportó las técnicas de transcripción de la lingüística descriptiva e inspiró métodos de trabajo microanalíticos para el comportamiento en la interacción (movimiento, espacio y postura), así como contribuyó a la búsqueda de unidades estructurales, cómo participan, cómo se definen en su nivel de organización y cómo se organizan en un todo. El psiquiatra Albert Schefflen fue uno de los principales defensores del Análisis del Contexto en los años 60 y 70 (Schefflen 1966) y lo aplicó a sesiones de psicoterapia (Schefflen 1973), junto a Condon, Duncan y Erikson.

³⁶ Publicado en 2002, pero basado en una conferencia en la American Anthropological Association en 1975.

³⁷ Esta revisión crítica puede considerarse una escuela nueva, denominada *Comunicación Multimodal* (Bohle 2014: 1303-1305).

estructura secuencial (Mondada 2013a, 2013b), las acciones cooperativas (Goodwin 2018) y las actividades comunicativas corporeizadas (Streeck 2009a, 2017).

Los recursos multimodales se han expandido de los gestos al contexto: el contexto espacial (C. Goodwin 2000, 2007, 2010, Mondada 2013a, 2013b), la traslación de los participantes por el espacio (Haddington, Mondada y Nevile 2013), la especialización en entornos de trabajo como la interacción doctor-paciente (Heath 1986), una sala de control del metro (Heath and Luff 1992), excavación arqueológica (C. Goodwin 2000), galerías (vom Lehn et al. 2005). Esta mirada hacia cómo el contexto y el complejo social afecta a los turnos y las secuencias difumina la distinción entre Análisis del Contexto, Análisis de la Conversación y (Micro)etnografía (Streeck 2009a: 28).

Aunque es evidente el uso de gestos en la comunicación, los gestos y el comportamiento no verbal han sido menos estudiados desde la lingüística que desde otras áreas. Ni el enfoque cognitivo ni el interaccional nace en la Lingüística sino en la Antropología, la Psicología y otras ciencias sociales. Sin embargo, recientemente ha aparecido un *enfoque lingüístico* del gesto, que considera cuestiones de significado y de funciones dentro del enunciado, así como las aportaciones pragmáticas del gesto.

Generalmente, los estudios lingüísticos sobre gestos se han restringido a la Pragmática (Payrató y Teßendorf 2014: 1531) y, aun así, de forma reciente (Kendon 2017). Esta restricción se debe que la Pragmática trabaja con elementos lingüísticos en su realización situada y la multimodalidad se interpreta como parte de ese contexto, no como parte del sistema lingüístico (Cienki 2017). De ahí la división de la comunicación verbal como lingüística y la comunicación no verbal como pragmática. Sin embargo, la Lingüística Cognitiva ve la Semántica y la Pragmática como dos finales de un *continuum* para caracterizar el significado (Cienki 2017: 61-63), por lo que la función del gesto tampoco debe clasificarse como pragmática y extralingüística directamente, ya que no siempre se aleja de la naturaleza semántica del lenguaje, tal y como han defendido Kendon (1980, 1994, 2004) y McNeill (1985, 1992, 2005). Por consiguiente, los enfoques lingüísticos del estudio de la multimodalidad sostienen que el habla y los gestos se basan en principios estructurales y semióticos similares (Bressem 2021: 6).

Müller, Ladewig y Bressem (2013) ofrecen una visión general del estudio de los gestos desde una perspectiva lingüística, de la que extraemos cuatro líneas de estudio: la

relación forma-significado de los gestos, la integración sintáctica y semántica del gesto en el enunciado, las metáforas multimodales y la sincronización del gesto con la prosodia.

Respecto a la relación forma-significado de los gestos, analiza como signos motivados o inmotivados, globales o segmentales, sintéticos o analíticos (McNeill 1992, 2005); también se ha intentado descomponer su significado hasta cierto grado en rasgos de forma convencionalizados (Calbris 1990, Kendon 2004, Müller 2004) a partir de los parámetros potencialmente significantes de las lenguas de signos. Sobre la “gramática” del gesto véase Fricke (2007, 2012), Müller (1998), Müller, Bresse y Ladewig (2013).

La perspectiva de una "gramática" del gesto apunta, pues, a una documentación lingüística sistemática y basada en la forma de los patrones gestuales. Se refiere a las propiedades básicas de forma de los gestos, a sus estructuras y a revelar su potencial para el lenguaje. Se asume que los gestos siguen principios de creación de significado, construyen unidades de diferente complejidad y muestran distintos grados de lexicalización y gramaticalización (*trad.* Bresse 2021: 7).

Este enfoque también ha llevado al estudio de la integración sintáctica y semántica del gesto en el enunciado. Por ejemplo, la relación con partículas verbales deícticas (de Rutter 2000, Fricke 2007, Kita 2003, Streeck 2002); la semantización y la gramaticalización de gestos, como los vinculados a la negación verbal (Harrison 2009, 2010); la integración de gestos en la estructura sintáctica (Andrén 2010, Bohle 2007, Clark 1996, Goodwin 1986, 2007, Enfield 2009, Langacker 2008, McNeill 2005, Slama-Cazacu 1976, Streeck 2009a); y la contribución semántica del gesto al enunciado (Bavelas, Kenwood y Phillips 2002, Bresse 2012, Calbris 1990, Engle 2000, Freedman 1977, Fricke 2012, Kendon 2004, Ladewig 2011, 2012, McNeill 1992, 2005, Slama-Cazacu 1976)³⁸. Se podría hablar de una Gramática Multimodal (Fricke 2012) a partir de los estudios lingüísticos que se enfocan en los gestos que pueden cubrir los mismos usos que las categorías verbales (véanse gestos de uso adverbial en Bresse 2012) y también funciones sintácticas (Ladewig 2012), pues han demostrado que los gestos muestran relaciones tanto paradigmáticas como sintagmáticas y tienen potencial para la combinatoria y las estructuras jerárquicas (Bresse 2012: 12). “Los gestos no se consideran unidades lingüísticas propiamente dichas, sino que asumen funciones de

³⁸ Fox et al. (2013: 730) también mencionan algunos estudios lingüísticos multimodales sobre la combinación de cláusulas (Fox 2001), la producción conjunta de enunciados (Hayashi 2003a), la risa (Ford y Fox 2010), las respuestas a preguntas de tipo wh (Fox y Thompson 2010) y la referencia (Eriksson 2009).

unidades lingüísticas en colaboración o intercambio con unidades lingüísticas vocales” (*trad.* Bressem 2021: 6).

Desde la Lingüística Cognitiva se pueden explicar los gestos que contribuyen a expresiones multimodales de eventos de movimiento lexicalizados (Talmy 1985, 1991), a la formación de metáforas multimodales en conjunción con el habla (Cienki y Müller 2008a, Müller y Cienki 2009) o al *conceptual blending* (Parrill y Sweetser 2004). Y, puesto que la prosodia es un modo que no ha sido desestimado por la Lingüística, se ha analizado la sincronización del gesto con la prosodia (Birdwhistell 1970, Loehr 2007, McClave 1994).

Finalmente, “connecting gesture form analysis with a pragmatic perspective addresses gestures as communicative action” (Müller en prep.). Desde la Pragmática, se ha estudiado la relación del gesto con la formación del enunciado (*composite utterances*), su uso en la interacción y en el discurso (Enfield 2009), la dependencia praxeológica de la forma y el significado del gesto (Streeck 2009a), y su relación con el dinamismo comunicativo (McNeill y Levy 1993: 365). Según Payrató y Tessedorf (2014), hay pocos estudios pragmáticos de gestos, excepto los de perspectiva interaccional de Bavelas et al. 1992, 1995, Goodwin 1986 y del AC (Heath 1992, Mondada 2006, Bohle 2007, Poggi 2004, Sherzer 1991, entre otros). Se podría decir que en cierto punto el enfoque lingüístico pragmático y el interaccional se encuentran porque atienden a los mismos fenómenos del discurso. La diferencia del enfoque pragmático es que aplica la teoría sobre los actos de habla (Austin 1962; Searle 1969) a los gestos, buscando su fuerza ilocutiva o la acción comunicativa que realizan. Un buen ejemplo de esto es Kendon, cuya perspectiva interdisciplinar le hace tomar unidades funcionales que provienen de la lingüística estructural para el análisis del comportamiento corporal (Kok et al. 2016). Atiende a las propiedades de formación de los gestos, su integración con el habla, sus propiedades semióticas y su papel en la interacción. Su estudio de la coordinación entre movimiento y habla (Kendon 1972, 1980) le lleva a definir el enunciado como objeto construido con el habla o con una acción corporal visible o por una combinación de ambas modalidades (Kendon 2004: 110). Algunos nombres que se han dado para los enunciados que combinan ambos modos son *composite signals* (Clark 1996, Engle 2000) o *composite utterances* (Enfield 2009). Con este enfoque se puede estudiar cómo los gestos contribuyen a las construcciones multimodales con un significado pragmático (Bressem

y Müller 2017, Ladewig 2020) o la construcción espontánea de actos gestuales a partir de acciones manuales (Streeck 2017, 2021).

En cuanto al ámbito hispanico, no se han desarrollado muchos estudios sobre multimodalidad, salvando algunos autores destacados (Poyatos, Cestero, Payrató). El interés es bastante reciente y se ha centrado en la enseñanza de español como lengua extranjera, aunque también destacan recientes trabajos de enfoque cognitivo.

Inicialmente, quien desarrolló teóricamente los gestos en español fue Fernando Poyatos (1972, 1987, 1992, 1994). Poyatos hace hincapié en el abordaje interdisciplinar, pero vemos una base semiótica importante por la inclusión de todo tipo de signos en su definición de comunicación no verbal: “las emisiones de signos activos o pasivos, constituyan o no comportamiento, a través de los sistemas no léxicos somáticos, objetuales y ambientales contenidos en una cultura, individualmente o en mutua coestructuración” (Poyatos 2003: 68). Asimismo, también sigue un criterio lingüístico por las equivalencias entre gestos y elementos verbales. Define el emblema como “un gesto que tiene un equivalente verbal sin ninguna ambigüedad en su propia cultura” (1994: 187-189) y señala *marcadiscursos kinésicos* que equivalen a pronombres, adjetivos, sustantivos, preposiciones y conjunciones (1994: 190-192). Su visión del gesto anclada a la palabra se observa en diez maneras en las que el paralenguaje y la kinésica influyen en esta (Poyatos 2018): sustituyendo, añadiendo información, como fuga de información; apoyando, realzando su significado, debilitando, contradiciendo, repitiendo lo expresado, enmascarando, economizando su uso. Aun así, Poyatos también selecciona gestos interactivos que participan en la gestión del turno de hablante y oyente (1994: 230-246).

Siguiendo los pasos de Poyatos, Ana María Cestero ha estudiado elementos no verbales (1996, 1998, 2006, 2016) y ha creado inventarios de gestos españoles: *Repertorio básico de signos no verbales del español* (1999) y *Diccionario audiovisual de gestos españoles* (Cestero et al. 2020). Este último ofrece muestras de signos no verbales en diálogos en formato audiovisual y tiene el objetivo de ser material complementario para el aprendizaje de ELE, pero también se concibe como material de consulta para antropólogos, lingüistas, psicólogos y especialistas en lenguas de signos. Frente a otros diccionarios similares (Coll, Gelabert y Martinell 1990, Martinell y Ueda 1998, Gaviño Rodríguez et al. 2011) tiene la ventaja de que el repertorio se estructura en torno a funciones, extraídas del del Plan Curricular del Instituto Cervantes (Instituto Cervantes 2006) y otros diccionarios o repertorios de gestos españoles específicos para ELE.

La mayor parte del estudio de los gestos en el ámbito hispánico enfoca sus objetivos en la enseñanza de ELE (Forment Fernández 1997, Martinell y Forment Fernández 1998, Jashiashvili 2017, Monterubbianesi 2013, Moreno García y Tuts 2006, Schmidt 2013, Yagüe 2004) y para ello contrastan emblemas españoles con los otras lenguas (García García 2004), con el portugués brasileño (Nascimento Dominique 2005, 2012), con el italiano (Betti 2007), con el japonés (Ueda 1998, Quintero García 2005), con el chino (Sánchez Benítez 2009), etc.

Desde la Lingüística Cognitiva, hay algunos trabajos sobre gestos rítmicos y su relación con la prosodia y el desarrollo del lenguaje (Esteve- Gibert 2016; Esteve-Gibert y Prieto 2013; Esteve-Gibert, Igualada y Prieto 2019); la expresión multimodal de la prominencia (Jiménez-Bravo y Marrero 2020); gestos que ayudan a interpretar la doble negación (Prieto y Borrás-Comes 2013); gestos que codifican eventos de movimiento (Peiró Márquez e Ibarretxe Atuñano 2021); gestos que expresan distancia temporal (Alcaraz Carrión y Valenzuela 2019); patrones multimodales de (des)alineación conversacional (Olza 2022); y gestos de acuerdo y desacuerdo (Yániz Ciriza 2022).

Lluís Payrató toma el enfoque pragmático con trabajos sobre los emblemas y las funciones pragmáticas (1993, 2003, 2014, 2016, 2019). También plantea la inclusión de los emblemas y los gestos coverbales en una gramática multimodal que parta de la Gramática Funcional Discursiva (*FDG*, Hengeveld y Mackenzie 2008), pues esta utiliza como unidad básica el acto (Alturo, Clemente y Payrató 2016).

Además, se pueden encontrar otros trabajos vinculados a la pragmática y la interacción: los gestos emblemáticos (Belío-Apaolaza 2019, 2022); la distinción de gestos referenciales y pragmáticos en su contraste entre L1 y L2 (López-Ozieblo 2016); los gestos empleados en la ironía (Cestero 2006, 2009), los gestos relacionados con funciones pragmáticas en el desarrollo discursivo conversacional (Muñoz Trujillo y Guerrero González 2019); los gestos en la organización del discurso (Montes 1994, Inhoffen 1996); los gestos coexistentes con marcadores temporales (López-Ozieblo 2019); las funciones discursivas de los gestos de autoseñalamiento (Montes 2009); los gestos de disconformidad (Montes 2002); el señalamiento mediante los labios (Alvarado Cantero 2016); la repetición del gesto en el discurso interactivo (Tabensky 2004); la multimodalidad en narraciones interactivas infantiles (Alam, Rosemberg y Scheuer 2020).

3.2.4. Conclusión

La *multimodalidad* es un término multidisciplinar para cualquier medio en el que se utilizan varios modos de comunicación, pero la definición que tomamos para nuestro estudio es la de Mondada (2013b: 580): “the integrated study of all the relevant linguistic, embodied, and material resources participants mobilize for organizing social interaction in an audible-visible intelligible way”. Considerando nuestro objetivo, la multimodalidad es una pieza fundamental en tanto que profundiza y amplía las cuestiones estudiadas desde la Pragmática y el Análisis de la Conversación. Las relaciones que se establecen entre las intervenciones de la conversación y la estructura dialógica que forman no pueden describirse adecuadamente sin tener en cuenta la multimodalidad.

En este apartado se ha mostrado cuáles son las modalidades que se suelen incluir cuando se hace referencia a la multimodalidad interaccional: modo verbal, paralingüaje, kinésica, proxémica y otros modos sociales y contextuales (apariencia física, ambiente, objetos, etc.). El modo verbal y el paralingüístico ya son considerados en la segmentación dialógica de las unidades conversacionales de Val.Es.Co., pues son las que pueden captarse por la grabación de audio. Son los modos visuales los que podemos captar en la grabación de vídeo y considerar en el análisis, a los que denominaremos a menudo *elementos visuales*, a semejanza de Selting (2013: 591).

Los modos vinculados a los objetos que no forman parte de acciones intencionales con funciones lingüísticas quedan excluidos. Es decir, la apariencia de los hablantes o el espacio físico puede tener efectos en la relación social de los hablantes, en lo que dicen y en como lo dicen, pero no son factores que vayamos a tener en cuenta en el análisis de la estructura de la conversación. Asimismo, la distancia física entre los interlocutores tampoco se considera, pues es un factor importante en estudios interculturales y sobre la relación social entre interlocutores, pero, de nuevo, no para la estructura de la conversación. Sin embargo, sí incluimos en nuestro estudio la postura, en tanto que orientación corporal entre los interlocutores que puede indicar la relación entre emisor y receptor de la intervención, así como los demás componentes kinésicos: gestos manuales, gestos con la cabeza, gestos con otras partes del cuerpo, expresión facial y mirada.

También hemos visto que cualquier movimiento corporal no es equivalente a un gesto con intencionalidad comunicativa. Los criterios para identificarlos son el contexto de acciones lingüísticas y sociales, la orientación del movimiento hacia un receptor, el esfuerzo para su realización, su grado de convencionalización y su vinculación temporal

y funcional con el habla. A lo largo del capítulo, desarrollamos algunos de estos criterios que sirven para crear tipologías de los gestos.

Entre los elementos visuales mencionados, establecemos una diferencia tanto en la revisión bibliográfica de este capítulo como en el propio análisis: por un lado, la postura y la mirada; por otro lado, el resto de los componentes kinésicos. Esta división se debe a las características formales, temporales, semióticas y funcionales de cada tipo de movimientos. De hecho, no son elementos que se suelen analizar con los mismos parámetros y se estudien en conjunto. La complejidad de describir todos ellos con la misma base teórica es la razón de esta división. En §3.3 se explica el uso de la postura y la mirada para establecer los marcos del diálogo y gestionar el turno. En §3.4 y sucesivos se explican las características y funciones del resto de los gestos.

Respecto a los enfoques vistos en §3.2.3, entre los dos enfoques preponderantes, nuestro estudio se encuentra dentro del interaccional, no del cognitivo. Puesto que tiene una concepción de la multimodalidad como herramienta para llevar a cabo la interacción social y organizar el orden secuencial y, además, utilizamos una metodología observacional y cualitativa. Sin embargo, aunque tomamos ideas del AC, tanto del orden de los elementos verbales como de los multimodales, nuestro enfoque también es pragmático. Nuestro análisis estructural de la conversación se sirve de las unidades conversacionales de Val.Es.Co., cuya jerarquía parte del acto como unidad mínima estructural. En consecuencia, una concepción lingüística del gesto, que se rige por principios estructurales y semióticos similares al lenguaje y que puede tener funciones lingüísticas dentro del enunciado en colaboración o intercambio con la información verbal, puede adaptarse a la jerarquía del modelo.

La revisión bibliográfica de los gestos, pragmática e interaccional, de este capítulo nos da la información necesaria para comprender los gestos como parte de acciones comunicativas y su uso en la interacción. Desde las funciones de los gestos de Kendon (2004), revisamos qué funciones semánticas y pragmáticas se les ha dado a los gestos en la formación de enunciados. Esto nos sirve para observar en nuestro estudio qué funciones y qué formas tienen los gestos que afectan a la segmentación dialógica de la conversación.

No se ha realizado anteriormente un estudio que combine una segmentación jerárquica de las intervenciones de la conversación con la multimodalidad; por consiguiente, nuestro estudio pretende encontrar resultados novedosos al explorar la conversación desde este marco teórico. Aún menos se ha realizado en español, pues el

estudio de la multimodalidad en esta comunidad se ha desarrollado más dentro de la enseñanza de E/LE o de la Lingüística Cognitiva.

3.3. Orientación interactiva: postura y mirada

Aunque dentro de la kinésica se incluye la mirada (Birdwhistell 1952) y la proxémica (Hall 1963), la función principal de la mirada y de la postura es reguladora, así que Kendon (1990, 2004: 8-9) considera que la acción de mirar a una persona, cambiar la postura o establecer la orientación espacial del encuentro no tiene significado intrínseco, solo se entiende en relación con otros elementos. Es decir, no es posible encontrar el significado de la mirada o de la postura, sino la función metadiscursiva que tiene en la estructura. También Argyle (1988: 1) dice de la mirada que es más un canal que una señal.

Los criterios que exponen Bavelas, Coates y Johnson (2002) para incluir la mirada como un acto de significado visible son, a nuestro parecer, exclusivamente reguladores del turno y dependientes de este: 1) la mirada es sensible a la disponibilidad de un receptor; 2) está sirviendo a una función simbólica, el hablante no mira al oyente meramente por verlo, sino para pedir una respuesta; 3) hay evidencia de que los oyentes entienden el significado de la mirada porque responden inmediatamente y apropiadamente; 4) la mirada está completamente integrada con las palabras y otros actos visibles.

En este apartado explicamos, por un lado, el uso de la postura y la mirada para comprometerse en la participación y, por otro lado, las funciones de gestión del turno de ambos modos.

3.3.1. Compromiso con el diálogo

La copresencia no es el único requisito para que la conversación se establezca. Goffman (1961) distingue una *interacción no enfocada*, que expresa acciones del hablante de manera inintencionada por el hecho de compartir el espacio, de una *interacción enfocada*, en la que los individuos se ponen de acuerdo para sostener una conversación u otra actividad. Este acuerdo se muestra mediante la organización de las acciones de su cuerpo, de manera que posibiliten la visibilidad hacia otros y, además, comuniquen su orientación al compromiso en la actividad con diferentes roles de participación. Se ha denominado *ecological huddle* (Goffman 1961) o *F-formation* (Kendon 1990) a las configuraciones

orientativas y posturales de unos cuerpos hacia otros en el espacio para la interacción. La postura y la mirada son los elementos fundamentales para indicar esta orientación (Goffman 1963). El área transaccional de los individuos se superpone para crear un marco de participación e informarse mutuamente de que comparten una atención común. Este espacio de interacción es adaptativo a los participantes involucrados, pudiendo rehacer su forma en un polígono de más o menos aristas para facilitar la entrada o salida de los involucrados (Müller y Bohle 2007, *apud.* Hagemann 2014: 1316).

Goodwin también ha mostrado interés en otros elementos externos a los individuos en este marco de participación: la manera en que la configuración contextual contribuye a la organización de la interacción (2000, 2003, 2007) y la orientación hacia objetos físicos relevantes en la interacción (Goodwin 2009). Predominan los estudios de interacción en un lugar estático, pero Mondada (2013b: 585) está interesada en la interacción en situaciones móviles, como paseando o yendo en coche, de manera que los interlocutores deben ajustarse al espacio público, organizarse mediante la proximidad o la velocidad.

La mirada mutua es la precondition que transforma la mera copresencia en la ratificación mutua del compromiso en una interacción social. En este sentido, mirar al hablante es una forma de demostrar que el oyente está involucrado en la interacción, mientras que la falta de mirada del receptor puede considerarse como una falta de interés o, al menos, sancionada como problemática por el hablante (Goodwin 1981). Según el contexto, se puede llegar a procedimientos invasivos hacia un receptor poco dispuesto a mirar, que implican tocarle, ponerse en su línea de visión e incluso agarrar su cara para girarla hacia sí (Kidwell 2006).

La mirada y la postura transmiten grados de compromiso (*engagement*), particularmente cuando uno está involucrado en actividades simultáneas (Goodwin 1981, Kendon 1990, Schegloff 1998). La torsión del cuerpo (Schegloff 1998), mediante la desalineación de cabeza, torso y piernas, indican que el participante sigue formando parte de la unidad de interacción, pero su atención está temporalmente en otro foco o actividad. Cuando todos los segmentos se alinean en la misma dirección, se comunica una única orientación dominante; cuando no, comunican orientaciones múltiples simultáneas que clasificables por nivel de estabilidad. Algunos estudios relacionados con esto son el de Tiittula (2007 [Hagemann 2014]), que estudia los cambios en el eje de orientación en contextos de trabajo y los relaciona con peticiones verbales de tareas específicas y

cambios entre acciones relacionadas con negocios o no; y el de Kamunen (2019), donde la jerarquización de las actividades se manifiesta en la torsión del cuerpo y la suspensión de una actividad manual para priorizar una reparación verbal.

Los participantes alternan entre la presencia y la no presencia y se monitorean entre sí para comprobar que el otro está dispuesto a participar o no (Goodwin 1981: 96). Mediante muestras de *engagement*, como la mirada y la orientación corporal, se hacen relevantes las acciones para la conversación (Goodwin 1981: 95-97). En cambio, muestras de *disengagement*, como no mirar a otros participantes en combinación con la realización de otras actividades o con una posición que no permite percibir las acciones de otra persona, indican que el individuo no está dispuesto a una actividad colaborativa (Goodwin 1981: 98-101). Cuando el hablante va a abandonar el compromiso, el destinatario muestra su entendimiento a la propuesta de cambio en el estatus de participación y lo acepta organizando sus propios movimientos para entrar en el mismo estado (Goodwin 1981: 107).

Aunque durante el encuentro puede haber lapsos de mutuo *disengagement*, la interacción no ha terminado si los participantes se mantienen físicamente cercanos entre ellos o si siguen orientados corporalmente; incluso con una falta de orientación hacia el otro, los participantes están, de hecho, monitoreando las acciones del otro por si se da una reanudación de la actividad (Goodwin 1981: 98-101). De alguna forma los participantes están *engaged* durante el *disengagement*, e incluso es posible producir habla que no propone un completo *engagement* y no exige respuesta (Goodwin 1981: 113). Por ejemplo, para dar por finalizada la interacción, las muestras de *disengagement* no tienen por qué coincidir con el fin: un participante puede empezar a alejarse del marco participativo e iniciar otras actividades y al mismo tiempo emitir su último turno. De esta forma anticipa el final de la interacción (Goodwin 1981: 107).

En una conversación con varios participantes no todos participan activamente. Para Kendon (1973) el hablante y el oyente principal están conectados por el eje de interacción, cuyo origen es la coordinación de movimiento y la sincronía interactiva; después distingue los *non-axial listeners* y aquellos que temporalmente se retiran de la interacción. Generalmente, la mirada del hablante sirve para determinar el destinatario principal (*participant*) (Lerner 2003), mientras que otros participantes (*party*) comparten el espacio, pero no son destinatarios directos y se quedan observando la interacción de los interlocutores principales. Sin embargo, en estas interacciones con más interlocutores

hay patrones en los que el uso de las distintas modalidades cambia. A veces la mirada del hablante no es suficiente para establecer eficazmente el destinatario, porque el receptor no ve que se dirigen a él o los otros participantes no ven el intercambio de miradas entre emisor y receptor y responden en su lugar (Lerner 2003: 180-181)³⁹. Mientras una conversación de dos hablantes la mirada puede ser suficiente, con más hablantes el giro de cabeza es más visible y relevante (Jokinen *et al.* 2009).

El comportamiento gestual difiere tanto entre hablante y destinatario como entre destinatarios primarios y secundarios. El destinatario primario gestualiza y se mueve más que el secundario y mira más al hablante, por lo que normalmente el siguiente turno lo coge el destinatario primario (Battersby 2011, Jokinen *et al.* 2013); mientras que los destinatarios secundarios tienen la misma probabilidad de mirar a cualquiera de las partes (Battersby y Healey 2010). Como las personas solo pueden mirar a un único participante a la vez, tanto la orientación de la cabeza (Jokinen *et al.* 2009) como la de las manos cobra importancia sobre la mirada por su precisión. La mirada del oyente no destinatario sirve para que el hablante lo incluya dirigiéndole la palabra (Kidwell 1997), pero los oyentes utilizan más la orientación de las manos o estas combinadas con la cabeza que solo la cabeza y los hablantes son significativamente más propensos a invocar una respuesta de los receptores con los cambios en la orientación de las manos (Battersby y Healey 2010).

Frente a la *mirada mutua* entre dos interlocutores, Healey y Battersby (2009) denominan *triangulación* a la disposición espacial de tres participantes para mostrar su participación simultánea. Restringiéndose a la orientación de la cabeza y los gestos, existen tres posibilidades espaciales básicas: 1) el hablante se orienta hacia el destinatario secundario con un gesto mientras sigue orientándose hacia el destinatario principal con la cabeza; 2) el hablante se orienta hacia el principal con un gesto y se orienta hacia el secundario con la cabeza o 3) el hablante utiliza una combinación de orientación de cabeza y gesto hacia el secundario. En el 63% de los casos, el secundario responde al cambio de orientación alejándose del principal y dirigiéndose al hablante.

³⁹ En esos casos también las palabras apelativas son más eficaces (Lerner 2003: 180-184): el pronombre “tú” requiere acompañamiento de la mirada para desambiguar el destinatario, pero favorece que los participantes comprueben la dirección de la mirada; el uso del nombre propio u otro término identificador es menos común, normalmente cuando el contexto no permite utilizar la mirada.

3.3.2. Gestión del turno

Mirada

En cuanto a la función reguladora de la mirada, Kendon (1967) provee de una explicación sobre la dirección de la mirada que ha sido la base para el microanálisis de las miradas en la interacción conversacional⁴⁰. Según este trabajo, hablante y oyente se dirigen miradas durante la interacción para monitorear sus acciones y su atención, pero el tiempo que el hablante mira al oyente es menor que el tiempo que el oyente mira al hablante⁴¹ y siguen un orden secuencial. En enunciados de más de cinco segundos, los hablantes tienden a retirar la mirada del oyente desde el principio y la redirigen al aproximarse al final del enunciado, entonces el nuevo hablante comienza su turno y aparta la mirada⁴². Cuando este patrón no se produce al final del turno, hay una tendencia significativa a que se retrase la respuesta o no se responda. Los receptores miran hacia los hablantes para indicar su atención y los hablantes hacia los oyentes para señalarlos como destinatarios (*Face Address System*, Kendon 1990) y juzgar si su turno ha sido recibido (Bavelas, Coates y Johnson 2002, Goodwin 1981, Kendon 1990, Kidwell 1997, Lerner 2003). En general, las miradas confluyen entre componentes del turno (cláusulas) y al final del turno; en cambio, durante el turno los hablantes retiran la mirada del destinatario y adoptan una de media distancia, así como el oyente (Goodwin 1981:86). El hecho de que la mirada elicite respuesta es la base de los estudios de la mirada como acción⁴³ (Goodwin 1980, 1981, Heath 1986, Bavelas, Coates y Johnson 2002, Kidwell 2006, 2013)⁴⁴.

⁴⁰ Siempre se debe tener en cuenta que, en el estudio de la multimodalidad, se suele partir de la cultura occidental, aunque también puede haber variedad dentro de esta misma. Los usos de la mirada tampoco son universales; por ejemplo, diferencias de la mirada en el rol del receptor entre personas de Italia, Papua Nueva Guinea y México (Rossano, Brown y Levinson 2009), o los asiáticos orientales dedican mucho más tiempo a la mirada mutua que los caucásicos occidentales (Haensel et al. 2022).

⁴¹ Esto se ha confirmado en experimentos, el oyente mira 1.6 veces más que el hablante (Vertegaal et al. 2001).

⁴² Duncan y Fiske (1977:216) denominan *speaker-state signal* a las señales del oyente para tomar el turno: desviar la mirada e iniciar la gesticulación.

⁴³ El pico en los estudios de la mirada estuvo entre 1970 y 1980, pero hay trabajos recientes sobre ella en la interacción multimodal (véase una recopilación en Streeck, Goodwin y LeBaron 2011).

⁴⁴ Como confirmación de la mirada como recurso elicitor, es interesante el análisis comparativo de la mirada del español de Colombia con el de la zona de Cochabamba (Satti y Soto Rodríguez 2021), en el que se comprueba que en esta zona los interlocutores son menos propensos a la mirada mutua y, en cambio, utilizan marcas verbales (preguntas confirmatorias y ciertas partículas del quechua) en mayor cantidad como sustitutos de la mirada y otros recursos visuales apelativos.

Estos patrones de la mirada se han confirmado y expandido en estudios de *eye-tracking* (Oertel *et al.* 2012, Jokinen *et al.* 2013, Brône *et al.* 2017, Zima 2017, entre otros). Efectivamente, en una toma de turno limpia, sin solapamientos, hay mirada mutua durante la pausa o mientras el primer hablante termina y el segundo empieza (Jokinen *et al.* 2013). La mirada mutua suele situarse un segundo después del inicio del turno del nuevo hablante y ahí se inicia el desvío (Oertel *et al.* 2012). La negociación sobre quién continúa ocurre mientras el hablante previo aún mantiene el turno. Cuando hay cambio de hablante, el previo mantiene la mirada al compañero durante el silencio; sin embargo, si no hay cambio, normalmente el hablante ya había empezado a desviar la mirada 2 segundos antes del inicio del silencio, pues ya había detectado que el compañero no parecía dispuesto a convertirse en el nuevo hablante (Oertel *et al.* 2012). En un turno que el hablante quiere mantener a pesar de la falta de fluidez, se utilizan recursos como las pausas llenas y el desvío de la mirada al oyente para que no sea un LTP (Jokinen *et al.* 2013). En cambios de turno durante un solapamiento, hay un incremento de la mirada del hablante previo orientadas al compañero hasta que se resuelve el solapamiento (Oertel *et al.* 2012). La mirada mutua la rompe quien toma el turno, aunque el primer hablante siga hablando, para no dar opciones a que lo mantenga (Jokinen *et al.* 2013).

Si el hablante no obtiene la mirada de su destinatario durante el curso de un turno de habla, como es prototípico, utiliza procedimientos que permite el sistema, como los reinicios, las pausas y las dubitaciones, para conseguir la mirada de oyente (Goodwin 1980, 1981). Por lo tanto, estos fenómenos no vienen solo de las dificultades del hablante en su proceso interno, sino que también funcionan interactivamente, atrayendo al oyente con un rediseño del enunciado y evidencian la necesidad que tiene el hablante de recibir la mirada del oyente.

En conversaciones de varios participantes, se encuentran usos especiales de la mirada a un oyente secundario: bien del narrador que pide verificación a otro participante que tiene conocimiento de los hechos a través de la mirada (Goodwin 1981; Mandelbaum 1987; Lerner 1992); bien de un oyente que mira a una persona distinta al hablante actual para que tome el turno (Heath 1986, Kidwell 1997).

Destaca el nuevo marco analítico que aporta Rossano (2012), en el que la unidad con la que la mirada opera no es el turno, sino la secuencia. Refuta la idea de consenso según la cual el orden de la mirada se debe al rol de cada participante. Las investigaciones previas no diferenciaban los tipos de acción del hablante (narrar una historia, preguntar,

afirmar...); al tener esto en cuenta se observa que no todos los turnos del hablante ni todos los turnos iniciales de secuencia están acompañados por la mirada del hablante ni requieren la mirada del receptor. Tampoco es cierto que los receptores miran más a los hablantes que los hablantes a los receptores. Algunas actividades requieren una mirada más sostenida del hablante o del destinatario que otras, por lo que la libertad relativa de los participantes depende de las expectativas de la mirada asociadas al desarrollo de la acción en curso. Por ejemplo, durante narraciones extensas (el tipo de actividad analizan Kendon y Bavelas), los receptores miran al hablante; en cambio, durante las secuencias pregunta-respuesta, los hablantes miran al destinatario mientras realizan la pregunta e incluso el inicio de la respuesta, pero los oyentes no miran mientras reciben la pregunta. Además, Rossano observa que, cuando la secuencia de acción (como un par adyacente pregunta-respuesta, narración-apreciación, petición-cumplimiento) ha terminado exitosamente y se considera completa, los participantes apartan la mirada el uno del otro; mientras que, si uno o ambos mantienen la mirada en el otro, la secuencia necesita una continuación o expansión.

Streeck (2014) también analiza la mirada en las secuencias de interacción a la luz de la teoría de Rossano. Sugiere que el objetivo de la mirada no es verificar la atención, sino el reconocimiento. Al final de una TCU, la mirada recíproca significa que el receptor declina una oportunidad de registrar cualquier postura negativa en la acción (cara de incredulidad, desinterés) (Streeck 2014: 52) y reconoce el acto que realiza el hablante y el impacto de lo que dice. Streeck entiende la mirada mutua como una interacción "contractual" mínima (y a menudo suficiente) por la que un acto individual se ratifica explícitamente como un *fait accompli* o hecho social. Por esto, los hablantes buscan la mirada del receptor durante los inicios de las secuencias, para que reconozca la nueva acción en curso. En cambio, que el receptor decida retirar la mirada del hablante cuando responde o dejando la pregunta sin responder, en vez de mirar recíprocamente, puede ser una forma de registrar que el acto inicial no genera un curso de acción preferido, al que solo se une de mala gana, condicionalmente o no se une. Mirar a otro lado se hace en este contexto como alternativa a la mirada recíproca.

Postura

Respecto a la postura, Hagemann (2014: 1312) se plantea qué funciones comunicativas puede cubrir, si siempre es una acción complementaria al habla o si puede

funcionar como sustituta de actividades verbales como otros gestos. A partir de la definición de reguladores de Ekman y Friesen (1969), Hagemann asume que la distancia y la orientación axial, como posturas corporales dinámicas, se entienden como reguladores.

Se aprecian cambios de postura asociados tanto al turno, como a la secuencia y a la interacción completa. En primer lugar, el cambio de orientación de la cabeza es un elemento clave en la regulación del turno asociado a la aceptación o la cesión del turno (Duncan 1972, Hadar et al. 1984); así como el torso cambia la orientación axial de posición de descanso hacia atrás a posición hacia delante y viceversa (Li 2014: 177). Las culturas mediterráneas son “culturas de contacto” en las que esa orientación de la postura y de la mirada puede llegar al contacto físico, apoyando la mano en el brazo o el hombro del interlocutor (Hall 1966).

En un nivel jerárquico superior, los cambios en la postura revelan cambios de secuencia, de tema o de actividad, se realizan de forma conjunta y se reorganizan colaborativamente (Streeck 2009a: 22). Streeck denomina *pre-enactments* a los cambios de postura que anticipan cambios en la conversación. El hablante mantiene la postura hasta que alcanza el objetivo de su proyecto interaccional, tal como reporta Li (2014: 183-190) en secuencias de pregunta-respuesta y en secuencias de argumentación. En el primer caso, el hablante mantiene la postura hasta que se le da la respuesta, de modo que orienta la continuación de la actividad y la expectación de la respuesta. En el segundo caso, el cambio de postura se relaciona con la secuencia argumento-respuesta-confirmación; la progresión hacia la respuesta preferida (acuerdo) es un logro interactivo, por lo que, una vez dado, el hablante relaja el cuerpo. La postura también acompaña a turnos extensos que cuentan una historia, aconsejan o dan explicaciones (Li 2014: 177). El cambio a una postura inestable o “torcida” es interpretada por los participantes como causa para limitar la expansión de la secuencia de habla, mientras que el alineamiento de la parte baja del cuerpo con la cabeza puede ser señal para su expansión (Schegloff 1998).

Por último, las orientaciones corporales mutuas son indicadoras de una unidad interaccional espacial (Kendon 1990: 12, Özyürek 2000, 2002), y el cambio de estas orientaciones indican el inicio y el final de la interacción completa (Kidwell 2013:105). Hay pocos estudios en profundidad sobre la orientación axial y la actividad verbal en términos de organización del discurso o control temático, pues se suele estudiar desde la sociología, la antropología, la etnografía y la psicología, que se fijan más en las

actividades no verbales que en su relación con aspectos verbales de la comunicación (Hagemann 2014: 1311).

3.4. Parámetros para la clasificación de gestos

Aunque la postura y la mirada son movimientos corporales que por su intencionalidad comunicativa también identificamos como gestos, durante este apartado con el término *gesto* nos referimos a los demás elementos kinésicos, ya sean estos realizados con las manos, la cabeza, la cara u otras partes del cuerpo.

Estos gestos tienen ciertas particularidades semióticas, temporales y funcionales distintas a la postura y la mirada y que permiten distintos sistemas de clasificación, que a menudo se confunden. Por ejemplo, a menudo se han confundido la convencionalización y la función, asignándoles en exclusiva unas funciones a gestos más o menos convencionales. De la misma forma, se ha vinculado el gesto no coverbal a la convencionalización. Cada categoría de la clasificación de Ekman y Friesen (1969) sigue un criterio diferente (Kendon 2004) y la tipología de McNeill (1992) de gestos icónicos, metafóricos, deícticos, emblemáticos y rítmicos persiste a pesar del malestar generalizado (Cooperrider 2017; cf. Kendon 2004, McNeill 2005, Enfield 2009). Además, la mayoría de las clasificaciones se centra en los gestos manuales, mientras que otros movimientos corporales se ajustan a veces a estas clasificaciones o se estudian separadamente.

Por tanto, en vez de explicar las tipologías de los principales investigadores por separado, vamos a centrarnos en los parámetros por los que se describen los gestos: el grado convencionalización (§3.4.1), la dependencia del habla (§3.4.2) y las funciones (§3.4.3).

3.4.1. Grado de convencionalización

Aunque el parámetro de la convencionalización del gesto haya servido para establecer categorías (emblemas) que también se asocian a otros parámetros (coverbalidad, intencionalidad, etc.), en este apartado se resume solamente lo que se ha dicho sobre su carácter semiótico, es decir, la convención establecida entre la forma de un gesto y su significado. Por un lado, hay gestos cuya forma no está fijada a un significado por convención lingüística y, por otro lado, hay gestos que sí tienen esta fijación.

Clasificación de gestos por su convencionalización	Gestos no convencionales o espontáneos
	Gestos convencionales o emblemas
	Gestos recurrentes

Tabla 7. Clasificación de gestos por su convencionalización.

3.4.1.1. Gestos espontáneos y emblemas

Los gestos espontáneos no convencionales son más dependientes del contexto secuencial o verbal y de la iconicidad de la forma para comprender cómo contribuyen al significado del enunciado verbal. No es una relación arbitraria entre forma y significado creada por convención social, sino motivada por una relación visual con el significado que transmiten, que, igualmente, depende también de la información verbal a la que acompaña. Por ello, se suele asumir que son necesariamente coverbales, mientras que los emblemas pueden aparecer de forma independiente al habla y entenderse como sustitutos traducibles de palabras (Ekman y Friesen 1969); sin embargo, en §3.4.2.4. se verá que no siempre es así, entonces, la dependencia del habla no debe ser un criterio excluyente en la convencionalización de los gestos.

Estos gestos no convencionales se han denominado *objetivos* (Efron 1970 [1941]), *ilustradores* (Ekman y Friesen 1969), *gesticulación* (McNeill 1992) o *referenciales* (Kendon 2004), y se han subclasificado por la manera en la que establecen su significado: describiendo visualmente la forma del referente (*iconografía*, Efron 1970 [1941]; *icono*, McNeill 1992; *pictografía*, Ekman y Friesen 1969; *depiction*, Kendon 2004), describiendo la acción físicamente o mediante un patrón de características comunes a la acción (*kinetografía*, Efron 1970 [1941], Ekman y Friesen 1969; *enactment*, Kendon 2004), sirviendo las manos como modelo del objeto (*modelling*, Kendon 2004) o de manera deíctica, es decir, señalando al objeto presente prototípicamente con un dedo, pero también con un objeto u otras partes del cuerpo (Efron 1970 [1941], Ekman y Friesen 1969, McNeill 1992, Kendon 2004). Enfield (2009:12) considera que el señalamiento, el *signo simbólico indexical*, se encuentra entre los gestos no convencionales y convencionales, ya que tiene propiedades de ambos: su forma es convencional, pero el referente depende del contexto. McNeill (1992) añade a los gestos no convencionales los gestos metafóricos, que son similares a los icónicos en que presentan una imagen, pero de un concepto abstracto, y especifica que los gestos deícticos no tienen por qué apuntar a entidades concretas sino también al espacio al que se le ha dado un valor referencial.

En cambio, los denominados *emblemas*⁴⁵ (Ekman y Friesen 1969) —*semiotic gestures* (Barakat 1969), *formal pantomimic gestures* (Wiener et al. 1972), *symbolic gestures* (Calbris 1990, Efron (1970 [1941]), Poggi 2002, Wundt 1973 [1900]), *autonomous gestures* (Kendon 1983, Payrató 1993), *quotable gesture* (Kendon 1992, Brookes 2001)—, están codificados léxicamente y no dependen del habla para su comprensión, ya que la forma es la que permite identificar su significado. Para que esa identificación funcione, el emblema tiene que cumplir un estándar de buena formación, es decir, pequeñas variaciones en la configuración formal del gesto hacen que ya no se interprete ese significado; de manera que se interpreta igual en variedad de contextos (Hanna 1996). Aun así, Payrató (2003) cuestiona la articulación de emblemas en componentes mínimos: la variación morfológica formal de muchos emblemas⁴⁶, su sinonimia y polisemia (al menos parcial) y la modalización que pueden recibir (especialmente de la expresión facial) y aconseja tratar estas unidades no como categorías cerradas, sino en términos de redes de semejanza familiar.

Generalmente, se asocian los gestos emblemáticos o convencionales con formas arbitrarias y los no convencionales con formas icónicas, y esta era la diferencia fundamental para Efron (1970 [1941]); sin embargo, el criterio principal para los emblemas es el seguido por Ekman y Friesen (1969): la convención del gesto con un significado léxico compartido. Es decir, aunque la forma arbitraria o icónica del gesto pueda ser un indicio, no es determinante, ya que un emblema puede tener rasgos de iconicidad como vestigio de su pasado⁴⁷. El emblema sigue un proceso de lexicalización, nace como gesto icónico espontáneo que se ritualiza y se establece como parte de un código, pudiendo mantener esa iconicidad o tan solo un esquema de su forma que se percibe como arbitraria (Kendon 1981, 1988, 2013).

La convención de los emblemas, al igual que la de los signos lingüísticos, depende de cada comunidad de habla, por lo que estos gestos normalmente no se entienden fuera de la cultura en la que se insertan⁴⁸, especialmente cuando no son icónicos. La

⁴⁵ Poyatos (1994: 187), como Scherer (1977), no considera que los emblemas solo puedan ser kinésicos, sino también paralingüísticos (ej. un bufido de impaciencia) o incluso proxémicos (ej. acercarse bruscamente para indicar ‘estamos juntos’).

⁴⁶ Para las variantes morfológicas de un mismo emblema (Payrató 2003: 76), Belío-Apaolaza (2019) sugiere la denominación *alomorfos emblemáticos*.

⁴⁷ Estos signos son más longevos que los formados por palabras, pues, al no tener composicionalidad, no cambian fácilmente (Kendon 1981).

⁴⁸ Morris et al. (1979) encuentran una relación inversa entre la distribución geográfica de un emblema y la cantidad de significados, es decir: cuanto más expandido está el emblema, tiende a un único significado

investigación de gestos con este enfoque etnográfico tiene como resultado los diccionarios o repertorios de gestos de una cultura⁴⁹, que suelen incluir exclusivamente emblemas.

Los emblemas son los gestos más cercanos a la lengua hablada, ya que esta convención los convierte en símbolos lingüísticos. Ekman y Friesen (1969), además del carácter convencional y cultural del emblema, destacan esta posición cercana a la lengua a partir de características como su realización con consciencia e intencionalidad, la traducción en palabras o frases y la sustitución del habla. No obstante, en §3.4.2.4. vemos que estas no son características exclusivas del emblema.

Respecto a la independencia verbal, Poggi (2002) diferencia entre *emblemas holofrásticos*, que equivalen a un acto de habla completo, y *emblemas léxicos/articulados*, que se comportan como componentes de un acto que también incluye elementos verbales o de otra modalidad (como una expresión paralingüística o un gesto facial que no forma parte del emblema). El hecho de que los emblemas puedan funcionar de esta manera holofrástica o independiente lleva a su caracterización desde una perspectiva pragmática como acto de habla con su propia fuerza ilocutiva (Kendon 1995, Payrató 1993, 2003, Poggi 2002) (§3.4.3.3).

3.4.1.2. Gradación de la convencionalización: gestos recurrentes

A pesar de que existe la distinción entre gestos convencionales y no convencionales, esta no siempre es clara y se asume que existe una gradación, no una dicotomía⁵⁰. Kendon (1988) observa que las propiedades lingüísticas de los distintos tipos de gestos cambian según las demandas comunicativas: los gestos en asociación con el habla representan aspectos de significado de una manera visual, pero en alternancia con el habla o en ausencia de esta, asumen todas las funciones de la lengua hablada, equivalen a una palabra o frase; especialmente cuando la forma es estable, llegando a tener propiedades estructurales análogas.

común; cuanto menos expandido, más significados; además, los gestos atados a una expresión idiomática o interjección no suelen cruzar fronteras lingüísticas.

⁴⁹ Algunos ejemplos en español: Cestero (1999), Cestero et al. (2020), Coll, Gelabert, y Martinell (1990), Green (1968), Poyatos (1970), Meo-Zilio y Mejía (1980); y en catalán: Payrató (1993).

⁵⁰ Por eso en su clasificación Payrató (1993) crea la categoría de *pseudo-emblemas* para aquellos gestos con menos convencionalidad y precisión de significado.

A partir de esta idea, McNeill (1992: 37) sintetiza lo que llama *Continuum de Kendon*, cuyo orden es: *gesticulación, gestos como lengua, pantomima, emblemas y lenguas de signos*. Según nos movemos de izquierda a derecha, la presencia obligatoria de habla decae, la presencia de propiedades de la lengua se incrementa y los gestos idiosincrásicos se remplazan por signos socialmente regulados. Sin embargo, Kendon (2004: 104) no está de acuerdo completamente con esta adaptación de su explicación a un *continuum* porque no siempre la coocurrencia con el habla o la iconicidad implica que el gesto no esté convencionalizado⁵¹. Kendon aboga por el uso de clasificaciones multidimensionales, es decir, en las que las categorías de gestos cumplan determinadas características que no son interdependientes: forma, iconicidad/arbitrariedad, in/dependencia del habla, convencionalización, etc. De ahí que McNeill (2005) divida el *continuum* en cuatro parámetros: la relación con el habla (obligatoria en la gesticulación, opcional en los emblemas), las propiedades lingüísticas (ausentes en la gesticulación, algunas en los emblemas), convención (ninguna en la gesticulación, parcial en los emblemas) y semiosis (global en la gesticulación, segmentada en emblemas). Quizás una descripción más completa es la de Iriskhanova y Cienki (2018: 30), con un *continuum* multivectorial de 12 parámetros: convencionalidad, semanticidad, arbitrariedad, transparencia pragmática, autonomía, *import* social y cultural (simbolismo), consciencia, recurrencia, iconicidad, metaforicidad, indexicalidad, saliencia. Iriskhanova y Cienki critican del *Continuum de Kendon* la tendencia lingüística de comparar los movimientos corporales con las palabras y, por tanto, de solo tomar en consideración el criterio de la convencionalidad, en vez de las características semióticas o no semióticas del gesto. El *continuum* semiótico de los gestos no debe establecerse en función de mayor o menor parecido con el signo lingüístico, ya que los signos pueden ser signos en una variedad de maneras.

Realmente, no todos los elementos del *Continuum de Kendon* (McNeill 1992) se van a encontrar en la conversación oral: ni gestos pantomímicos, pues forman parte de una actividad artística particular, ni signos de las lenguas de signos, ya que no son elementos de las lenguas orales. Müller (2010b, 2017) y sus discípulas proponen otro

⁵¹ Además, Müller (2018) añade que la adaptación no es correcta porque Kendon (1988) incluye los gestos *depictive* y pantomímicos dentro de la gesticulación y afirma que ambos pueden funcionar como palabras en una oración verbal.

*continuum*⁵² con los gestos que suelen aparecer en la comunicación en lengua oral utilizando como parámetros la forma, el grado de convencionalización y las funciones lingüísticas y comunicativas⁵³. Estos gestos son los *gestos singulares*, los *gestos recurrentes* y los *gestos emblemáticos*.

Los gestos singulares son creados al momento, son libres y espontáneos, aunque están basados en un repertorio de técnicas de creación compartido culturalmente. Estas gesticulaciones tienen significados fáciles de interpretar en su contexto sin haberlos conocido anteriormente, pero algunos por su recurrencia están a medio camino en ese proceso de convencionalización hacia el emblema (Ladewig 2014a). Los gestos recurrentes combinan el carácter idiosincrásico del gesto singular con el convencional de los emblemas. Su significado “is schematic rather than word-like” (Ladewig 2014a: 1560) y todavía es transparente. Algunos ejemplos son cepillar el aire con el brazo hacia un lado para rechazar una idea o un gesto cíclico para formular. Su forma y su núcleo semántico son estables en diferentes contextos por diferentes hablantes de una comunidad de habla, o sea, más estables que en los gestos icónicos o metafóricos que se dan en un momento dado. Sin embargo, para que el significado de estos gestos recurrentes se distinga, basta con utilizar uno o dos parámetros de forma, pues se agrupan en familias de gestos⁵⁴ que comparten un tema semántico común (véase Müller 2004 y Kendon 2004). En consecuencia, su articulación no es tan restrictiva como la de los emblemas. Dependiendo del contexto de uso, muestran diferencias de forma que se correlacionan con variantes de significado y función (Ladewig 2014a, 2014b, Müller 2004, Teßendorf 2014).

La mayoría de los recurrentes son gestos coverbales que se clasifican funcionalmente como pragmáticos en una cultura, que evolucionan desde un significado icónico o metafórico. Cienki (2013) sugiere un fenómeno paralelo al que Traugott (1988) describió para el cambio semántico diacrónico en el lenguaje verbal: una expresión con referencia física o metafórica se usa en relación con el contenido mismo del discurso en

⁵² Por no extender más este apartado, no explicamos el *continuum* de Rossini (2012: 46-7) a partir de la tipología de gestos de McNeill (emblemas arbitrarios → emblemas icónicos → iconos (beats) → metáforas), descritos mediante cinco parámetros para medir el nivel de lexicalización: intencionalidad, conciencia, abstracción, arbitrariedad y extensión.

⁵³ El estudio comparativo de Kendon (1995) entre emblemas y gestos coverbales aparentemente convencionales, con funciones pragmáticas principalmente, es el que inicia la investigación de los gestos recurrentes (Teßendorf 2013).

⁵⁴ Frente al concepto de *familia de gestos*, Bressemer y Müller (2014a: 1588) proponen el término onomasiológico *gestural field*: una compleja red de relaciones entre gestos del repertorio que comparten un tema semántico motivado por aspectos de forma compartidos.

un proceso de vaciamiento semántico y reforzamiento pragmático. De forma prototípica, se puede describir la función de los recurrentes como pragmática, la de los singulares como semántica y la de los emblemáticos como actos de habla completos⁵⁵ (Müller 2018). No obstante, la vinculación de los gestos recurrentes con el habla es variada (Ladewig 2014a: 1562-3), desde una función referencial que complementa el contenido proposicional del enunciado, pasando por una función discursiva o modal, hasta una función performativa que regula el comportamiento de otros o realiza la fuerza ilocutiva de un acto de habla. Con esta última función, pueden aparecer sin vinculación con el habla; la función discursiva, en cambio, puede ser independiente en menos ocasiones (un caso sería el gesto cíclico con función formulativa). Centrándose en la función referencial, Streeck (2021) describe el proceso de emancipación por el que gestos referenciales espontáneos se vuelven recurrentes y denomina a este tipo *gesto conceptual*.

3.4.2. Dependencia temporal con el habla

El siguiente parámetro para clasificar los gestos es su dependencia temporal con el habla. En §3.4.1. algunos investigadores asocian esta propiedad al carácter convencional del gesto, pero en este apartado vamos a explicarla por sí misma.

El grado de dependencia del habla supone tres tipos de enunciado (Kendon 2004: 110): acciones compuestas por habla, por acciones visibles o por una combinación de ambas. En el último caso, el enunciado es una unidad compleja con una estrategia retórica única en la que se coordinan temporalmente ambas modalidades y sus significados.

3.4.2.1. Temporalidad del gesto

La lengua es temporalmente lineal y se estructura de forma jerárquica segmentando el significado global en palabras, mientras que los gestos son simultáneos y sintéticos, por lo que presentan el significado complejo sin segmentación, “each symbol is a complete expression of meaning unto itself” (McNeill 1992:21). Entonces, no se puede segmentar el significado del gesto de forma jerárquica, pero la realización de los movimientos sí puede dividirse temporalmente en *gesture phrases* y *gesture units* (Kendon 2004: 108-112).

⁵⁵ Véase Fricke (2013) para las funciones de los gestos singulares y Teßendorf (2013) para las de los emblemáticos.

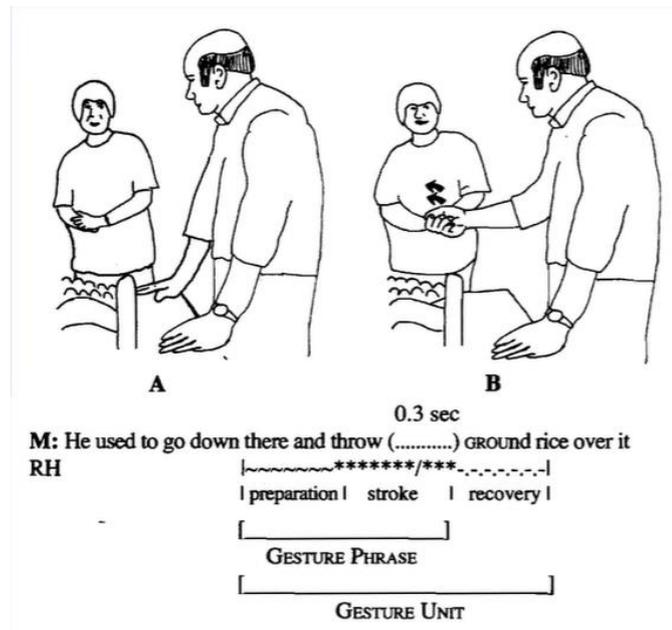


Fig. 5. Ejemplo de unidad gesto y su estructura interna en coordinación con un enunciado verbal (Kendon 2004: 114)

La *frase gesto* es la unidad de acción corporal visible identificada mediante características kinésicas que corresponden a unidades de acción significativas como apuntar, dibujar, actuar o hacer pantomima. Sin embargo, en su realización se producen además movimientos colaterales que junto a la frase forman la unidad gesto. Esta unidad incluye tres unidades mínimas o fases⁵⁶: 1) todo el movimiento desde la posición corporal de reposo hasta el gesto propiamente dicho, llamado *preparación*; 2) la expresión del gesto, cuya forma y significado se manifiesta con claridad, llamada *fase álgida*⁵⁷ (*stroke*); y 3) el regreso a la posición de reposo, es decir, *recuperación*. La fase álgida es el núcleo del gesto y se puede expandir al añadir entre la fase álgida y la recuperación un *post-stroke hold* (Kita 1993), que sostiene la articulación de la fase álgida, a menudo para cubrir el segmento de habla. La preparación y la fase álgida forman la frase gesto; la frase gesto junto a la recuperación forman la unidad gesto, que puede contener una o más frases gesto.

⁵⁶ A estas fases Poyatos (2018) las denomina: *formativa, central, desarticuladora*. Véase Ladewig y Bresem (2013) para una descripción detallada de la forma y de la función de cada fase.

⁵⁷ No existen muchos trabajos en español que traten estas unidades, pero tomamos la traducción de *stroke* como *fase álgida* de Esteve-Gilbert (2016).

3.4.2.2. Sincronía de gesto y habla

La correlación entre gesto y habla ha sido ampliamente documentada en los estudios del gesto (McNeill 1985, Clark 1996, Bavelas y Chovil 2000, Kendon 2004). Los enunciados compuestos se interpretan mediante el reconocimiento y la unión de múltiples signos bajo un principio de correlevancia que los une pragmáticamente (Enfield 2009). Kendon observa que “the pattern of movement that co-occurs with the speech has a hierarchic organization which appears to match that of the speech units” (Kendon 1972: 190). La sincronización es una herramienta metacomunicativa para señalar lo que está en la misma unidad integrada de significado (Engle 2000). Bavelas, Holt y Allison (2000, *apud.* Bavelas y Chovil 2016) muestran que normalmente no se utilizan marcadores lingüísticos para mostrar la relación, simplemente la sincronía (70% de los gestos datados).

Generalmente, la fase álgida coincide con palabras semánticamente clave, por lo que la preparación del gesto comienza antes del habla. Esto ilustra que el hablante ya ha planificado el habla para coordinar el gesto (Kendon 1972, 1980, 2004, McNeill 1985, 1992, 2000). La imagen surge primero (fase de preparación) y se transforma en una estructura compleja en la que se integran gesto (fase álgida) y estructura lingüística (sílabas prosódicamente prominente) (McNeill 1992).

Según el concepto de *co-expresividad* de McNeill (1992, 2005), el uso del gesto y el habla en un momento dado expresa conjuntamente una misma idea, aunque el contenido de cada uno no sea exactamente el mismo, sino complementario. McNeill (1992: 27-29) considera que no hay ejemplos, en principio, de que con el habla se presente una idea y con el gesto otra al mismo tiempo, sino que siempre hay una sincronía semántica o pragmática. Sin embargo, es perfectamente posible la situación que ejemplifica Poyatos (1994: 135) en la que una persona se dirige a dos receptores a la vez, a uno verbal y paralingüísticamente por teléfono y a otro kinésicamente en persona. En tal caso, sin que los respectivos oyentes se reconozcan entre sí como coparticipantes, estaríamos hablando de que el hablante está en dos conversaciones simultáneamente.

En el caso de la sincronía semántica, los dos modos, el habla y el gesto, presentan significado al mismo tiempo para formar una *idea unit* (McNeill 1992: 27) (véase la función referencial del gesto en §3.4.3.2). Aun así, no siempre coinciden temporalmente los modos debido a tres causas (McNeill 1992: 27-29): las pausas, los gestos múltiples y los gestos que corresponden a más de una cláusula. En la primera, la fase álgida continúa

durante la pausa verbal, mostrando que la estructura semántica del habla interrumpida permanece intacta. En la segunda, para una cláusula se utilizan varios gestos, cada uno cubre una unidad de idea desde un ángulo diferente. En la tercera, ocurre lo contrario, la fase álgida, a pesar de haber terminado, presenta información que se desarrolla durante más de una cláusula. En situaciones en las que la desincronía dificulta la identificación de la vinculación entre las modalidades, son las actividades de la interacción las que permiten a los hablantes relacionarlas entre sí (Sidnell 2006).

Un estudio reciente de Urbanik y Svennevig (2021) señala que las representaciones genéricas de una acción se inician en su mayoría antes de la verbalización de la acción o se sincronizan con ella, porque facilitan el reconocimiento de la acción verbalizada; mientras que las representaciones contextualizadas, que incluyen referencias deícticas al entorno espacio-material o referencias icónicas del modo específico de realización de la acción, se inician en su mayoría simultáneamente con la verbalización y se extienden más allá de la frase verbal o del turno, para añadir información específica que complementa el significado de la frase verbal, asegurando la comprensión de los aspectos específicos de la acción.

Si la sincronía no es semántica, el gesto y el habla realizan juntos una función pragmática (McNeill 1992) (véase §3.4.3.3 sobre gestos pragmáticos). Además, Streeck (1993, 2009b) observa que los gestos pragmáticos que se inician ligeramente antes del habla proyectan el acto comunicativo de los enunciados antes de verbalizarlos, por ejemplo: *stance markers* que muestran una actitud hacia lo que van a decir, como falta de compromiso encogiendo los hombros; un gesto de señalar con el dedo para autoseleccionarse como hablante (Mondada 2007); o una sonrisa que advierte una evaluación, así como se mantiene después de esta (Ruusuvuori y Peräkylä 2009).

3.4.2.3. Sincronía de gesto y turno

Diversos estudios han demostrado que la sincronía, o mejor dicho, la asincronía, del gesto con el habla resulta ser especialmente explicativa de la estructura y la gestión del turno (Müller 1998, Streeck 2009b, Bohle 2007, Li 2014, Mondada 2015). Son prácticas coordinadas sistemáticas, en las que cada modalidad tiene su lugar y adapta su temporalidad mediante alargamientos o repeticiones de un gesto, de una fase del gesto, de una sílaba o de una palabra (Kendon 2004: 127).

El gesto no solo anticipa la *idea unit*, sino también el turno verbal cuando el gesto o su fase preparatoria se produce antes de este. Schegloff (1984) es el primero en identificar esta posición previa del gesto como un método deliberado de proyección, también denominada *forward-gesturing* (Streeck 2009b), que tiene como objetivo facilitar la coordinación interpersonal en la gestión del turno.

También el final del turno es un momento de gran complejidad en el que se produce la transición al turno de otro hablante. Los participantes muestran su coordinación señalando e interpretando el final del turno utilizando varias modalidades: sintaxis y entonación (Wells y Macfarlane 1998, Couper-Kuhlen 2001, Selting 2005, Ruitter et al. 2006), gestos (Mondada 2007, Streeck y Hartge 1992), mirada (Bavelas, Coates y Johnson 2002, Rossano 2012). El fin del gesto tiende a coincidir con el fin del turno, al volver la parte del cuerpo a su posición de reposo (señal ya mencionada por Duncan y Fiske 1977), no siempre a la vez que el final sintáctico. El siguiente hablante no suele tomar el turno hasta que no termina el gesto, hasta que el turno no ha terminado por completo (Bohle 2007, 2013).

Esto lleva a que, durante pausas o ante competición por el turno, los gestos se alarguen para indicar una continuación del turno del hablante actual (Bohle 2007, 2013). No obstante, la información sobre la gestión del turno que proviene del gesto y del habla puede contradecirse: bien el gesto indica la continuación del turno al mantenerse a pesar de la conclusión sintáctica o pragmática, bien el gesto termina mientras que el final no es explícito por la entonación y la sintaxis (Li 2014: 172). Nótese también que los gestos relacionados con turnos de varios TCU tienen estructuras más complejas (Li 2014: 146-153): la unidad gesto puede coexistir con una sola o con varias TCU del turno y cada TCU puede coincidir con una o más frases gesto. Además, el gesto también puede proyectar el inicio de una TCU que no sea la primera, es decir, se sitúa entre dos TCU (Streeck 2009b: 171). Por tanto, desde un análisis multimodal, conceptos clave del AC se ven limitados y se redefinen (Bohle 2007, 2013): las frases gestuales deben formar parte integral del TCU y cada modo (sintaxis, prosodia, gesto) puede indicar un *completion point*, pero este puede ser suspendido como LTP si otro modo señala la continuación.

Por último, Goodwin (1979, 1980, 1981) demostró que en la construcción del turno participan actividades multimodales de otros participantes con las que el turno se sincroniza. Concretamente, el hablante produce reinicios o pausas para atraer la mirada del receptor, de manera que no continúa el turno hasta que no se sabe escuchado (§3.3.2).

En vista de lo argumentado en este apartado, los estudios de la interacción deben ser necesariamente multimodales y, además, no solo se deben estudiar los gestos en relación con su afiliación verbal, sino con su posición en el turno (Streeck 2009b: 177).

3.4.2.4. Gestos no coverbales

Una cantidad menor de gestos se producen sin que el hablante produzca ninguna emisión verbal al mismo tiempo. Podemos diferenciar dos situaciones diferentes en las que esto sucede: a) el gesto sustituye parte del habla en un acto verbal, b) el gesto constituye un acto gestual completo.

En el primer caso, los gestos se integran linealmente en el enunciado hablado, ocupando huecos sintácticos, preferiblemente el de un nombre o un verbo, a mitad del enunciado o al final del enunciado sin terminar (Ladewig 2014a). No es una información suplementaria de una unidad verbal, sino que la reemplaza; por consiguiente, no es lo mismo que un gesto que acompaña a una expresión anafórica como “así” o “de esta forma”⁵⁸, con el que mantiene una relación atributiva y restrictiva (Ladewig 2014a). No hay gestos que sean inherentemente sustantivos o verbos, es la posición de sustantivo o verbo la que hace resaltar la información semántica de objeto o de acción de ese gesto, que reside en mayor medida en la forma y en el movimiento de la mano respectivamente (Ladewig 2014a). A esta relación del gesto con las funciones sintácticas Fricke (2012) la denomina *gramática multimodal*—Slama-Zazacu (1976), *mixed syntax*; Poggi (Poggi y Magno Caldognetto 1996), *gestos léxicos*; McNeill (2005), *speech-linked gestures*; Beck Nielsen (2020), *silence gestures*—. Además, se pueden yuxtaponer gestos que casi forman una construcción predicativa de una lengua de signos (LLAMAR + DESPUÉS) (Payrató 2014). Poyatos (1994) explica que este tipo de sustitución se produce por deficiencia verbal, o sea, por la dificultad del hablante para recuperar la palabra del léxico, pero habría otras razones (palabras tabú, rapidez, expresividad...).

Aunque generalmente se suelen asociar los emblemas a los gestos independientes del habla y los gestos ilustradores o icónicos a los que acompañan al habla (§3.4.1), el análisis de Ladewig (2014a) demuestra que cualquier tipo de gesto puede ocupar una función sintáctica; pero, además, la mayoría de los que se integran sintácticamente son gestos referenciales no convencionalizados (Ladewig 2011), por lo que debería

⁵⁸ Para el uso de *so* y *like* como “*body quotation*”, véase Streeck (2002b).

desecharse esa relación entre la convencionalización del gesto y la independencia del habla. También Poyatos (1981: 39-40) observa que los emblemas suelen acompañar al habla, al menos en la cultura hispánica, y Kendon apunta a observaciones similares sobre Nápoles (2008: 360).

En cambio, hay gestos que no se integran en la estructura sintáctica de un enunciado verbal, sino que son completamente independientes del habla y por sí mismos transmiten un significado semántico o pragmático de un acto de comunicación o un enunciado no verbal. Han adquirido fuerza ilocutiva (Payrató 2016, Payrató y Clemente 2020) (§3.4.3.3.3. sobre gestos performativos). Estos gestos se han denominado *segmentables* (Cestero 2016, 2017), *autónomos* o *citables* (Kendon 2004), *gestures of silence* (McNeill 2014) u *holofrásticos* (Poggi 2002).

Las causas que motivan los gestos autónomos (Kendon 2004: 343) son, a nivel práctico, que su emisión es más rápida que la del enunciado hablado y es visible a larga distancia si es necesario; a nivel discursivo, que permiten intercambios externos al diálogo principal o evitando crear uno “oficialmente”. Es decir, por un lado, en conversaciones con más de dos interlocutores, uno puede comunicar algo a otros más allá de los márgenes del intercambio oral actual sin interrumpirlo; por otro lado, se favorece el uso del gesto en ciertas conversaciones mínimas o conversaciones que no se quieren establecer como tales imponiendo la estructura del discurso dialogal, por ejemplo, al cruzarse con otra persona de camino a un lugar o al relacionarse con extraños y con personas que dan un servicio.

3.4.3. Función

Vistos los parámetros de la convencionalización y de la dependencia verbal, que no tienen relaciones unívocas entre sí, este apartado se centra en las funciones que pueden cumplir los gestos, que se han adelantado en parte. Puesto que nuestro objetivo es identificar cómo afectan los gestos a la estructura dialógica y a la validación de turnos, la perspectiva funcional es la más adecuada para comprender el papel que cumplen en la interacción.

La clasificación de Ekman y Friesen (1967) ha sido criticada por vincular cada tipo de gesto a una parte del cuerpo (Schönherr 2014: 1337): los emblemas y los ilustradores con los gestos manuales, los reguladores con las miradas y las muestras de emociones con la expresión facial; aunque haya afinidad entre ciertas áreas y ciertas

funciones. Es más útil una clasificación funcional, acorde a la función del gesto en el curso de la interacción, que una clasificación por el área del cuerpo empleada (manos, expresión facial, mirada, proxémica...), ya que por separado dicen poco sobre el significado del patrón (Schönherr 2014: 1334, Bavelas 1994: 202). Las clasificaciones funcionales permiten una visión global independiente de los enfoques basados en los canales (Schönherr 2014: 1340). Sin embargo, aunque haya funciones que sean comunes, debido a la complejidad de una categorización global, las investigaciones suelen centrarse en estudios de caso en profundidad (Deppermann 2013: 3), en familias de gestos o partes del cuerpo. Algunos de ellos los mencionaremos en el apartado de cada función.

Los principales criterios para clasificar la función de un gesto son su forma y el contexto secuencial. A los emblemas se les asocia un significado sin necesidad de recurrir al contexto, mientras que otros gestos espontáneos dependen de la iconicidad de la forma y del contexto para interpretarlos, y también hay gestos recurrentes cuyas formas convencionales y variantes no están tan definidas (§3.4.1). La identificación del significado del acto gestual depende en gran parte del contexto verbal y secuencial, de sus antecedentes y de sus consecuencias, que son los únicos datos analizables de la intencionalidad del hablante (Argyle 1988: 7). La metodología de *Methods for Gesture Analysis* (MGA), que tomamos para nuestro análisis (§5.5), cuenta con una identificación de la función que se divide en cuatro bloques (Bressemer y Müller 2014a: 1577-8): 1) descripción detallada de la forma⁵⁹ y su motivación metafórica; 2) estructura secuencial, es decir, en relación con el habla y otros gestos; 3) contexto de uso local, es decir, la relación con aspectos sintácticos, semánticos y pragmáticos del habla; y 4) distribución en diferentes contextos de uso para distinguir el núcleo semántico del gesto de su significado local y de sus variantes contextuales. En cambio, Bavelas (1994) se aleja de enfoques taxonómicos que vinculan la forma del gesto con su significado, dando la importancia en la identificación de las funciones en el contexto local: “In short, the goal of analysis should not be to decide in which category we should put a gesture (or all gestures) but rather to discover at least some of the things a gesture is doing at its particular moment in the conversation” (Bavelas 1994: 204).

Los gestos son multifuncionales por naturaleza, un mismo gesto puede cumplir una función distinta en cada situación o incluso contribuye a múltiples funciones

⁵⁹ Utilizan *Linguistic Annotation System for Gestures* (Bressemer et al. 2013) (§4.4.2).

semánticas y metacomunicativas simultáneamente (Ekman y Friesen 1969, Bavelas 1994, Kendon 2004, 2013, Kok et al. 2016, Cienki 2017). Por ejemplo, un gesto que sustituye una palabra puede servir para completar el enunciado verbal porque el hablante no encuentra la palabra (función referencial) y/o una forma de pedir ayuda al oyente para completar el enunciado (función interactiva). Aun así, las ambigüedades de las *marcas posturo-mimo-gestuales* se reducen cuando se tienen en cuenta el resto de los elementos con los que aparece en el contexto específico (Bouvet y Morel 2002: 130).

Ante casos multifuncionales o ambiguos, se suele señalar una función principal y otras secundarias (Müller 1998), pero el resultado es que se mantienen las segundas en la oscuridad (Kok et al. 2016). En general, el comentario crítico a la clasificación funcional es que es un método fuertemente interpretativo y a veces es difícil distinguir qué funciones se cumplen y encajarlas en una etiqueta; por ello, estas clasificaciones deben definir las categorías con precisión y ponerlas en práctica de forma comprensible para mantener la mayor intersubjetividad posible (Schönherr 2014: 1340). La falta de una descripción unificada es lo que dificulta la comparación de los resultados de la investigación. Aunque haya ciertas ideas en común, cada sistema de clasificación refleja la disciplina del investigador y se basa en criterios diferentes, por lo que un mismo gesto puede clasificarse de forma diferente en función del aspecto que se considere más relevante (Bohle 2014: 1459). McNeill (2005: 41-42) trata de evitar una categorización discreta y propone describir el gesto mediante dimensiones que se mezclan en un mismo gesto (iconocidad, metafóricidad, deixis, ritmo, interacción social y otras características) de manera no jerárquica; sin embargo, raramente se ha adoptado en estudios cuantitativos (Kok et al. 2016).

A pesar de que existen varias coincidencias entre la lingüística funcional y la descripción funcional de los gestos en Kendon (2004), Streeck (2009a), Enfield (2009) y muchos otros, las referencias explícitas o la adaptación de los modelos funcionales concretos son escasas (Kok et al. 2016).

Las funciones que se explicarán a continuación parten de la clasificación de Kendon (2004), inspiradas en la lingüística estructural, pero en cada una mencionaremos autores previos en los que se basa y posteriores que han utilizado nomenclaturas diferentes, que han añadido alguna característica particular o que se oponen de distintas maneras (Tabla 8).

Kendon (2004)	<i>Gestos representacionales</i>	<i>Gestos pragmáticos parsing</i>		<i>Gestos pragmáticos performativos</i>	<i>Gestos pragmáticos modales</i>
Müller (1998, 2013, en prep.)	<i>Depictive gestures</i>	<i>Gestos pragmáticos discursivos</i>		<i>Gestos pragmáticos performativos</i>	<i>Gestos pragmáticos modales</i>
(Streeck 2005)		<i>Speech handling gestures</i>	<i>Gestos metapragmáticos</i>		
MFD (Bavelas et al. 1992, Bavelas 1994)	<i>Gestos temáticos</i>	<i>Gestos interactivos o gestos conversacionales</i>			
Scherer (1977)	<i>Gestos parasemánticos</i>	<i>Gestos parasintácticos</i>	<i>Gestos parapragmáticos</i>		
Gilbert Austin (1966[1806])	<i>Gestos significantes</i>	<i>Gestos no-significantes</i>			
Wehling (2017)	<i>Gestos narrativo-referenciales</i>		<i>Gestos gestores del discurso</i>		
Otros		<i>Batons</i> (Ekman y Friesen 1969, Efron 1970 [1941]), <i>beats</i> (McNeill y Levi 1982, McNeill 1992) o <i>marcadiscursos</i> (Poyatos 1994a). <i>Ideografías</i> (Efron 1970 [1941]) <i>Cohesivos</i> (McNeill 1992)	<i>Reguladores</i> (Ekman y Friesen 1969) <i>Conductas conversacionales</i> (Poyatos 1996a) <i>Sincronizadores</i> (Cosnier y Vaysse 1997)		

Tabla 8. Comparativa funcional de los gestos por varios autores.

3.4.3.1. Gestos referenciales y gestos pragmáticos

En primer lugar, se diferencia entre funciones referenciales y funciones pragmáticas. Por un lado, los gestos con contenido semántico contribuyen al significado del enunciado repitiendo el significado de las palabras, restringiéndolo, añadiendo otros significados o representando un objeto⁶⁰ (Kendon 2004: 176-185). Se han denominado *gestos representacionales* (Kendon 2004), *gestos temáticos* (Bavelas et al. 1992), *gestos*

⁶⁰ Una descripción similar de la relación entre conducta no verbal-lenguaje es la de Poyatos (1994: 80): confirmar, duplicar, reforzar, debilitar, contradecir, camuflar.

parasemánticos (Scherer 1977), *gestos significantes* (Gilbert Austin 1966[1806]), *gestos narrativo-referenciales* (Wehling 2017).

Por otro lado, los gestos pragmáticos no tienen significado referencial ni contenido proposicional, sino un uso procedimental (Kendon 1995, 2004), "rather than contributing to the propositional content of the utterance, [...] embody the illocutionary force or the communicative action which often remains verbally implicit" (Müller 2008: 225). También se han denominado *gestos no-significantes* (G. Austin 1966[1806]), *gestos parapragmáticos* (Scherer 1977) *gestos interactivos* (Bavelas et al. 1992), *gestos conversacionales* (Bavelas 1994) y *speech handling gestures* (Streeck 2005). Mientras que el contenido sustantivo de cada enunciado no tiene límites (también así gestualmente), el tipo de acto que es, el tipo de estructura organizativa en la toma de turno y la forma en la que el discurso se estructura es mucho más limitada (Kendon 2004: 282). Podría decirse que son marcadores de los actos de habla (Kendon 2017), pues esta definición semántica coincide con la de los marcadores del discurso en no presentar un contenido referencial o denotativo, sino un significado de procesamiento (Martín Zorraquino y Portolés 1999).

Sin embargo, adelantamos que estas funciones semánticas y pragmáticas no son mutuamente exclusivas (Kendon 2013). Bien un gesto puede tener funciones referenciales y pragmáticas simultáneamente, bien tiene una u otra según el contexto. Por eso los "gestos pragmáticos" no son un registro cerrado, sino que son gestos que en un contexto tienen funciones pragmáticas (Kendon 2004). Uno de los ejemplos que da Cienki (2017: 64-65) es el gesto conocido por las siglas *PUOH* (*palm up open hand*) (Kendon 2004, Müller 2004): unas veces su significado es más semántico al mostrar un objeto o señalar un referente visible; otras veces es más pragmático al presentar una idea o indicar que se desconoce algo. Las categorizaciones demasiado estrictas llevan a errores, como excluir un gesto con rasgos icónicos o metafóricos de los gestos pragmáticos, o excluir una variante con función referencial de un grupo de gestos pragmáticos que comparten una forma y/o un significado similar (Ladewig 2014a: 1560-1). Aunque en este apartado se expongan las características de cada función, en la práctica cualquier gesto puede ser analizado como pragmático en algún grado (Payrató y Tebendorf 2014).

Esta distinción entre referencial y pragmático podemos relacionarla con la propuesta inicial de Austin (1962) de enunciados *constatativos*, que describen estados de la realidad y son susceptibles de ser verdaderos o falsos, y enunciados *realizativos*, que conllevan

una acción simultánea a su emisión lingüística. Sin embargo, esta teoría evolucionó en las tres perspectivas de cualquier acto: locutivo, ilocutivo y perlocutivo. Por tanto, podríamos decir que los gestos referenciales y pragmáticos son componentes del acto que ayudan a interpretar su sentido locutivo e ilocutivo/perlocutivo, respectivamente.

Asimismo, el concepto de *significado de procesamiento* o *procedimental* proviene de la teoría de la Relevancia (Sperber y Wilson 1986) y en Lingüística se suele utilizar para los conectores discursivos (Llopis Cardona 2016). Es Blakemore (1987) quien propone distinguir entre formas lingüísticas con significado procedimental y con significado conceptual. Por tanto, parece que Kendon traslada este tipo de significado estudiado en pragmática de los marcadores discursivos a los gestos.

Por último, aunque Kendon (2004) y Müller (1998, 2013, en prep.) incluyen en este nivel de clasificación funcional los gestos deícticos, nosotros lo hemos incluido dentro de los referenciales o pragmáticos según el contexto. Pues el hecho de que un gesto apunte es una cuestión relacionada con la forma del gesto y no con la función que desempeña; puede tener ciertas funciones asociadas, pero no son exclusivas y dependen del contexto secuencial, como otras formas gestuales.

3.4.3.2. Gestos referenciales

Los gestos referenciales realizan una contribución semántica al enunciado. Pueden ser una expresión paralela de lo significado por las palabras y, por tanto, redundante⁶¹, lo cual suele ocurrir en contextos en los que se quiere asegurar la comprensión de receptor (Kendon 2004: 179). Scherer (1977) denomina a esta relación *amplificación*, que apoya, refuerza o ilustra el contenido verbal. En otras situaciones, el gesto no es un equivalente quinésico de la palabra (Kendon 2004: 176-177, 185), sino que restringe su significado mostrando una forma específica de la acción u objeto; por ejemplo, utilizando el verbo *ir*, pero ilustrando con el cuerpo la acción de correr; o utilizando el verbo *pagar* con el gesto de *dinero*. El gesto también puede hacer referencia a algo no presente en el componente verbal o que sea una implicatura de este. Scherer (1977) diferencia este uso de *modificación*, que debilita o modifica el significado de la expresión verbal, de la *contradicción*, vinculada a la ironía⁶². Wehling (2017) incluye dentro del gesto narrativo-

⁶¹ Poyatos (1994a: 82) advierte contra el concepto de redundancia: “lo que parece ser redundante es la mayoría de las veces complementario”.

⁶² Un ejemplo de gesto irónico (De Fornel 1992: 172): las palabras (“irse a dormir”) y el gesto (“beber”) no se corresponden). Hay un significado primario del gesto (gesto + vínculo con el componente verbal) y

referencial, además de los que describen, modifican o amplían el contenido, los movimientos que presentan espacialmente los objetos en la narración.

Hay experimentos que corroboran el valor complementario de los gestos en el sentido del enunciado. En el de Melinger y Levelt (2004), dos grupos describen una escena con gestos y sin gestos y como resultado la información que se expresaba con gestos tendía a omitirse en el habla, frente a la descripción únicamente verbal. En el de Bavelas et al. (2008), en condiciones de no visibilidad los gestos referenciales mantienen la frecuencia, pero también son redundantes con el habla; mientras que cara a cara los gestos son significativamente menos tendentes a la redundancia con las palabras a las que acompaña temporalmente y aumentan las expresiones deícticas verbales y los gestos que señalan su referente.

En gestos referenciales que describen acciones, objetos y eventos y que son esenciales para comprender el enunciado, la mirada del hablante se dirige al gesto que él mismo realiza como acto deíctico (junto a expresiones verbales deícticas) que alerta al receptor de la relevancia del gesto para el apropiado entendimiento del habla (Streeck 1993, 2009a: 84-104). Los gestos manuales suelen encontrarse próximos a la cara del hablante para asegurar estar en el campo de visión del oyente. Gullber y Holmqvist (2006) confirman por *eye-tracking* que el cambio de mirada del recipiente al gesto descriptivo no está motivado por factores perceptuales o cognitivos, sino sociales. Mediante este cambio, los interactantes explícitamente y mutuamente lo reconocen y ratifican como oficial, como componente del acto comunicativo.

Los gestos referenciales no suelen ser convencionales y requieren contexto para interpretarlos (§3.4.1), aunque también existen repertorios de acciones ampliamente utilizadas (Kendon 2004: 161). Pero hay diferentes estrategias para vincular la forma al referente. Por ejemplo, se puede hacer referencia al objeto del tema creando una versión del objeto con las manos que ilustren sus características (forma, tamaño), su relación espacial o patrones de acción (Kendon 2004: 176, 185). Estas referencias pueden ilustrar expresiones verbales deícticas: “era así”, “la de aquí arriba”, “están como así”, “haciendo así”; o reemplazar a una palabra (Scherer 1977), tal como se ha visto en §3.4.2.4., con gestos léxicos o sintaxis mixta.

un significado secundario determinado por el significado primario y la implicatura conversacional que permita respetar la relación de congruencia.

Otra manera de referirse a un objeto es mediante gestos de señalamiento, ya sean con el dedo, la mano u otras partes del cuerpo como los ojos, las cejas, la cabeza o los labios fruncidos (véase para los ojos Goodwin 1981; véase para los labios Sherzer 1973, Enfield 2001), de manera que se crea un camino lineal invisible hacia los objetos físicos o reportados a los que se dirige la atención. En el estudio de Kendon (2004: 199-224) en Nápoles, se encuentra una relación entre la forma de apuntar y el uso del referente en el discurso: el dedo índice para distinguir un objeto específico y la mano extendida para hacer un comentario sobre el objeto, entre otros usos. Los participantes son flexibles, se adaptan a las oportunidades que tienen (Bangerter 2004): a mayor distancia, se recurre a la descripción verbal de su localización o sus características; a menor distancia se entiende sin ambigüedad el objeto presente y se señala sin acompañamiento de palabras, pero a menudo se combina con deícticos verbales. También pueden indicar objetos que están fuera de la vista por algún obstáculo visual o que se sitúan en un mapa invisible creado por el hablante (*espacio narrativo*, Haviland 1993; *deixis abstracta*, McNeill, Cassell y Levy 1993).

Respecto al autoseñalamiento, Montes (2009) identifica una serie de significados: uno mismo (acompaña a formas pronominales como *yo* o *a mí*); el hablante como ejemplo de una persona genérica; un grupo con el que el hablante se identifica; estados internos emocionales (se ubican en el centro del pecho) o cognitivos (en la cabeza); el punto cero de la deixis espacial; el cuerpo humano. El gesto de señalamiento es uno de los que pueden combinar una función pragmática (Kendon 2004: 199) (§3.4.3.3), pues no solo se refiere al objeto, sino que también indica que el hablante quiere hacer algo con ese objeto, bien ofrecerlo, bien que se lo den, especialmente cuando no acompaña al habla.

Aunque generalmente las clasificaciones de gestos se basan en gestos manuales, los gestos referenciales también son faciales (Bavelas y Chovil 2006, Bavelas *et al.* 2014). Como ya se ha dicho, se puede señalar con la cabeza, los ojos o los labios, pero además, si bien con la cara no se puede modelar o dibujar un objeto, la cara permite representar en una narración la expresión que uno mismo u otra persona puso, pondrá o pondría en una situación pasada, futura o hipotética. La expresión de emociones en este caso no representa la emoción real del interlocutor, sino la emoción de un personaje de su narración. La diferencia entre estos dos niveles de enunciación es clara en la expresión porque se seleccionan las características relevantes de las emociones reportadas o se

exageran y se borran las irrelevantes; además, el cambio de la expresión genuina a la reportada es rápido y se sincroniza con el discurso verbal.

Una motivación para utilizar gestos referenciales es el dinamismo comunicativo (Levy y McNeill 1992), ya que la información ya conocida por el oyente tiende a formas más atenuadas como gestos rítmicos o a no utilizar ningún gesto, mientras que la información nueva es alta en dinamismo comunicativo y se utilizan más gestos icónicos. Lo mismo confirman Gerwing y Bavelas (2004) en un experimento describiendo juguetes: el hablante utiliza gestos menos complejos, precisos e informativos cuando el oyente ya conoce la información. A lo largo de la conversación el grado de conocimiento se va equiparando y observan que la referencia a los elementos tiene una forma más esquemática, pues es suficiente para el propósito comunicativo.

3.4.3.3. Gestos pragmáticos

Aunque tanto en los gestos representativos como en los pragmáticos las manos actúan como si manipularan objetos, difieren en que los primeros corporeízan el objeto del que se habla y los segundos realizan una acción comunicativa (Müller 2013, en prep.). Los gestos pragmáticos son acciones que no pertenecen al contenido proposicional del discurso del hablante: “any of the ways in which gestures may relate to features of an utterance’s meaning that are not a part of its referential meaning or propositional content.” (Kendon 2004: 158). Además de esta definición negativa (“lo que no es referencial”), Müller (en prep.) explica que en estos gestos existe una fuerza ilocutiva vinculada a la realización de actos comunicativos. A menudo estos gestos aparecen sin acompañar a una intervención verbal, así que son medios eficientes para involucrar al destinatario sin ocupar el turno y coordinar las contribuciones a la conversación (Bavelas 1994).

Resulta útil la subclasificación de gestos pragmáticos para comprender las distintas funciones que entran dentro de esta categoría. Partimos de la tipología de Kendon (2004) en gestos discursivos (*parsing*), gestos modales y gestos performativos, pero añadimos un cuarto tipo, los gestos interactivos, pues su función, descrita por otros autores (Ekman y Friesen 1969, Bavelas et al. 1992, Cosnier y Vaysse 1997, Streeck 2005, Bohle 2013, Wehling 2017), no encaja en la tipología de Kendon.

De todas maneras, no se puede tomar de forma estricta esta tipología porque, una vez más, en los gestos la frontera entre funciones no siempre es clara y un gesto puede asumir varias de estas funciones pragmáticas a la vez, dependiendo de su contexto de uso

(Kendon 2004, Streek 2009, Bressemer y Müller 2014a). Kendon (2017) ejemplifica que una función modal a veces no es fácilmente distinguible de una función performativa y declara que todavía no está del todo satisfecho con su clasificación de funciones, así que hay margen de mejora. Por ello, en cada categoría se mencionan descripciones y otras formas de clasificación de otros autores.

3.4.3.3.1. Gestos rítmicos y discursivos

Los gestos rítmicos siguen el ritmo prosódico sin referirse al contenido. A menudo no entran en la categoría de gestos (Iriskhanova y Cienki 2018: 27) porque no se ven afectados por las manipulaciones de visibilidad (Alibali et al. 2001) y se les atribuye semiconsciencia, así que parecen ser gestos de fondo (Cooperrider 2017); sin embargo, otros autores que vemos en este apartado le atribuyen una función pragmática similar a la prosodia.

Se han denominados *batons* (Efron 1970 [1941], Ekman y Friesen 1969), *beats* (McNeill y Levy 1982, McNeill 1992) y *marcadiscursos* (Poyatos 1994). Poyatos los describe como “conductas (sobre todo kinésicas) conscientes o inconscientes que puntúan y refuerzan (con movimientos de cabeza, cejas, mirada, manos y tronco), la sucesión acústica y gramatical de palabras y frases, según su posición y relevancia, y coinciden con los símbolos de puntuación” (1994: 189). No son segmentables, o sea, no pueden ser independientes, no se separan del turno de habla (Cestero 2016).

Son normalmente movimientos de mano bifásicos rápidos, pequeños y de baja intensidad, aunque también se realizan con la cabeza (McNeill 1992); en el caso de los realizados con las manos, se producen donde estas estén en ese momento, a veces incluso con la configuración formal de otro gesto en su fase de mantenimiento. La cabeza se mueve casi constantemente con esta función, enfatizando la sílaba tónica (Maynard 1989). Estos movimientos suelen situarse también después de pausas entre oraciones o cláusulas, pero la cabeza tiende a permanecer quieta durante pausas y cuando se tiene el rol de oyente; además, hay una correlación entre los movimientos rápidos y los picos de voz alta (Hadar *et al.* 1983).

Las expresiones faciales también acompañan al habla enfatizando, subrayando y modulando el contenido (Ekman 1976). Se puede decir que los gestos faciales tienen una función sintáctica que marca pregunta o énfasis (Chovil 1991, Bavelas y Chovil 1997) siendo redundantes con la prosodia; por ejemplo, un levantamiento de cejas que enfatiza

una palabra o frase y sonrisas para puntuar partes de una narración o su final (Bavelas et al. 2014).

En general, la ratio de gestos representacionales decrece cuando el oyente no los mira, pero no la de los gestos rítmicos (Alibali et al 2001), lo cual es un indicio de que es una herramienta del hablante para la producción del discurso.

Respecto a la clasificación de esta función, Bavelas (1994) incluye los gestos rítmicos dentro de los interactivos (termino que corresponde a los pragmáticos en su conjunto), ya que enfatizan palabras, no contienen significado y no avanzan el tema de conversación; así como Ekman y Friesen (1969) los incluyen en el término *reguladores*. Pero Scherer (1977) los coloca en una categoría diferente, la parasintáctica, pues las cuestiones de pausa y tempo entran dentro de la prosodia, que considera dentro de la Lingüística, no de la Pragmática.

Los gestos rítmicos suelen incluirse en una misma categoría con los gestos que bosquejan en el aire la dirección y la estructura del discurso o pensamiento, organizando el flujo de habla sin hacer referencia a los objetos o ideas. Por ejemplo, mostrando los componentes lógicos del discurso con movimientos de mano a izquierda y derecha, o con golpes de mano que tienen la función de listar elementos (Müller 2004: 254). A estos gestos Efron (1970 [1941]) los llama *ideografías* y los incluye junto a los *batons* en la categoría de *gestos lógicos o discursivos*. Müller (1998) y Kendon (2004, 2013) también los unen con las denominaciones *gesto discursivo* y *parsing* respectivamente.

Un ejemplo muy estudiado de gesto convencionalizado con esta función discursiva es el *PUOH*, cuya forma (la mano abierta con la palma hacia arriba) se puede explicar por la extensión funcional de la acción instrumental de dar y recibir un objeto, presentando un objeto abstracto y discursivo como una entidad concreta y manipulable (Müller 2004). Debido a este proceso metafórico McNeill (1992) clasifica este gesto y otros con función discursiva como metafóricos. Una de sus funciones principales se puede clasificar como interactiva porque apela al interlocutor, por consiguiente, volveremos a mencionar este gesto en §3.4.3.3.4; pero este gesto presenta otras variantes⁶³ que pueden encajar mejor en esta función discursiva: (Müller 2004): 1) *PUOH* repetido que se sincroniza con el habla para ofrecer un listado de argumentos; 2) *PUOH* rotatorio que

⁶³ Müller (2004) considera que el *PUOH*, por su amplitud de funciones y variantes, forma por sí mismo una familia de gestos que comparte como características kinésicas la forma de la mano y la orientación.

indica la continuación de los argumentos ofrecidos; 3) *PUOH* alternante, que alterna ambas manos hacia abajo para ordenar los argumentos ofrecidos —los dos últimos son menos frecuentes que el resto—; 4) *PUOH* con movimiento lateral antagonista (los dos brazos se entrecruzan como unas tijeras) que no solo presenta algo, sino que indica parte del contenido: un corte, una ruptura, un abandono; 5) *PUOH* de ambas manos con movimiento lateral que las aleja entre sí para cubrir un amplio rango de objetos discursivos ofrecidos. Asimismo, los gestos deícticos también pueden señalar una palabra o enunciado para ayudar a la interpretación de la información (Jokinen y Vanhasalo 2009).

Las ideografías también se pueden realizar con la cabeza (McClave 2000), por ejemplo, moviéndola por cada ítem nombrado en un listado o en una presentación de alternativas entre referentes físicos o abstractos; a menudo el movimiento se realiza hacia una posición contraria. Kendon (1972) fue el primero en notar patrones de cabeza según la función discursiva del enunciado. Señala que un cambio en la posición de la cabeza se relaciona con el nivel jerárquico de las unidades de habla, pues tiende a aparecer al inicio y al final de una locución, mientras que las locuciones que son comentarios o paréntesis siguen otros patrones.

La categoría de gestos interactivos/pragmáticos de Bavelas (1994) no crea subdivisiones estrictas, pero algunas funciones que especifica sobre gestos que marcan el lugar de la información en el discurso del hablante quizás podrían entrar mejor en esta categoría de gesto discursivo que en otras que veremos posteriormente. Pues Kendon (1995: 164, 2004: 225-247) observa que los gestos que destacan aspectos del discurso o de la estructura del discurso lo hacen porque la información es relevante o para diferenciar el tópico del comentario. Las funciones que menciona Bavelas (1994) relacionadas con la información son indicaciones sobre: 1) que es información nueva y relevante, 2) que ya es conocida, 4) que debe tratarse como un aparte del tema principal o 4) que el oyente debe completar el significado de la información por sí mismo (Bavelas 1994: 213).

3.4.3.3.2. Gestos modales

El gesto modal aporta un marco interpretativo para lo que se dice verbalmente, de manera que se interpreta lo dicho como hipótesis, aserción, cita... (Kendon 2004:159). Uno de los múltiples usos de el *PUOH*, que metafóricamente muestra el objeto hablado en la palma de la mano, expresa que algo de lo que se habla es claramente visible y, por

tanto, obvio (Müller 2004). Mediante el gesto *throwing away* se cualifica algo dicho como irrelevante (Bressemer y Müller 2014b; Müller en prep.). En el sur de Italia, el gesto de comillas con los dedos se usa para citar lo dicho por otro o indicar que lo que se dice es hipotético (Kendon 2004). En relación con esta función citativa, (McClave 2000) observa que los movimientos de cabeza del hablante marcan cambios de estilo indirecto a directo, de narrador a personaje (aunque en citas largas puede volver a la posición neutral mirando al oyente hasta el final de la cita). Además, durante el discurso reportado los movimientos de cabeza también cumplen funciones referenciales deícticas, pues se adaptan a la situación espacial creada e indican la posición de los personajes o la imagen que se proyecta de ellos (por ejemplo, sus alturas: el adulto que habla al niño mira hacia abajo y viceversa).

Otro movimiento de cabeza con función modal es la sacudida lateral que no niega la declaración, sino que reconoce otras posibilidades, falta de información o incertidumbre, y se combina con marcas modales como “I guess”, “I think”, “whatever”, “whoever” (McClave 2000).

Partiendo de la clasificación de Kendon, el grupo de Bavelas (Bavelas et al. 2014) indica las funciones pragmáticas de algunos gestos faciales y considera modal el *facial shrug*, o sea, el equivalente facial al encogimiento de hombros, con levantamiento de cejas y retracción de una esquina de la boca, que transmite que algo dicho no importa (por ejemplo, que se ha dicho lo suficiente o que se ha dicho bien). Véase una descripción más detallada de este gesto en §3.4.3.3. Asimismo, las sonrisas pueden indicar equivocación o humor (Bavelas y Chovil 2018).

Respecto a la expresión facial coverbal, esta muestra la reacción personal del hablante sobre lo que está diciendo en ese momento (Bavelas y Chovil 2018). Tienen características formales distintas a las emocionales: son más rápidas y se sincronizan con el comportamiento verbal concomitante (Ricci Bitti 2014). Cabe mencionar un estudio sobre la distinción de oraciones interrogativas absolutas y oraciones interrogativas de incredulidad en hablantes de valenciano y holandés (Crespo Sendra et al. 2013). En dicho estudio se concluye que los valencianos se fijan en los gestos faciales para interpretar estos significados cuando no es suficiente el criterio acústico, mientras que las marcas acústicas en holandés son más claras y no requieren la confirmación visual.

También podríamos interpretar como modales lo que Poyatos (1994) observa como rasgos informativos añadidos al enunciado verbal, que refuerzan su sentido

(acercarse mucho a alguien con quien se está discutiendo, abrazar con cariño a una persona que se despide), lo debilitan (ligera sonrisa diciendo “te has equivocado”) o lo contradicen (expresión triste o preocupada con mirada abajo diciendo “no pasa nada”). La contradicción entre el habla y el gesto también puede ser un recurso sarcástico o de broma (Bugental, Kaswan y Love 1970, Coates 1991)

En clasificaciones posteriores a la de 2004, Kendon (2013, 2017) incluye otra función pragmática, la operacional, que no es común en las de otros autores. Estos gestos operan en relación con lo que se expresa verbalmente, confirmando, negándolo o desmintiéndolo. Un ejemplo de negación es el giro de la cabeza a izquierda y derecha, que, además de negar semánticamente la proposición, se puede relacionar con la propia enunciación en autocorrecciones o para expresar duda sobre lo dicho o cómo se ha dicho (Kendon 2002). Para Bouvet y Morel (2002: 104), esta negación sobre la capacidad de decir bien lo que se quiere decir es modal. A la recuperación léxica le suelen preceder sacudidas laterales o pequeños temblores laterales como muestra de ese proceso cognitivo (McClave 2000). Con el mismo sentido, también serían operacionales los gestos de la familia *Away* que se detallan en §3.4.3.3.3. Sin embargo, a falta de más ejemplos de Kendon o de otros autores, quizás esta función se podría subsumir en la modal o en la performativa, en tanto que afecta al sentido pragmático del enunciado, o en la discursiva, pues contribuye a interpretar el lugar del enunciado dentro del discurso.

3.4.3.3.3 Gestos performativos

Los gestos performativos manifiestan el acto de habla o la fuerza ilocutiva del enunciado verbal (Kendon 2004: 159), si es pregunta, petición, ofrecimiento, ejemplo, queja, evaluación, rechazo, ataque, etc. “It becomes clear that gestures also perform speech acts contributing to the verbal speech acts in a variety of ways” (Bressem et al. 2013: 1113). Este gesto puede aparecer incluso antes del enunciado verbal, anticipando el tipo de acto que va a ser (Streeck y Hartge 1992, Streeck 2009b). Fenómeno que aparece también en las narraciones, cuando el hablante reacciona antes a lo que va a contar para preparar al oyente (Ruusuvuori y Peräkylä 2009).

Además, esta función performativa puede llevarse a cabo como otro acto diferente al verbal dentro del turno. Por ejemplo, Kendon (2002) explica que una sacudida de cabeza en un enunciado verbal se puede interpretar como otro tipo de acto, bien expresando una actitud negativa hacia la información dada verbalmente, bien

respondiendo a una pregunta con el gesto y completando el turno con una opinión expresada verbalmente. Son dos actos distintos pero relacionados.

El gesto que no acompaña a ningún acto verbal realiza el único acto de ese turno (Kendon 2004: 225). La acción subyacente no se muestra ni se hace referencia a ella, sino que se realiza, se cumple (Müller 1998). En las conversaciones cotidianas suelen aparecer turnos totalmente gestuales, como una sacudida de cabeza, una palma de la mano hacia arriba o un señalamiento que bastan como respuesta a una pregunta (Cooperrider 2017).

En general, gestos menos y más convencionalizados pueden realizar actos de habla (Bressem, Ladewig y Müller 2013: 1113). No obstante, hay gestos pragmáticos más convencionalizados que parecen especializarse en actos de habla (Streeck 1994, Kendon 1995, 2004: 225). Tradicionalmente, se afirma que los emblemas se suelen utilizar como actos gestuales para saludos, insultos, direcciones interpersonales, réplicas, comentarios sobre el estado físico de uno mismo, expresiones de afecto y apariencia (Ekman y Friesen 1969). En la mayoría de los repertorios de emblemas, el significado es más pragmático que léxico⁶⁴ (Teßendorf 2013), pero, para Kendon, los gestos en posición independiente del habla normalmente no tienen significado nominal, sino que se refieren a una petición o actividad (Kendon 2004: 339-340); por ejemplo, el emblema de *dinero* para pedirlo o el de *abrir y cerrar tijeras* para pedir callar. Asimismo, los gestos referenciales realizados mediante la forma convencional del señalamiento pueden servir, en esta posición, como respuesta a una pregunta (Ehlich 2013: 652).

Payrató (2014) propone que las características del emblema no son las que lo convierten en un acto de habla, sino que su función como acto de habla es la que genera estas características del emblema; es decir, es su uso como acto pragmático el que establece un núcleo semántico básico —dependiente de factores situacionales y valores funcionales, como todo acto ilocutivo—, el que favorece el proceso gradual de convencionalización y el que permite su autonomía y su equivalencia con un acto verbal.

Payrató (1993) aplica la clasificación de actos de habla de Searle (1979) para el repertorio de emblemas del catalán y concluye que los gestos directivos, los de control interpersonal y los que se basan en acciones interactivas son los que tienden a la emblematización. Quizás por esto encuentra una relación entre los emblemas y el habla coloquial, mientras que en discursos más formales son más frecuentes los gestos

⁶⁴ Al emblema léxico Kendon (2004) lo denomina *narrow gloss gesture*.

coverbales no convencionales que complementan el significado y marcan el ritmo del discurso (Payrató 2004). Bressemer y Müller (2014a) también clasifican una serie de gestos recurrentes pragmáticos a partir de los actos de Searle y encuentran que la mayoría de los gestos son asertivos y directivos, los expresivos son frecuentes y los compromisivos son raros.

Más allá de los cinco tipos de Searle, las clasificaciones más detalladas de actos de habla son infinitas (véase Kerbrat-Orecchioni 1990: 230-231). Tomando como referencia las funciones asignadas a los emblemas de Kendon (1981) —quien toma en cuenta emblemas de Colombia, Estados Unidos, el Este de África, el Sur de Italia, Francia e Irán—, de Matsumoto y Hwang (2013)— quien los toma de Estados Unidos, el África Sub-Sahariana, el este de Asia, el sur de Asia y Latinoamérica— y de Cestero (2016: 9, Cestero et al. 2020), intentamos resumir las funciones performativas de los gestos:

- Influir en el interlocutor (Cestero et al. 2020): ordenar, pedir, insultar, amenazar y protegerse (Kendon 1981, 2004, Matsumoto y Hwang 2013), jurar o prometer (Kendon 1981).
- Facilitar las interacciones sociales mediante convenciones y normas, por ejemplo, el agradecimiento, los saludos y despedidas, los actos y símbolos religiosos, pedir disculpas, (Matsumoto y Hwang 2013), presentar, felicitar (Cestero 2016).
- Expresar ideas: comentar o describir a alguien o a sus acciones (Kendon 1981); expresar actitudes, opiniones y evaluar algo (Matsumoto y Hwang 2013; Cestero et al. 2020); dar respuestas evaluativas (Kendon 1981); expresar gustos, deseos y sentimientos (Cestero et al. 2020); hacer declaraciones sobre las relaciones personales y sobre estados mentales y físicos (Matsumoto y Hwang 2013) — quizás equivalente a anunciar del estado actual de uno mismo (Kendon 2004)—.

También se mencionan algunas funciones para los emblemas que podríamos relacionar más con las referenciales —indicar o referirse a algo o a alguien (Matsumoto y Hwang 2013); designar objetos y acciones (Kendon 1981)— o las interactivas (Cestero 2016).

A continuación, mencionamos algunos gestos o familias de gestos que se han descrito como performativos en la bibliografía.

Performativos manuales

Algunos gestos manuales que acompañan al habla informan sobre el acto que se está llevando a cabo. A veces su forma deriva de acciones cotidianas reales, cuyos

movimientos se extienden a entidades abstractas: “speakers visually conceptualize their conversational actions in analogy to instrumental actions” (Streeck 1994:1-2).

Kendon explica las funciones de la familia *Open Hand Supine*, cuyos tres miembros comparten la orientación de la palma hacia arriba (Kendon 2004: 265). El gesto *Palm Addressed*, orientado hacia el interlocutor, es una apelación al otro que refuerza ciertos enunciados como pregunta u otro acto que requiera la respuesta del interlocutor. El hecho de que en ocasiones este gesto no se retraiga hasta que la otra persona tome el turno es una evidencia de que se espera una respuesta (Streeck 2009b: 175). La variedad *Palm Presentation*, en la que simplemente se gira la muñeca para orientar la mano hacia arriba, realiza los actos similares de presentación, solicitud y ofrecimiento; mientras que el gesto *Palm Lateral* suma a la orientación hacia arriba el movimiento lateral (o incluso hacia atrás) para realizar un acto de rechazo.

Siguiendo el campo semántico del rechazo, la familia *Away* incluye gestos que comparten un movimiento que se aleja del cuerpo y se asocia con la noción de exclusión (Bressem y Müller 2014a, 2014b): *Throwing away*, que evalúa negativamente el tema de conversación molesto y está formado por la mano ahuecada orientada en sentido vertical, con la palma de la mano en dirección contraria al cuerpo del hablante, que se agita hacia abajo desde la muñeca como si lanzara un objeto molesto (Bressem y Müller 2014a: 1581); *Sweeping away* gesto de rechazo mediante la palma plana hacia abajo con movimiento lateral (Bressem y Müller 2014a: 1582); y *Brushing aside*, gesto de barrido usado en España con el dorso de la mano para rechazar una idea, “apartando” metafóricamente los objetos discursivos o el comportamiento de otros (Teßendorf 2014). La variante de este último realizada a la altura de los hombros suele funcionar como enunciado por sí mismo sin acompañar al habla o al final del enunciado verbal (Payrato y Teßendorf 2014: 1536).

Por otro lado, el *gesto cíclico* (Ladewig 2014b), formado por el movimiento circular de la mano desde la muñeca, además de usos referenciales, se utiliza performativamente en la búsqueda de palabras y para animar a un interlocutor a continuar una actividad (de habla) en curso.

Los gestos déicticos pueden complementar semánticamente el enunciado, como hemos visto en §3.4.3.2, pero también pueden completarlo pragmáticamente, aumentando la probabilidad de que se interprete como una petición indirecta (Kelly *et al.* 1999: 577). Normalmente se utiliza la mano, pero también se pueden utilizar los ojos, las cejas, la

cabeza o los labios fruncidos. Algunos estudios sugieren que hay diferencias en la forma de la mano según la función. Calbris (1990) apunta que se utiliza el dedo índice para dar órdenes y la mano con todos los dedos extendidos para presentar u ofrecer, mientras que utilizar la cabeza o el pulgar resulta maleducado. Kendon (2004: 199-224) se centra en la orientación de la palma: es la posición neutra para señalar es la palma en vertical, mientras que con palma arriba el hablante presenta, indica que está listo para recibir o reconoce a otro el origen del dato, que tiene razón o que está de acuerdo con él; el señalamiento mediante la mano abierta oblicua hacia arriba y el antebrazo parcialmente hacia arriba se dirige a un objeto sobre el que se ha hecho un comentario, sobre todo cuando es hacia una persona y es negativo o crítico.

Performativos con hombros

El gesto *shrug* no se puede traducir porque implica distintos movimientos corporales para los que no existe una palabra en castellano. Generalmente, se refiere al encogimiento o alzamiento de uno o ambos hombros, pero también se combina con otros movimientos: la cabeza inclinada lateralmente hacia un hombro elevado; cejas elevadas; el encogimiento de la boca, con la barbilla levantada y las comisuras de la boca bajadas (Morris 1994: 165) o los labios separados (Streeck 2009a: 190) o haciendo pucheros (Givens 1977: 13); una rotación hacia arriba de uno o ambos antebrazos; a veces con la palma abierta hacia arriba y los dedos extendidos; un movimiento lateral de la mano alejándose del cuerpo (Kendon 2004: 275); la sacudida de la cabeza en un contexto de obviedad expresa una negación implícita (Jehoul *et al.* 2017). El gesto varía formalmente en el uso de más o menos de estos elementos corporales. Por esto se ha dicho que es gesto kinésico compuesto o complejo (Streeck 2009a: 189; Morris 1994: 165, Givens 1977: 26).

Su significado se ha explicado como descargo de responsabilidad, traducible por "no sé", "no tiene nada que ver conmigo" o "no lo entiendo" (Morris 1994); muestra de distanciamiento y desentendimiento (Streeck 2009a: 189-191); indisposición para actuar o para ofrecer una solución (Kendon 2004: 175); en algunas culturas del sudeste asiático señala vergüenza (Elfenbein y Ambady 2002). Debras (2017: 23-27) identifica cinco significados: 1) incapacidad y 2) no responsabilidad (inacción), que suelen formarse mediante los antebrazos en posición supina; 3) distancia afectiva o indiferencia (y también rechazo), tiende a realizarse mediante un gesto más completo (inclinación lateral

de la cabeza + elevación del hombro + antebrazos en posición supina); significados epistémicos como 4) la expresión de indeterminación (e ignorancia), altamente relacionada con la boca, y 5) el *common ground*, que incluye la expresión de obviedad y conocimiento compartido, que no se relaciona con ninguna forma específica. A menudo es repetido por el interlocutor como muestra de la comprensión compartida en forma de reformulación verbal o alineación gestual.

Performativos con cabeza

Cambiando ya de gesto, las sacudidas de cabeza de izquierda a derecha y de arriba abajo, son emblemas expandidos geográficamente⁶⁵ que equivalen al significado semántico de negación⁶⁶ y afirmación, respectivamente, en ausencia de habla (McClave 2000). Si los adverbios *sí* y *no* son equivalentes a enunciados, “ya que pueden responder de forma aislada a una interrogativa total y son portadores de fuerza ilocutiva” (Pons Bordería 2022), consideramos que estos gestos de cabeza pueden funcionar de igual manera. Kendon (2002) explica los usos de la sacudida de cabeza y entre los performativos encontramos la negación del receptor mientras el hablante habla a modo de comentario y el gesto de negación que sustituye a la partícula verbal como respuesta a preguntas absolutas. Propone algunos factores que propician la aparición de este último uso: 1) se responde a una pregunta, pero se da más importancia a la información que va a añadir a continuación⁶⁷; 2) no se niega explícitamente a un oponente para evitar disputas; 3) lo emite un niño hacia una persona mayor. Otro movimiento de cabeza es el tirón repentino hacia atrás para mostrar sorpresa o un tipo particular de comprensión (M.H. Goodwin 1980: 309).

Performativos faciales

Generalmente, las emociones expresadas mediante movimientos faciales no se consideran un gesto comunicativo ostensivo, sino reveladores de información

⁶⁵ No son universales, como afirmó Darwin (1872). Jakobson (1972: 93) confirma que estos movimientos funcionan así en la mayoría de las culturas europeas, salvo en Bulgaria y algunos grupos étnicos de la península balcánica y Oriente Próximo.

⁶⁶ La negación mediante movimiento lateral también se puede realizar con el dedo índice estirado (Bressem y Müller 2014a: 1583).

⁶⁷ McClave (2000) Goodwin (1980: 3169) y Schegloff (1987:105-106) observan que los hablantes sacuden sus cabezas con enunciados afirmativos si contrastan con el enunciado inmediatamente anterior de un interlocutor.

intrapésica del individuo —ansiedad, tensión, dolor, triunfo, alegría, etc. (Knapp 1972)—, así como procesos cognitivos de atención o concentración (De Sanctis 1902, *apud*. Ricci Bitti 2014: 1344). Sin embargo, algunos distinguen una intencionalidad interactiva en ciertas expresiones. Ekman (1979) identifica algunas funciones socialmente orientadas que son reacción a lo que se dice o no se dice. Aunque Kendon (2004: 8-9) indica que ciertas risas, sonrisas y llores son comunicativos en un uso fingido o performativo⁶⁸, se centra en los gestos manuales. Pero Bavelas (Bavelas y Chovil 2006, Bavelas, Gerwing y Healing 2014) y su grupo traslada la tipología de funciones de los actos visibles de Kendon (2004) a los gestos faciales conversacionales, siguiendo el estudio de Chovil (1991), que encuentra que solo un cuarto de los 720 gestos faciales analizados eran reacciones emocionales personales y que el resto se relacionaban semántica o sintácticamente con el tema del diálogo. También desde estudios neuropsicológicos se argumenta la función lingüística de algunas expresiones faciales coverbales (Ricci Bitti 2014): las expresiones faciales emocionales se procesan en el hemisferio derecho y las lingüísticas, en el izquierdo, pues, según el daño en una zona u otra, los pacientes de afasia pueden realizar unas u otras funciones (Adolphs *et al.* 1996, Borod *et al.* 1998).

Algunos ejemplos de emblemas realizados con la cara son: el guiño para transmitir complicidad y acuerdo recíproco (Ricci Bitti 2014); tirar la cabeza para atrás, elevar las cejas y entrecerrar los párpados para señalar negación en culturas mediterráneas (Morris *et al.* 1979); y otros en combinación con la mano (Morris *et al.* 1979), como besar la punta de los labios para expresar apreciación o como empujar hacia abajo el párpado inferior con el dedo índice para avisar o alertar.

Algunos gestos que se utilizan para preguntar acompañando o no al habla son el *grappolo* o *finger bunch* en Nápoles (Kendon 2004: 159), el levantamiento de cejas (Ekman 1979: 185, Nota *et al.* 2022) y el fruncimiento del entrecejo. Este último también puede mostrar incompreensión o desacuerdo a modo de retroalimentación (Poggi 2013). Hömke *et al.* (2022) muestran que las iniciaciones de reparación verbal con cejas fruncidas tienen más probabilidades de ser respondidas con aclaraciones, que además son más largas y aparecen rápidamente (alrededor de 230 ms), y que las cejas fruncidas por

⁶⁸ Este mismo criterio es el que se ha seguido en la segmentación del Corpus Val.Es.Co. 3.0 (Pons Bordería, 2022, 2023)

sí solas pueden ser suficientes para ocasionar una aclaración; por consiguiente, este movimiento desempeña un papel comunicativo y no es meramente un epifenómeno.

El gesto *eye roll* o *poner los ojos en blanco* (Clift 2022) suele ser una reacción independiente del habla. Indica disconformidad y/o exasperación con algo que se acaba de decir o hacer, pero no pretende interrumpir el progreso de la secuencia. El gesto puede estar dirigido al sujeto que provoca la protesta, pero a menudo se persigue la afiliación eligiendo a un tercer destinatario o no se dirige con la mirada en nadie en particular como invitación abierta a la colusión.

Una de las modalidades claves para realizar los actos evaluativos es la expresión facial (Peräkylä y Ruusuvuori 2006), mostrando una postura afectiva (como "divertido" o "sorprendente") ante el referente de su evaluación. Así como la simetría de las expresiones faciales entre los participantes reflejan la relación afectiva momentánea durante la actividad de evaluación, la falta de expresión facial de uno de los participantes parece expresar la falta de implicación afectiva mutua entre ellos. La coordinación de estas expresiones entre los interlocutores se lleva a cabo gracias al desarrollo secuencia y la mirada mutua en los momentos necesarios. La renuncia a la expresión facial positiva mutua está precedida por la retirada mutua de la mirada y el paso a una nueva acción verbal. La expresión facial del hablante también se utiliza interactivamente para solicitar la respuesta adecuada a la postura del narrador (Ruusuvuori y Peräkylä 2009).

3.4.3.3.4 Gestos interactivos

Aunque Kendon (2004: 159) menciona una función interactiva o interpersonal de gestos que indican mantenimiento del turno o a quién se dirige el turno, se le da o se le pide; no forman un grupo por sí mismo en su clasificación. Los *gestos gestores del discurso* son los que funcionan explícitamente para regular las interacciones discursivas con respecto al intercambio de contenidos y la toma de turno (Wehling 2017), también denominados *reguladores* (Ekman y Friesen 1969), *interactivos* (Bavelas et al. 1992), *sincronizadores* (Cosnier y Vaysse 1997) o *metapragmáticos* (Streeck 2005), *conductas conversacionales* (Poyatos 1996: 230-246). Su función es pragmática porque gestionan el desarrollo de la interacción y no aportan contenido semántico. Ayudan a mantener la conversación como un sistema social y hacen referencia al interlocutor (Bavelas et al. 1992: 469). Aun así, todos los gestos se pueden describir de alguna manera como

interactivos en el sentido de que cualquier tipo de gesto adapta su extensión para tomar, mantener o ceder el turno (Bohle 2013) (§3.4.2.3).

Estos gestos tienen consecuencias sociales directas en la organización local de la toma de turno. Pueden posicionarse a mitad, al final o al principio de la TCU y pueden funcionar para iniciar o finalizar el turno (Bavelas *et al.* 1992, Bohle 2007, Streeck y Hartge 1992) o incluso como herramienta para mantener el turno (Bohle 2007, Streeck y Hartge 1992).

El Microanálisis del diálogo cara a cara (*MFD*) ha estudiado especialmente el funcionamiento de los gestos interactivos. El carácter interactivo de estos gestos se demuestra en el subsiguiente e inmediato comportamiento del destinatario de forma estadísticamente significativa, en su incremento en el discurso dialogado y en condiciones de visibilidad entre los interlocutores (Bavelas *et al.* 1992, 1995, 2008). La variable más significativa es la de diálogo/monólogo, pero la visibilidad también aumenta los gestos interactivos, mientras que sin visibilidad tienden a ser redundantes con la palabra. Además, los gestos interactivos tienen menos tendencia a ser redundantes con las palabras que los gestos tópicos, o sea, que aparecen sin una referencia verbal que le corresponda; por consiguiente, son medios eficientes para involucrar al destinatario sin ocupar el turno y coordinar las contribuciones a la conversación (Bavelas 1994).

Bavelas *et al.* (1992, Bavelas 1994) identifican como indicio formal de la función interactiva la orientación del gesto manual hacia el oyente, normalmente simple y rápida, mostrando la palma hacia arriba o apuntando con el índice directamente al destinatario, por ejemplo, para pedir ayuda, para referirse a algo que él dijo o a algo que sabe. Por consiguiente, el PUOH, entre otras funciones vistas en §3.4.3.3.4, cuando se realiza con un movimiento direccional hacia el coparticipante, también se interpreta como interactivo porque realiza la cesión del turno al siguiente hablante en un lugar de transición relevante (Streeck y Hartge 1992).

Sin embargo, los gestos interactivos también se realizan mediante otras partes del cuerpo como como cabeza, labios, ojos, pies, pues con ellos también se puede apuntar sin intenciones referenciales (Wehling 2017). Además, los gestos interactivos pueden no ser deícticos hacia el interlocutor, aunque sí se mantienen dentro del eje gestualizador-interlocutor; por ejemplo, un gesto de escudo manual para interrumpir, el gesto de

inclinarse o alejarse de un interlocutor para tomar o ceder el turno, o incluso la no-acción cuando se le ha ofrecido a uno hablar (Wehling 2017).

Wehling (2017) encuentra la razón de esto en la motivación metafórica de estos gestos⁶⁹: ESPACIO DEL DISCURSO ES ESPACIO FÍSICO, HABLAR ES MOVIMIENTO HACIA DELANTE, COMUNICACIÓN ES INTERCAMBIO DE OBJETO, IDEAS SON OBJETOS. Por ello, los interlocutores usan sus manos como contenedores de las ideas que se ofrecen el uno al otro y el hablante asume su papel ocupando el espacio gestual compartido y bloqueando el acceso del interlocutor. La conversación se construye metafóricamente como un intercambio de fuerza física, que se manifiesta de cinco maneras: 1) la fuerza de empuje, con gestos hacia el espacio compartido para tomar el rol de hablante; 2) la fuerza de bloqueo activa, con gestos de escudo para interrumpir el turno de otro hablante; 3) la fuerza de bloqueo latente, mediante la ausencia de movimiento; 4) la eliminación de la fuerza cediendo el espacio, por ejemplo, inclinándose hacia atrás; y 5) la fuerza de atracción, con gestos que incitan a hablar al interlocutor, por ejemplo, el movimiento giratorio de mano autodirigido (véanse Müller 1998, Ladewig 2011). Dependiendo del contexto, estas fuerzas regulan cooperativamente una conversación amistosa (CONVERSACIÓN ES COREOGRAFÍA) o son gestos de control en el discurso argumentativo (DISCUSIÓN ES LUCHA FÍSICA). Los gestos inclusivo-cooperativos tienden a usar menos intensidad, mientras que los de control⁷⁰, más.

Sin embargo, no todos los gestos interactivos derivan de un sentido espacial metafórico. Además de los gestos *directly embodied metaphoric*, están los gestos *directly embodied non-metaphoric*, que parten de movimientos naturales como las expresiones emocionales (cara enfadada o de disgusto como señal para interrumpir) o muestras de intenciones (abrir la boca como señal de que se desea hablar), y los gestos *conventionalized embodied*, que son movimientos convencionalizados culturalmente, como levantar el índice para pedir el turno o asentir para mostrar acuerdo (Wehling 2017). Aun así, explica que las distinciones no son tan claras porque, al fin y al cabo, los gestos con motivación metafórica también están convencionalizados y porque los

⁶⁹ Véase McNeill (1992), Cienki (1998), Cienki y Müller (2008a, 2008b) sobre la construcción de la comunicación mediante metáforas conceptuales que se visualizan con los gestos.

⁷⁰ Hay pocos estudios que incluyan los gestos de control en contextos de conflicto: Bressemer y Müller (2014b), Calbris (2011), Müller (2017), Wehling (2010, 2012, 2013).

convencionalizados también pueden ser multifuncionales, no se restringen a un solo sentido.

Aparte de la coordinación de cualquier gesto con el habla para gestionar el turno, hay gestos parcialmente convencionalizados que realizan un acto relacionado con la interacción. En consecuencia, pueden catalogarse tanto como performativos como discursivos (Payrató y Teßendorf 2014: 1535). Una forma a la que le atribuyen varias funciones interactivas es el *PUOH*⁷¹, utilizado tanto por el hablante como por el oyente, ya sea para presentar algo, dar una explicación, un ofrecimiento o para recibir estos mensajes (Bavelas *et al.* 1992, Kendon 2004, Müller 2004⁷²). Según Müller (2004), presenta un objeto discursivo al interlocutor, se sugiere su inspección o acuerdo y le invita a compartir la perspectiva propuesta —realizarlo con ambas manos intensifica el ofrecimiento de acuerdo—. Para Streeck (1992, 2007) el hablante utiliza esta forma para cambiar el rol de hablante con su interlocutor y solicitar respuesta. A su vez, el oyente puede usar el *PUOH* para tomar la palabra ya sea interrumpiendo o aceptando la cesión del hablante actual (Lopez-Ozieblo 2020).

A continuación, repasaremos algunas descripciones y análisis sobre gestos interactivos. Como los estudios suelen centrarse en formas o en partes del cuerpo concretas, este resumen parte de una serie de funciones interactivas y en cada una se menciona cómo cada parte del cuerpo puede marcarlas.

Autoselección

Los gestos, incluyendo la mirada, la expresión facial y, especialmente, la boca, se utilizan como movimientos preparatorios (Duncan 1973, Kendon 1973) que proyectan el inicio de habla para tomar el turno, lo cual tiene el beneficio de ser menos intrusivo que una expresión verbal (Schegloff 1996). En esta forma de autoselección el hablante inicia un gesto justo antes del posible *completion point*, y no solo asegura el derecho a hablar, sino que adelanta el significado semántico y/o pragmático del acto (Streeck 2009b). En consecuencia, cualquier tipo de gesto puede adaptar su lugar de aparición para cumplir esta función en la toma de turno. No obstante, se han analizado algunos gestos específicos para la autoselección, como la palma hacia arriba, la *[a]-face* (Streeck y Hartge 1992), el

⁷¹ Véase más sobre este gesto con función performativa en §3.4.3.3.4.

⁷² Müller estudia el *PUOH* en una conversación entre una española y un alemán.

humedecimiento de labios, la orientación/acercamiento/inclinación hacia el oyente (Poyatos 1994: 233), el gesto de señalamiento (*pointing gesture*) (Mondada 2007), y otras manifestaciones corporales complejas (Schmitt 2005). La *[a]-face* es la articulación silenciosa de una vocal, que no se trata de una acción preparatoria para articular la palabra realmente porque también se produce para iniciar enunciados por una consonante, así que es una práctica convencional para informar de que se está listo para hablar. Por su parte, Mondada (2007) muestra que, al menos, en un ambiente profesional en el que se está prestando atención a unos documentos u objetos, el gesto señalamiento también sirve para reclamar el habla mientras el turno del hablante actual está en proceso. De manera que en estos casos no es un recurso referencial o deíctico, sino de autoselección.

Mantenimiento del turno

En situaciones de lucha por el turno, en las que dos emisores tratan de tomar la palabra, ya sea de una forma más amistosa o más conflictiva, el hablante mantiene el gesto que había iniciado para mantener sus derechos de habla⁷³ (Bohle 2014), o repite el gesto para demostrar que una secuencia no está completa (Streeck 2009b). En general, la gesticulación inhibe cualquier señal de cesión de turno y reduce los reclamos de turno del oyente (*gesticulation signal*, Duncan y Fiske 1977:188-189). Entonces, gestos con otra función referencial, discursiva, etc., se utilizan a la vez con esta función interactiva. Un gesto especializado en este uso puede ser la *cara de pensar* (Kendon 2004: 159). La lucha por el turno puede convertirse literalmente en una lucha de manos (Li 2014: 175-176), pues una forma muy directa de parar el turno de otro hablante cuando no acepta otras formas de interrupción es agarrar las manos con las que está gesticulando.

Cesión del turno

Entre otras marcas lingüísticas y prosódicas, el fin del gesto manual o la relajación desde una posición de mano tensa indican que el hablante cede el turno (*speaker turn signal*, Duncan y Fiske 1977: 183); aunque también pueden indicar el fin de una TCU (Mondada 2007). Li (2014: 158-162) encuentra en sus datos que para terminar el turno además se puede utilizar un autoadaptador (Ekman y Friesen 1969) en secuencias en las

⁷³ Por ejemplo, en el estudio recién mencionado de Mondada (2007), a veces los hablantes mantienen el gesto de apuntar a través de pares adyacentes o secuencias más largas, mostrando que los derechos del hablante no terminan en los límites de los turnos.

que el hablante da una información y opina sobre ella⁷⁴: cruza la mirada con el oyente y usa el autoadaptador en una zona visible (cara o cabeza). Si los autoadaptadores son propios de los oyentes, mientras que los hablantes gesticulan (Schegloff 1984), el paso de la gesticulación al autoadaptador indica este cambio de rol y la cesión del turno. Por eso mismo, los gestos atraen la mirada y los autoadaptadores la repelen (Goodwin 1986: 40), pues cada uno indica si se tiene o no se tiene algo que decir. Poyatos (1996: 230-246) también nombra este ocuparse con algo como marca de cierre del turno, así como el humedecimiento de labios, el saca-mete de lengua, la relajación postural y echarse hacia atrás.

Interrupción

Para interrumpir se puede utilizar el gesto *Open Hand Prone 'vertical palm'* (*OHP-VP*) (Kendon 2004, Bressemer y Müller 2014a, 2014b⁷⁵, Kamunen 2018), formado por una mano plana con la palma en posición vertical en dirección hacia el destinatario y el dorso hacia el gesticulador. Incluye variantes con la misma función en las que la palma no está completamente vertical, sino en ángulo oblicuo. Entre otros usos, este gesto controla la toma de turnos de un coparticipante, ya sea impidiendo la continuación de un turno en curso o adelantándose a una inminente autoselección, y reclama el turno de habla para uno mismo.

Este uso es frecuente en programas televisivos de tertulia, pero la interrupción también se utiliza en conversaciones mundanas con objetivos colaborativos: para evitar que el interlocutor hable sobre un contenido problemático o para interrumpir el turno de interlocutor porque el gesticulador está comprometido en una actividad diferente a la conversación (Kamunen 2018). Respecto a esta interrupción colaborativa, Poyatos (1994: 242) incluye como una actividad secundaria del oyente las señales apuntadoras que avisan de un problema al hablante: guiñar un ojo, abrir los ojos muchísimo, dar un codazo, señalar al alguien disimuladamente con el dedo, la cabeza o la mirada, revolverse en el asiento para distraerle, cogerle del brazo.

⁷⁴ Bouvet y Morel (2002: 109) vinculan estos *mouvements autocentrés de grattage* con las unidades del discurso: en sus ejemplos el hablante utiliza el autoadaptador cuando la introducción de un nuevo objeto del discurso resulta infructuosa y este termina cuando consigue reanudar el tema.

⁷⁵ Bressemer y Müller (2014a, 2014b) lo denominan *holding away*.

Elicitación del hablante al oyente

Una función pragmática que el hablante en su turno puede realizar mediante un gesto es la elicitación (*seeking gestures*, Bavelas et al. 1992, *función apelativa*, Müller 2013), bien para pedir algún tipo de respuesta relevante para la secuencia, bien para pedir la retroalimentación del oyente.

La retroalimentación puede mostrar que el oyente está escuchando, que está entendiendo o que está de acuerdo con lo que ha dicho el hablante. Tradicionalmente, se asumía que el oyente realizaba esta retroalimentación cuando quisiese, que era un proceso internamente motivado, pero el microanálisis revela que son, de hecho, respuestas a las peticiones del hablante. Las formas visuales utilizadas con este objetivo son el señalamiento, la mirada, el mencionado *POUH*, el asentimiento del hablante, gestos faciales... El grado de proyección del acto en la secuencia determina la necesidad de una técnica de selección de hablante más explícita, como la apelación verbal, o menos, como la mirada (Auer 2021).

Un gesto deíctico personal, es decir, apuntar a una persona que participa en la interacción, además de hacer referencia a esa persona en el contenido, apela a los participantes tal como lo hace un pronombre personal u otras expresiones de apelación. No solo se apunta con la mano, también, con la cabeza (a veces se combina con alzamiento de cejas), con los ojos, con la boca. Sin embargo, a menudo el gesto es más complejo porque hay una forma o un movimiento adicional de la mano que se combina con el componente deíctico para producir un significado específico (Bavelas *et al.* 1992: 473). Algunos estudios sugieren que hay diferencias en la forma de la mano según la función; por ejemplo, Calbris (1990) apunta que se utiliza el dedo índice para dar órdenes y la mano con todos los dedos extendidos para presentar u ofrecer.

Cuando el hablante está buscando una palabra adecuada y el oyente le ayuda a conseguirla para que continúe el turno, se produce la coconstrucción del turno de habla. Durante este proceso, el hablante puede utilizar herramientas multimodales para invitar al oyente a unirse a la búsqueda, como la *thinking face*, gestos manuales giratorios, mostrar la palma para recibir la palabra, levantar la palma hacia el frente para mantener el turno (Goodwin y Goodwin 1986, Streeck 1993, Hayashi 2003, Dressel 2020). También se ha evidenciado la importancia de la mirada en esta tarea: los sujetos tienden a apartar la mirada en los puntos de codificación difíciles (Kendon 1967), pero miran al oyente si desean una respuesta colaborativa. El oyente propone palabras y puede

acompañar esta búsqueda con la mimesis del gesto referencial utilizado por el hablante, contribuyendo así a la continuidad entre los turnos (Kimbara 2006).

El asentimiento mediante el movimiento de la cabeza de arriba abajo es una señal muy reconocible en el rol de oyente, pero también la usan los hablantes, o incluso más (McClave 2000: 874-875), para pedir retroalimentación al oyente, respuestas en momentos relevantes o para enfatizar información (Aiko 2011). Sin embargo, ya se ha observado en el estudio de este gesto que existe variación cultural en su uso (Maynard 1987, McClave 2000, Aiko 2011, Szatrowski 2000, Szatrowski 2003). También la inclinación lateral mantenida de la cabeza marca la solicitud de atención (Bouvet y Morel 2002: 104).

Los gestos faciales también se utilizan para pedir de retrocomunicación, bien con una expresión interrogativa sutil (Poyatos 1994: 230-246), bien con una expresión facial evaluativa (Ruusuvoori y Peräkylä 2009). Entrando en la multifuncionalidad, incluso los gestos rítmicos realizados de forma más marcada pueden elicitar la retroalimentación porque subrayan esa información (McNeill 1992: 171).

Gestos de retroalimentación del oyente

Además de otras formas de retroalimentación verbales y paralingüísticas (§2) y con la mirada (§3.3.2), el oyente también puede utilizar la cabeza, las manos y la cara. De hecho, el número de expresiones visuales de retroalimentación encontradas en el corpus ALICO supera el número de expresiones verbales (Włodarczak *et al.* 2012).

Uno de los gestos que comúnmente se vinculan a la retroalimentación es el asentimiento mediante el movimiento de la cabeza de arriba abajo, que ya se incluye en la primera mención a los *backchannels* (Yngve 1970), así como en Duncan (1972). El estudio de Stivers (2008: 31-35) observa que, durante las narraciones, los asentimientos del oyente se especializan en la expresión de afiliación con el punto de vista del hablante, mostrando su apoyo y respaldando la postura del narrador; mientras que los apoyos verbales muestran alineamiento formal con la actividad en progreso del hablante mediante continuadores y reconocedores. En general, los oyentes distraídos producen menos retroalimentación, pero además la proporción de retroalimentación no verbal (asentimiento) con respecto a la verbal aumenta significativamente en condición de distracción, lo que sugiere que los oyentes distraídos eligen una modalidad más básica para expresar la retroalimentación (Włodarczak *et al.* 2012). Los oyentes, además, a veces

mueven la cabeza en sincronía con el discurso del hablante a modo de gesto rítmico (Kendon 1970).

Otros gestos manuales más convencionalizados que se pueden considerar retroalimentadores son: gestos de cita, que señalan al hablante haciendo referencia a su discurso, a algo que ha dicho, mostrando acuerdo (Bavelas *et al.* 1992); así como el *PUOH*, que también es un gesto del receptor (Bavelas *et al.* 1992, Kendon 2004, Müller 2004).

También sirven como retroalimentaciones las expresiones faciales, como movimientos de las cejas o de los labios que confirman la atención u otras expresiones más pronunciadas que transmiten la reacción personal del oyente (Bavelas *et al.* 1992). De la misma manera que las sonrisas (Brunner 1979), con la misma sincronización, posición y función que los retroalimentadores verbales. Un parpadeo largo por sí mismo —con duración de más de 410 ms— también sirve como retroalimentación (Hömke, Holler y Levinson 2017), pues aparecen en la misma posición que otros continuadores: por un lado, al final de TCU, mostrando una recepción continuada, que deja pasar la oportunidad de tomar el turno y se alinea con el relato en curso; y, por otro lado, como respuesta a interrupciones y disfluencias intra-TCU, presumiblemente para camuflar estas disfluencias y reducir la carga cognitiva del hablante.

Ya se ha mencionado antes que el oyente puede contribuir con gestos manuales referenciales a la búsqueda de palabras del hablante. Pero también puede contribuir al turno aunque no se le solicite (Bolden 2003, Hayashi 2005, Iwasaki 2009) y estos elementos se pueden clasificar como retroalimentadores. Estos gestos del oyente se anticipan al habla o los gestos del hablante y hacen una proyección conceptual, evidenciando que los interlocutores habitan un mundo imaginado compartido (Hutchins y Nomura 2011: 42). De Fornel (1992) denomina a estos gestos icónicos con valor comunicativo *return gesture* cuando aparecen casi al mismo tiempo (durante el punto álgido) o después (en la recuperación) que el mismo gesto del hablante. El *return gesture* confirma el seguimiento o entendimiento de una manera más explícita que otros gestos de retroalimentación, mostrando una involucración más fuerte en la interacción. El significado del gesto referencial no viene de las palabras del oyente, sino del hablante, “they are in a contextual dependent relationship” y su juxtaposición no es fortuita. Aunque el gesto icónico del hablante no tenga una pretensión de elicitar la retroalimentación del oyente, este decide que es relevante para la organización secuencial

y lo incorpora en la retroalimentación. En actividades como el cumplimiento de una tarea, esta retroalimentación gestual sirve para mostrar el entendimiento mutuo de las instrucciones y hace el proceso más eficiente (Furuyama 2000, Clark y Krych 2004).

En la Tabla 9 se muestran todos los gestos kinésicos (incluyendo mirada y postura) que Poyatos (1994: 230-246) considera ejemplos de retrocomunicación.

Retrocomunicación	
Retrocomunicación	Sonrisa intermitente o continua, toda clase de gestos emblemáticos y exteriorizadores y, típicamente, imitando automáticamente algunos del hablante, especialmente los de reacciones (exteriorizadores) positivas o negativas Postura retrocediendo y avanzando o reorientándose
Atención	Silencio con la mirada fija y abierta Contacto ocular intermitente, asentimientos de cabeza, sonrisas, apoyando la barbilla en la mano+mirada fija, enarcamiento de cejas, inclinación anterior cruzando los brazos+mirada fija Inclinación del cuerpo hacia adelante
Inatención	Silencio aparentemente mirando al hablante, pero con falsa fijación (mirada vacía, perdida), mirando silenciosamente hacia otro sitio o explorando el entorno Fijación falsa incluso añadida a la mirada o a los asentimientos jugueteando con objetos Inclinación del cuerpo hacia atrás
Interés	Cabeceo frecuente, a veces repetidamente, a veces parpadeo más rápido y contacto ocular más prolongado, cambio postural al principio del turno del hablante o acompañando puntos importantes de su discurso
Desinterés	Movimientos inquietos, recostándose con la mano aplastando la mejilla+fijación falsa
Aprobación	Aplaudir, lanzando la palma de la mano oblicuamente hacia adelante
Desaprobación	Pateando, lanzando la palma de la mano oblicuamente hacia adelante

Tabla 9. Conducta conversacional de retrocomunicación (Poyatos 1994: 230-246).

Respecto al efecto de la retroalimentación en los gestos del hablante, la retroalimentación negativa provoca la hipergesticulación del hablante (Hoetjes *et al.* 2014); mientras que la que la tasa de gestos disminuye (así como sus cualidades formales) cuando la retroalimentación verbal es confirmatoria y se mantiene constante cuando la retroalimentación verbal es de aclaración, elaboración o corrección (Holler y Wilkin 2011). Al igual que el hablante abandona el turno verbal si no recibe ninguna ratificación del destinatario, los gestos también se abandonan y se vuelven a retomar cuando se recibe la mirada (Streeck 2009a: 114).

La función de los gestos que dan retroalimentación al hablante es claramente pragmática, pero en la subclasificación se pueden incluir tanto en los gestos interactivos como en los performativos, pues, como en Bavelas *et al.* (2014), el gesto de retroalimentación se considera un acto en sí mismo. Este hecho es aún más evidente en

reacciones del oyente que comportan un acto evaluativo, pues incluyen un sentido más concreto que la simple atención.

3.5. Recapitulación

A lo largo de este capítulo hemos revisado las propiedades y las funciones de los gestos que forman parte de la información multimodal que, desde un enfoque pragmático e interaccional, afecta a la interacción conversacional.

En primer lugar, hemos decidido marcar una diferencia entre las propiedades de la postura y la mirada con las propiedades del resto de movimientos corporales (cabeza, manos, cara, etc.). Con el objetivo de identificar los movimientos corporales que por sí mismos se pueden considerar turnos, los gestos manuales, faciales o de cabeza se aproximan a las características de la intervención verbal: unas limitaciones temporales breves y una fuerza ilocutiva concreta. En cambio, la mirada y la postura son movimientos que se inician en un momento concreto, pero se extienden durante varias intervenciones verbales o secuencias, por lo que no resulta claro segmentarlos como intervención. Tampoco tienen un significado claro, aunque sí se podría decir que tienen una función metacomunicativa que afecta en dos niveles: en un nivel macro, establecen el espacio de interacción, los interlocutores que participan y su grado de compromiso con la actividad; en un nivel micro, contribuyen en la gestión local de cada turno, junto a otras partes del cuerpo. Evidentemente, estas funciones de la mirada y la postura van a ser útiles para analizar la relación entre intervenciones de los hablantes y la validación de intervenciones, pero los consideraremos siempre como elementos coverbales que confirman la orientación de la intervención verbal y no como elementos independientes que constituyen intervención.

En segundo lugar, hemos revisado las propiedades de los *gestos* para comprender su uso en la interacción. La relación más o menos convencional entre la forma y el significado de los gestos nos interesa en tanto que se relaciona con la identificación de la función del gesto en la interacción y su capacidad para aparecer acompañando a la emisión verbal o por sí solo. La relación temporal entre el gesto y el turno verbal es importante para considerar su posición estructural en el modelo jerárquico desde el que analizamos la conversación. Por último, la clasificación funcional de Kendon (2004) y otros autores es necesaria porque seguimos un enfoque pragmático, en el que los gestos, como otros elementos lingüísticos, sirven tanto para transmitir contenido semántico como

para interpretar el contenido dentro de la interacción. Estas funciones pragmáticas de los gestos son: la discursiva, que relaciona los contenidos del discurso de un hablante; la modal, con la que el hablante cualifica el contenido; la interactiva, que vincula el contenido con la estructura dialógica entre hablante y oyente; y la performativa, que contribuye a la interpretación del contenido (o del propio gesto) como un tipo de acto. Siendo este un estudio sobre la estructura de la conversación, era necesario incorporar la función interactiva descrita por otros autores en la tipología general de Kendon.

Esta información nos ayuda a proponer en §5 los elementos visuales que forman parte del análisis de la segmentación visual de la conversación y la manera en la que encajan en la estructura de este género.

CAPÍTULO 4. METODOLOGÍAS PARA UN CORPUS MULTIMODAL

4.1. INTRODUCCIÓN	185
4.2. INCORPORACIÓN DEL VÍDEO EN EL ANÁLISIS DE LA INTERACCIÓN	185
4.3. CORPUS DE CONVERSACIONES COLOQUIALES EN VÍDEO	188
4.4. ANOTACIÓN DE GESTOS.....	192
4.4.1. SISTEMA DE ANOTACIÓN MUMIN.....	196
4.4.2. SISTEMA DE ANOTACIÓN LASG.....	198
4.4.3. SISTEMA DE TRANSCRIPCIÓN MULTIMODA DE MONDADA.....	203
4.5. RECAPITULACIÓN	204

4.1. Introducción

En el capítulo 3 se revisan los estudios sobre interacción multimodal para contribuir al objetivo de averiguar la manera en la que los gestos afectan estructural y socialmente a la conversación. Si bien esta información sirve de base para hacer una propuesta, esta no sería válida sin un estudio de corpus que la respalde.

Antes de explicar la metodología de nuestro estudio en el capítulo 5, este breve capítulo 4 trata algunas cuestiones sobre la metodología de recolección de datos y la anotación multimodal utilizada para el campo de los gestos: contextualiza el uso del vídeo en el estudio de la conversación, nombra algunos corpus existentes, plantea los problemas de la anotación formal de los gestos y muestra algunos sistemas de anotación y transcripción que sirven de base para nuestro estudio.

4.2. Incorporación del vídeo en el análisis de la interacción

El interés en grabar interacciones sociales que ocurren de forma natural tiene una larga tradición en la etnografía y así también en el AC (Sacks et al. 1974, Sacks 1984), ya que preserva las características de la interacción en su contexto y permite escuchar repetidamente el contenido y transcribirlo correctamente. Esta decisión inspira a otras perspectivas de estudio de la conversación en la utilización de transcripciones rigurosas que detallan todo lo que sucede momento a momento (Cortés Rodríguez 2002: 194, Fox et al. 2013: 729).

Las grabaciones de audio no permiten recoger los elementos visuales que participan en una conversación en presencia, por lo que el análisis de estas conversaciones siempre será parcial, frente al de la grabación en vídeo. Una grabación en audio es suficiente solamente en conversaciones sin presencia, como las telefónicas (Schegloff 1972, 1979, Francis y Hunston 1992⁷⁶), ya que las miradas y los gestos no forman parte del contexto de ese tipo de interacciones, los participantes y el analista comparten los mismos canales de información.

El primero en aplicar métodos empíricos con grabaciones en vídeo de situaciones naturales como fuente fue Efron (1970 [1941]) en su estudio de los gestos

⁷⁶ Francis y Hunston (1992) utilizan llamadas telefónicas porque su objetivo es la lengua, para que las miradas y los gestos no entren en juego.

conversacionales de la primera y segunda generación inmigrantes de judíos del Europa del Este y del sur de Italia en Nueva York, que podemos enmarcar dentro de una Etnografía Microanalítica (Streeck 2009a: 29). La escuela de Palo Alto en los 50 (Kendon 1990) también usó grabaciones en vídeo para entender el comportamiento no verbal de pacientes psiquiátricos y vio la necesidad de tratar la comunicación como un sistema integrado del comportamiento corporal y lingüístico en su contexto.

El vídeo ofreció una gran cantidad de datos sobre actividades corporales, objetos implicados en las actividades, características del entorno espacial al que se orientan los participantes, dispositivos con propiedades semióticas (como pantallas de ordenador, paneles de control, materiales escritos) y acontecimientos que ocurren en el fondo de la interacción que, de un modo u otro, parecen incidir en ella (Deppermann 2013: 4). Asimismo, la grabación en vídeo permitió un análisis en detalle, hasta entonces imposible, de la coordinación entre participantes y entre los sistemas que el participante utiliza (Goodwin 1981). En este análisis los actos audibles y visibles no son modos de comunicación separados, sino integrados (Bavelas et al. 2016), alejándonos de una visión del lenguaje logocéntrica (Mondada 2013c: 988). Los modos simultáneos que aparecen en los datos en vídeo demuestran un sistema más complejo de comunicación y provocan que lo que parecían casos similares en el habla se interpreten de forma muy diferente (Deppermann 2013: 4). Por consiguiente, esta herramienta ha favorecido tanto al desarrollo de los estudios sobre el gesto y la CNV como a los de la interacción en general.

Estamos totalmente de acuerdo con las siguientes palabras de Poyatos, pues coinciden con la premisa de este trabajo: “si una muestra de un corpus conversacional presenta solo el nivel verbal y el paralingüístico echaremos de menos el visual, es decir, el kinésico, sin el cual no puede completarse siquiera la estructura básica del discurso” (Poyatos 1996: 223).

Además, el avance de la tecnología permite hoy en día herramientas de monitoreo más precisas que el vídeo, como el *eye-tracking* o, incluso, la captura de movimiento 3D (Battersby et al. 2008), que ahorran tiempo de anotación manual; aunque los aparatos electrónicos necesarios pueden ser invasivos en el espacio de interacción (Kousidis et al. 2012).

Respecto al tipo de eventos grabados para el estudio del gesto, los centrados en un movimiento concreto o un tipo de gesto con una función pragmática (retroalimentación, conclusión de turno) o un valor semántico concreto (acuerdo, sustitución de un verbo),

sacrifican la espontaneidad en favor de que los gestos se capten desde un buen ángulo y con buena calidad, mediante grabaciones de televisión o de un experimento *ad hoc*, o combinando varias fuentes de datos. Algunos ejemplos de este tipo de estudios son los proyectos TimeGest (Pagán et al. 2020) o MultiNeg (Manero et al. 2023), ambos utilizan la base de datos de noticias televisivas NewScape de la UCLA (<http://newsscape.library.ucla.edu/>), el primero estudia patrones multimodales (expresión verbal + gestos) de representación del tiempo y el segundo, los patrones de desacuerdo y negación en interacción; el segundo, además, combina los datos con conversaciones grabadas en su laboratorio con un gancho. Otro estudio que toma datos equilibrados de varios géneros es el de los gestos repetitivos de Bressemer (2021: 13): conversaciones naturales, programas de juegos, debates, sesiones parlamentarias, juegos y datos experimentales.

Para analizar cómo los elementos visibles afectan a la estructura de la conversación se abren dos caminos metodológicos: el experimento y el trabajo de campo. El experimento es adecuado para dar validez a hipótesis sobre un fenómeno de menor aparición o sobre una variable muy concreta. Por ejemplo, Bavelas et al. (1992) realizaron dos experimentos para comprobar que el hablante realiza más gestos interactivos cuando habla hacia otra persona que en solitario y que realiza menos cuando no ve al oyente. Estas variables no requieren que la conversación tenga lugar en una situación natural y conserve las características prototípicas, como la negociación libre del turno y de los roles conversacionales. Sin embargo, “Without the precise reciprocity and collaboration inherent in a real dialogue, we cannot be sure that the gestures produced in the laboratory have anything in common with the gestures people use in everyday dialogues” (Gerwing y Bavelas 2013: 826). Los métodos que no incluyen interacción social real en el diálogo cara a cara tienen un carácter general cuestionable al uso natural de los gestos conversacionales (Gerwing y Bavelas 2013: 834).

Nuestra hipótesis afecta a la estructura de la conversación coloquial, por lo que es necesario dar libertad a los interlocutores para que desarrollen la conversación sin condiciones preestablecidas. No es posible seleccionar el material de medios públicos como programas de telerrealidad porque en ellos es común que se editen las pausas de los hablantes y los planos que muestran a cada participante, de manera que no contamos con el material en bruto. La selección de unos planos u otros es ya un trabajo de interpretación sobre el orden de la interacción (Mondada 2009). Tampoco encajan en la conversación

coloquial prototípica los pódcast con grabación en vídeo, que, si bien algunos presentan una gestión del turno más cercana a las características coloquiales, no dejan de tener el objetivo de entretener a un público. Téngase en cuenta que, respecto a otros géneros orales más formales, los hablantes en contextos formales tienden a inhibir gestos que usualmente sí realizan en contextos coloquiales, especialmente los arbitrarios son los más inhibidos (Rossini 2012: 48). La única manera de conseguir conversaciones genuinas es un trabajo de campo que aporte validez ecológica (Enfield 2013: 976).

Los datos naturalistas, grabados en el mismo entorno en el que se desarrollan habitualmente las actividades, son especialmente defendidos por la etnometodología y el análisis de la conversación, pues las acciones de los participantes y su temporalidad dependen fuertemente del contexto, del marco de participación, que no debe ser orquestado ni perturbado por los investigadores (Mondada 2013c: 987). Aunque la cámara de vídeo invada el espacio de interacción, se intenta mantener una situación de cotidianeidad. Por ejemplo, Peräkylä y Ruusuvuori (2006) tratan de imitar la situación cotidiana en el laboratorio, dando a los informantes de comer y grabando con tres cámaras y un micrófono; en cambio, Selting (2013) adopta la misma técnica, pero graba en domicilios particulares conversaciones entre dos o tres personas. Otra opción es la de Kendon (2002), quien graba en Nápoles y Salerno a personas que ya estaban hablando, pidiéndoles permiso; por consiguiente, son naturales en el sentido de que no se elicitaron, hubieran tenido lugar sin que hubiera aparecido el investigador para grabar la situación cotidiana. Sin embargo, complementa su material con algunas entrevistas o conversaciones solicitadas.

Estas grabaciones creadas para el estudio de los gestos se recopilan para su análisis. Algunos permanecen privados para los investigadores y otros se publican en corpus. En el siguiente apartado, nombraremos algunos corpus de español coloquial en audio y en vídeo, así como de otros idiomas.

4.3. Corpus de conversaciones coloquiales en vídeo

Entre los corpus orales en español, contamos con varios consultables de conversaciones coloquiales: Val.Es.Co. 3.0 (Pons Bordería 2022), con conversaciones de Valencia; Corpus Ameresco (Albelda y Estellés) con conversaciones de Argentina, Colombia, Cuba, Chile y México; COLA (Jørgensen 2007), con conversaciones de adolescentes de Madrid, Buenos Aires, Santiago de Chile y Managua; C-ORAL-ROM

(Cresti y Moneglia 2005), que entre otros tipos de discurso incluye conversaciones en español; los corpus CALLHOME y CALLFRIEND (Canavan y Zipperlen 1994a, 1996b) de conversaciones telefónicas entre familiares o amigos del área caribeña.

	<i>Corpus de conversaciones para E/LE</i> de la Universidad de Columbia (Ruiz Fajardo 2016)	GestINF (Rodríguez Rosique y Cifuentes Honrubia en línea)	Corpus del Grupo de Investigación de Lingüística Aplicada (COGILA) (Barros 2012)	Nijmegen Corpus of Casual Spanish (NCCSp) (Torreira y Ernestus 2012)	Freiburg SofaTalks (FST) (Pfänder).
Acceso	Libre	Libre	En libro con CD	Requiere solicitud	Requiere solicitud
Lengua	Español (multinacional)	Español (España)	Español (España)	Español (España)	Multilingüe (incluye español multi-nacional)
Nº de conversaciones	33	8	10	22	18
Duración media de las conversaciones	5 minutos aprox.	1 hora	Entre 3 y 21 minutos	90 minutos	1 hora
Duración total	162 minutos	8 horas	134 minutos	30 horas	20 horas
Nº de hablantes por conversación	Entre 2 y 10	4	Entre 2 y 4	2 y un gancho	2 y un gancho
Lugar	Espacios públicos	Laboratorio	Domicilio	Laboratorio	Domicilio
Nº de cámaras	1	3	1	1 + micrófonos de cabeza	1 (no se graba al gancho)
Genero discursivo	Conversación	Conversación. Tema sugerido pero no obligado.	Conversación. Libre.	Conversación. 20 minutos libre secreta + 35 minutos libre + 35 minutos tarea de negociación	Conversación. Libre.
Transcripción	Oral adaptada a estudiante de ELE	Oral de Val.Es.Co.	Oral de Val.Es.Co.	Oral ortográfica	GAT2 aplicada a 5 horas
Secreta	No	No	No	Parcialmente	No

Tabla 10. Corpus de conversaciones en español grabadas en vídeo.

Sin embargo, no existen muchos corpus de conversaciones en español grabadas en vídeo (véase Tabla 10). Hasta hace poco, el único de acceso libre era el *Corpus de conversaciones para E/LE* de la Universidad de Columbia (Ruiz Fajardo 2016), con fines didácticos. A una de ellas se recurrió en un estudio preliminar a este (Cabanes Pérez 2020) para realizar un primer análisis de la estructura de una conversación grabada en vídeo.

En 2022 se ha publicado en línea el corpus GestINF (Rodríguez Rosique y Cifuentes Honrubia; Rodríguez Rosique 2021), que alberga ocho conversaciones de una hora cada una, grabadas con gran calidad de imagen y sonido en laboratorio. Cada diálogo cuenta con cuatro participantes de distintos colectivos de la Universidad de Alicante (estudiantes, profesorado, personal de administración y personal de servicios) porque el objetivo de su proyecto es estudiar el conocimiento compartido comunitario.

Sin acceso libre, se deben mencionar tres corpus de conversaciones en español en vídeo: el Corpus del Grupo de Investigación de Lingüística Aplicada (COGILA) (Barros 2012), el Nijmegen Corpus of Casual Spanish (NCCSp) (Torreira y Ernestus 2012) y el Freiburg SofaTalks (FST) (Pfänder). El corpus COGILA de la Universidad de Granada recoge 10 conversaciones de estudiantes andaluces. La duración de grabación oscila entre los tres y los veintiún minutos y se llevó a cabo en espacios tranquilos y libres de ruido ambiental, con un fuerte predominio de las casas particulares (80% de las grabaciones). Aparecen en grupos de dos, tres o cuatro personas y les une una relación de amistad; en algunos casos, también son compañeros de clase (30%) o de piso (10%).

El corpus NCCSp forma parte de la base de datos The Language Archive del Max Planck Institute for Psycholinguistics. Participan en estas conversaciones casuales 52 hablantes de español durante 30 horas de grabación en vídeo con audio de alta calidad, transcrito ortográficamente en formato xml (Transcriber), TextGrid (Praat) y eaf (ELAN). Todos los hablantes son estudiantes universitarios y crecieron en la Comunidad de Madrid. En cada sesión se grabó durante 90 minutos a un grupo de tres personas del mismo sexo, dos hablantes nativos y un gancho. Los ganchos, ocho durante todo el corpus, reclutan entre sus amigos a los otros participantes y se aseguran de mantener viva la conversación.

El corpus FST (Pfänder), desarrollado en el Seminario de Romanística de la Universidad de Friburgo, reúne más de 200 grabaciones en audio y video en distintas lenguas, entre ellas distintas variedades de español americano. En estos vídeos los informantes están sentados juntos y hablan sobre experiencias compartidas. El moderador tiene en todos los casos una relación personal con ellos y es parte de su comunidad de habla. La cámara se coloca en una posición fija frente a los participantes, de modo que las acciones de ambos son visibles en todo momento.

Tres de estos corpus recogen muestras del español de España (Alicante, Granada, Madrid), mientras que los otros dos, de varios países hispanohablantes: ambos de

Argentina y Colombia; el de Columbia también de Cuba, Perú, Uruguay, EE. UU. y Guinea Ecuatorial; el de FST también de Bolivia e inmigrantes variados en Alemania. Los dos creados desde universidades españolas (COGILA, GestINF) han utilizado las convenciones de transcripción de Val.Es.Co. (Briz y Grupo Val.Es.Co 2002a), mientras que los de Columbia y Nijmegen han optado por la transcripción ortográfica.

Tres de ellos se realizan en espacios naturales donde suele darse una conversación coloquial (COGILA, Columbia, FST), lo cual beneficia a la comodidad de los informantes y a la naturalidad de la muestra. Las conversaciones del de Columbia, siguiendo una metodología similar a Kendon (2002), normalmente son conversaciones ya existentes en las que la investigadora se acerca a los hablantes y les pide permiso para grabarles. No obstante, los otros dos (GestINF y NCCSp), grabados en laboratorio, aseguran mejor calidad de audio. Para ganar naturalidad, el corpus NCCSp graba los primeros 20 minutos de media de manera secreta: se les dice a los participantes que el micrófono del gancho no funciona adecuadamente y este debe dejar la habitación; esto permite grabar la conversación espontánea que surge entre ellos sin que sepan que ya se está grabando. La segunda parte de la grabación consiste en una conversación libre entre el gancho y sus amigos, que dura sobre 35 minutos; y en la tercera parte se les pide a los participantes que elijan tres preguntas de interés general de un listado y que negocien una respuesta en grupo; esta parte también dura 35 minutos. En el resto de los corpus, las conversaciones son enteramente libres, aunque en GestINF a cada grupo se le proporciona dos temas que pueden servir como puntos de partida opcionales.

Por último, en el Anexo 1 observamos las características de otros corpus en otros idiomas de conversaciones grabadas en vídeo creados para el análisis multimodal de la interacción. La primera característica que destacamos es que todos se han realizado en laboratorio, lo que permite, generalmente, mayor cantidad de cámaras para captar la interacción desde varios ángulos, micrófonos de cabeza y otros aparatos de monitoreo. Los corpus holandeses son los más avanzados en esto: el Dutch Face-to-Face⁷⁷ (DF2F) (Hömke et al. 2022) y el Insight Interactive (Brône y Oben 2015) utilizan *eye-trackers* para aportar datos precisos sobre el uso de la mirada y el AliLeGra (Feyaerts et al. 2012),

⁷⁷ Dutch Face-to-Face (DF2F). Cada sesión de grabación constaba de tres fases de 20 minutos: Durante una fase de 20 minutos, los participantes no llevaban los micrófonos montados en la cabeza y el audio se grababa únicamente con un micrófono de techo. Durante una segunda fase de 20 minutos, el audio se grabó con los micrófonos montados en la cabeza, y durante una tercera fase de 20 minutos, el audio se grabó con los micrófonos montados en la cabeza y, además, los participantes llevaban gafas de seguimiento ocular.

recopilado por la KU Leuven y la Karl-Franzens-Universität de Graz, biosensores para relacionar las funciones fisiológicas con las acciones de la interacción.

Suelen participar diadas, a excepción de la tríada con rol play del NOMCO estonio y el grupo de 4 jugando a las cartas del NOMCO finlandés (Navarreta et al. 2011). Además de estas situaciones, generalmente, encontramos conversaciones libres, como en el DF2F y en el IFA Dialog Video (IFADV) (van Son et al. 2008). Algunas conversaciones libres se diseñan específicamente entre desconocidos, como los subcorpus danés y sueco del NOMCO (Navarreta et al. 2011, Paggio y Navarreta 2017, Allwood et al. 2000), o de manera secreta previa a una falsa prueba (AliLeGra). Pero también hay conversaciones orientadas en el Active Listening Corpus (ALICO) (Buschmeier et al. 2014; Malisz et al. 2016), tareas en el MULTISIMO (Koutsombogera y Vogel 2018) o una mezcla de conversaciones libres y orientadas en el Insight Interaction y en el Corpus of Interactional Data (CID) (Bertrand et al. 2008); así como de géneros variados en el NOMCO sueco y en el Corpus de Langue Parlée en Interaction (CLAPI⁷⁸) (Groupe ICOR 2016).

Al contrario que los corpus en español, la mayoría de los corpus en lenguas extranjeras cuentan con algún sistema de anotación de gestos. Un corpus multimodal, en un sentido amplio, es una colección digitalizada de material relacionado con el lenguaje y la comunicación que se basa en más de una modalidad; pero, en un sentido estricto, podríamos exigir que el material audiovisual vaya acompañado de transcripciones y anotaciones o codificaciones basadas en el material (Allwood 2008). Aunque ALICO y MULTISIMO⁷⁹ no sean corpus de conversaciones libres, los hemos incluido en la tabla porque cuentan con un sistema de anotación de risas y gestos.

4.4. Anotación de gestos

Las transcripciones son inherentemente selectivas y parciales. Al imponer el formato escrito a la lengua, se pierde parte de la relación temporal. Sin embargo, es evidente su utilidad para buscar elementos de interés y mostrarlos en formatos escritos como este. Especialmente, con la transcripción de un vídeo se pierde información porque no es posible anotar cada detalle visual con precisión sin perder legibilidad. Al fin y al cabo, percibimos información a través de varias modalidades simultáneamente y les

⁷⁸ Visto 31/08/2021: <<http://clapi.ish-lyon.cnrs.fr/>>

⁷⁹ <https://multisimo.eu/datasets.html>

damos un sentido, pero la palabra escrita es una sola modalidad que se caracteriza por su linealidad y debemos hacer un esfuerzo de procesamiento mayor para relacionar la información temporalmente.

Generalmente, la anotación de los elementos no verbales simultáneos se suele marcar con una o más líneas de cada modalidad paralelas a la transcripción. En este formato partitura, la dirección de izquierda a derecha representa los eventos en el tiempo y la dimensión vertical captura cómo se solapan los elementos en cada punto particular en el tiempo verbal (Bohle 2013:1001).

A menudo, los términos transcripción y anotación se utilizan indistintamente en la bibliografía. Una opción es distinguir la transcripción que describe la forma de enunciados y gestos de la anotación de otras características, por ejemplo, funcionales (Bressem, Ladwig y Müller 2013: 1100). Sin embargo, la transcripción del gesto no tiene el mismo grado de convencionalización que el habla. Si bien las letras no representan todas las variables que pueden aparecer en la producción real, generalmente, sí representa las esenciales. En cambio, según el detalle o granularidad de transcripción del gesto, podríamos hablar de anotación si solo incluyen ciertas características de interés para el objetivo del estudio. Al final, la anotación es producto de un pre-análisis que no se lleva a cabo en las transcripciones verbales, el investigador anota lo que es importante y no otros contenidos de la escena grabada dependiendo de la pregunta de investigación (Deppermann 2013: 3). “Ya se está interpretando los datos a medida que se transcriben” (Moore y Llompart 2017: 427). La transcripción multimodal es por tanto altamente reflexiva e interpretativa (Mondada 2007b), es un proceso hermenéutico mediante el cual el anotador reconstruye los actos en un sistema escrito (Bohle 2013: 993-994): al seleccionar los elementos que se incluyen en la anotación, los anotadores reducen los posibles análisis e interpretaciones de los datos. Por tanto, la función de la anotación es pivotante: sirve como base para el análisis, pero también es el resultado de un análisis.

Los sistemas de categorías creados para este fin pueden ser más descriptivos/globales o más interpretativos/restrictivos. No es factible codificar cada mínimo movimiento físico porque es una labor intensa y poco práctica. Cuanto más detalle, mayor coste de anotación y menor nivel de abstracción (Kipp et al. 2007). Por ejemplo, el sistema FACS (Facial Action Coding System) de Ekman y Friesen (1978), que describe cada detalle de la expresión facial teniendo en cuenta la acción de 58

músculos tiene el objetivo de construir un sistema libre de teoría; pero a la hora de aplicar una teoría se deben seleccionar los movimientos que son útiles para la investigación.

El sistema de kinégrafos de Birdwhistell (1952) también es muy completo. Presenta una codificación formada por 400 elementos que combinan dibujos, letras y números para representar movimientos de ocho partes del cuerpo: cabeza, cara, tronco, hombros/brazo/muñeca, manos/dedos, cadera/pierna/rodilla, pie/andar, cuello. Como se observa en la Figura 6, es un sistema complejo que requiere entrenamiento del lector para interpretarlo.

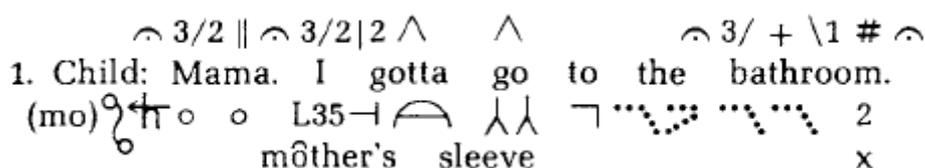


Fig. 6. Ejemplo de transcripción con los kinégrafos de Birdwhistell (1970: 283).

Por otro lado, una transcripción puramente interpretativa puede suprimir detalles que podrían contribuir al análisis. Desde el AC, Jefferson (1984) solamente anota el comportamiento corporal si influye de forma evidente en las orientaciones verbales y comunicativas de los hablantes y los destinatarios y lo hace en líneas de comentario con términos rudimentarios. Sin embargo, la terminología cotidiana no es adecuada para el análisis científico por su falta de precisión (Ehlich 2013: 654). El sistema HIAT (HalbInterpretative ArbeitsTranskriptionen) de Ehlich y Rehbein (1979a) también es rudimentario/cotidiano en su forma o función, aunque describa los movimientos mejor: parte del cuerpo, cualidad y acciones complejas (Bressem 2013a: 1044).

Por último, incluimos como ejemplo de codificación rudimentaria una que parte del sistema de transcripción verbal de Val.Es.Co. (Briz y Grupo Val.Es.Co. 2002): el Corpus PerLA (Percepción y Lenguaje Afásico) (Gallardo y Sanmartín 2005), que añade información multimodal en letra versalita y entre paréntesis junto a la emisión verbal. Esta información es una descripción breve y no sistemática del gesto, a veces centrada en la función y a veces en la forma (NEGACIÓN, MOVIMIENTO ASCENDENTE CON LA MANO HORIZONTAL). También utiliza para gestos repetitivos una R en superíndice y para la mirada una flecha y la inicial del interlocutor que se mira o hacia dónde se dirige la mirada (⇒E) (Gallardo y Sanmartín 2005: 49). Nuestra codificación para el sistema de

Val.Es.Co. (§2.3.4) pretende mayor sistematicidad en la codificación y el alineamiento temporal.

- 0082 I: que no sé/ ((en estas/ aní que)) (SE SEÑALA EL OÍDO)
- 0083 H: ¿pero de qué estás– de qué– de qué estás hablándonos?
- 0084 I: claro/ claro que mee– porque yo noo–/ no puedoo (LA MANO SUBE POR EL ESTERNÓN Y SALE DESDE LA BARBILLA)/ [((pero))=]
- 0085 H: [pe– pero]
- I: = no me–/ quan era joveneta– quan me vaa vore a mii/ quan ((xx)) vore/ els dos/ ella ha segut de tots
- 0086 H: ¿cuando te dio el infarto↓ te refieres?
- 0087 I: síi
- 0088 H: vale
- 0089 I: síi// ((ellaa–// on estava jo?))
- 0090 H: estabas en el Clínicoo/ te pasó aquí↓ en casa
- 0091 I: ((¿dónde yo?// ¿yoo? ¿dónde?))
- 0092 H: pero fui– te llevamos al Clínicoo/ ¿no te acuerdas?
- 0093 I: ((¿eso yoo?)) (EXPRESIÓN DE INCREDELIDAD O DISCONFORMIDAD)
- [Corpus PerLa, JFG: 0082-0093]

(Gallardo y Sanmartín 2005: 49-50)

La metodología MFD (Bavelas 1994) anota la función de los gestos directamente sobre el fragmento de vídeo, sin describir su forma. No significa que la forma no sea un criterio para identificar la función, sino que se ve directamente en el vídeo y no se dedica tiempo a codificarla. A este método Bohle (2013) lo denomina *evaluación directa*.

Respecto a su temporalidad y fases, algunos sistemas de anotación incluyen los tres niveles de gesto -*gesture phrase*, *gesture phase* y *gesture unit*- porque entre sus objetos de estudio está la sincronización del habla con el gesto: el de McNeill (1992), el Linguistic Annotation System for Gestures (LASG) (Bressem et al. 2013), el CoGest (Conversational Gesture Transcription system) (Gut et al. 2002), MultiModal MultiDimensional (M3D) (Rohrer et al. 2020). El nivel de detalle depende de la pregunta de investigación.

Algunas codificaciones optan por una perspectiva “fonológica”, es decir, basada en los componentes articulatorios significativos de las lenguas de signos, como LASG (Bressem et al. 2013) y CoGest (Gut et al. 2002). En cambio, otras describen las

características formales que son significantes para el estudio de funciones comunicativas interpersonales, como la codificación MUMIN (Nordic Network for MultiModal Interfaces) (Allwood et al. 2007).

Nuestro sistema (§5.5.3.) se basa principalmente en el de MUMIN porque también estamos interesados en las funciones interpersonales, pero, además, utiliza códigos de LASG para describir con mayor detalle los gestos manuales. Describimos estos dos sistemas de codificación a continuación.

4.4.1. Sistema de anotación MUMIN

El código de MUMIN (Allwood 2001, Allwood et al. 2005, Allwood y Cerrato 2003, Allwood et al. 2007) está orientado hacia las funciones comunicativas y es el utilizado por todo el corpus NOMCO (<http://sskkii.gu.se/nomco/>), así como otros materiales en griego (Koutsombogera et al. 2008) y en japonés (Jokinen et al. 2009). La anotación de gestos del corpus CID (Blache et al. 2017) también se basa en MUMIN, pero incluye características de Kipp et al. (2007) (cuyo propósito es generar y animar automáticamente gestos manuales de personajes virtuales) y de McNeill (1992) e incluye códigos para otras partes del cuerpo (brazos, hombros, tronco, piernas).

Explicamos los componentes de la descripción formal de MUMIN a partir de Allwood et al. (2007) y la información actualizada de la versión 4 del manual de anotación de la página web de MUMIN (<https://www.cst.dk/mumin/resources.html>). Tras las actualizaciones del esquema en 2009, se describe la forma y la función comunicativa de la expresión facial, de la mirada, de los movimientos de cabeza, de los gestos manuales y de las posturas corporales. Respecto a los gestos manuales, en la cuarta versión de MUMIN se incluyeron atributos sobre la orientación y la forma de la mano que no estaban en Allwood et al. (2007).

La codificación de la forma realiza a través de 21 atributos que describen el comportamiento de distintas partes del cuerpo (cara, cejas, ojos, mirada, boca, cabeza, cuerpo, hombros, manos). Algunas partes requieren más atributos que otras y para cada atributo hay una serie de valores codificados que los creadores han perfeccionado con su experiencia de transcripción (Tabla 11).

Los valores en los que se incluye “Other” indican que ninguno de los otros valores es apropiado y debe añadirse un comentario. La línea “Eyes” no es un atributo

Atributo	Valores
Face	Smile, Laughter, Scowl, FaceOther
Eyebrows	Frown, Raise, BrowsOther
Eyes	X-open, CloseBoth, CloseOne, CloseRepeat, EyesOther
GazeDirection	GazeForward, GazeBackward, GazeUp, GazeDown, GazeSide, GazeDirectionOther
GazeInterlocutor	ToInterlocutor, AwayFromInterlocutor
MouthOpenness	OpenMouth, CloseMouth
MouthLips	CornersUp, CornersDown, Protruded, Retracted, LipsOther
HeadMovement ⁸⁰	Nod, Jerk, HeadBackward, HeadForward, Tilt, SideTurn, Waggle, HeadOther ⁸¹
HeadRepetition	Single, Repeated
BodyInterlocutor	BodyToInterlocutor, BodyAwayInterlocutor
BodyDirection	BodyForward, BodyBackward, BodyUp, BodyDown, BodySide, BodyDirectionOther
Shoulders	Shrug, ShouldersOther
Handedness	BothHands, SingleHand
Palm	PalmOpen, PalmClosed, PalmOther
PalmPos	PalmUp, PalmDown, PalmSide, PalmPosOther
Fingers	IndexExtended, ThumbExtended, AllFingersExtended, FingersOther
Amplitude	Centre, Periphery, AmplitudeOther
TrajectoryRightHand	RightHandForward, RightHandBackward, RightHandSide, RightHandUp, RightHandDown, RightHandComplex, RightHandOther
TrajectoryLeftHand	LeftHandForward, LeftHandBackward, LeftHandSide, LeftHandUp, LeftHandDown, LeftHandComplex, LeftHandOther
HandRepetition	Single, Repeated

Tabla 11. Sistema de anotación formal de los gestos de MUMIN

de la mirada, sino que se refiere a expresiones que se realizan con la obertura o cierre de los ojos. La postura corporal se incluyó en la versión 4 con el atributo “BodyDirection”, pero en Navarretta y Paggio (2013) también se introdujo “BodyInterlocutor”. La aparición del valor “BodyDirectionOther” en Navarretta y Paggio (2017) fue de un 27%,

⁸⁰ El corpus ALICO (Buschmeier et al. 2014) se centra en la anotación multimodal del *feedback* (Malisz et al. 2016) mediante movimientos de cabeza. Utilizan una codificación similar: nod, jerk, tilt, shake, protrusión, retraction, turn, bobble, slide, shift, waggle.

⁸¹ En Allwood et al. (2007) también aparecen: Down, Down-Repeated, BackUp, BackUp-Repeated, BackUp-Slow, Forward, Back, Side-Tilt, Side-Tilt-Repeated, Side-Turn, Side-Turn-Repeated, Waggle, Other.

por lo que sugieren que los valores de este atributo necesitan más desarrollo. En Navarretta y Paggio (2013) también se incluyó el encogimiento de hombros (Shrug, ShouldersOther).

Solo se anotan las expresiones con valor comunicativo específico, por ejemplo, las que dan retroalimentación o se utilizan en la gestión del turno. Por tanto, los adaptadores no se anotan y de la expresión facial no se anotan los movimientos que forman parte de la expresión “neutra” (el pestañeo natural de los ojos, los movimientos de la boca debidos a la articulación del habla o los causados por una luz molesta). Además, cuando ciertas partes permanecen neutrales mientras otras se mueven, solo se anotan las que muestran movimiento o cambio de estado.

Además de la codificación de la forma, MUMIN incluye atributos y valores para las funciones interpersonales, para la tipología semántica del gesto -basada en la teoría de Peirce: *indexical, iconic, symbolic*- y para la relación con el contenido verbal. Aunque no partimos de la codificación funcional de MUMIN, la añadimos como alternativa (Tabla 12). Contempla distintas maneras de dar o pedir retroalimentación, expresar emociones o actitudes, gestionar el turno (tomar, aceptar, ceder, ofrecer, terminar, mantener) y la secuencia (iniciarla, retomarla, continuarla, cerrarla).

Atributo	Valores
FeedbackBasic	FeedbackGive, FeedbackElicit
FeedbackUnderstanding	Understand, NonUnderstand
Emotion/Attitude	Happy, Sad, Surprised, Disgusted, Angry, Frightened, Certain, Uncertain, Interested, Uninterested, Disappointed, Satisfied, Other
Turn	TurnTake, TurnAccept, TurnYield, TurnElicit, TurnComplete, TurnHold
Sequencing	SeqOpen, SeqResume, SeqContinue, SeqClose

Tabla 12. Sistema de anotación funcional de los gestos de MUMIN

4.4.2. Sistema de anotación LASG

El sistema de anotación LASG (Linguistic Annotation System for Gestures) (Bressem 2013b, Bressem, Ladwig y Müller 2013) sigue un enfoque lingüístico-semiótico que asume una separación heurística entre la forma, el significado y la función. Se centra en un análisis sistemático de la forma del gesto, que es independiente del enunciado para evitar la descripción a través de paráfrasis del significado basadas en la intuición.

Tanto los sonidos de las lenguas como los movimientos corporales de los gestos son infinitos y su anotación requiere de cierta abstracción. Al igual que la transcripción fonética, la representación articulatoria de las manos es moderadamente estrecha. La forma se describe mediante los parámetros de las lenguas de signos (Stokoe 1960): forma de la mano, orientación, movimiento, posición en el espacio (Bressem 2013b: 1083). Esto se debe a que, de la misma manera que las modificaciones de estos parámetros afectan en la diferenciación de signos, también crean diferencias de significado en los gestos. Esta anotación propone una terminología lo más objetiva posible que capture cualidades fisiológicas específicas, evitando la paráfrasis por su significado y función. Es posible incorporar nuevas formas para expandirlo (Bressem 2013b: 1083).

Respecto a la configuración de la mano, hay cuatro categorías básicas: *fist*, *flat hand*, *single fingers*, *combination of fingers* (Fig. 7). Si uno o más dedos tienen una forma concreta, se indica mediante su numeración del 1 al 5 (1-pulgar, 5-meñique) y la forma (Fig. 8, Fig. 9)



Fig. 7. Cuatro categorías básicas de configuración de la mano (Bressem 2013b: 1085).

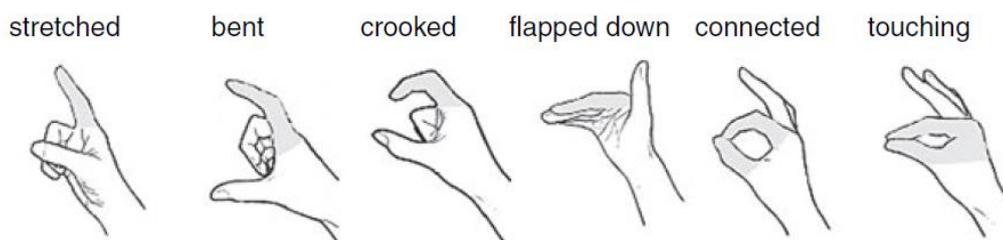


Fig.8. Resumen de las seis formas de los dígitos (Bressem 2013b: 1085).

Combination of fingers						
1+2 connected	1+3 connected	1+2 crooked	1+2 bent	1-5 crooked	1-5 bent	1-5 spread bent
2-5 flapped down	2-5 flapped down 1 stretched	2-5 bent	1-5 touching	1+5 connected	1+2 touching	

Fig.9. Ejemplos de formas de la mano con combinación de dedos (Bressem 2013b: 1086).

La orientación del gesto (Bressem 2013b: 1087) se indica, por un lado, mediante cuatro ángulos para la palma: *palm up (PU)*, *palm down (PD)*, *palm lateral (PL)*, *palm vertical (PV)*, además de diagonales (di); y, por otro lado, mediante el espacio del gesto: *towards center (TC)*, *away center (AC)*, *towards body (TB)*, *away body (AB)* (Fig. 10). Además, se puede especificar la orientación de los dedos si es necesario (*fingers down*).

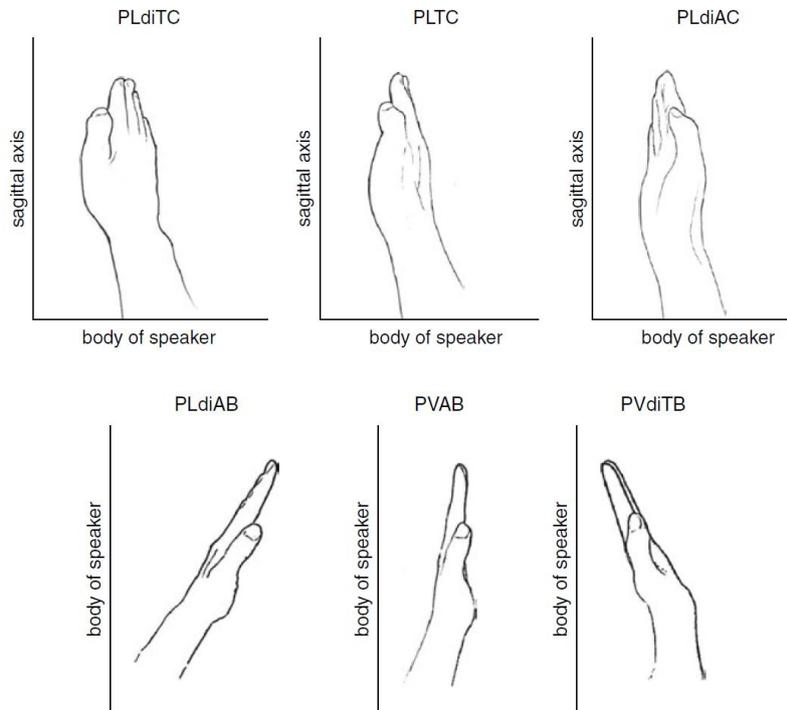


Fig. 10. Ejemplos de orientación del gesto (Bressem 2013b: 1087).

El movimiento del gesto (Bressem 2013b: 1088-1089) se describe mediante la forma que traza la mano en el aire (Figura 11) y su dirección en el eje horizontal, vertical, diagonal o sagital (o hacia/contra las agujas del reloj, en el caso de los movimientos

circulares). Además, hay movimientos específicos de la muñeca (Figura 12), de un dedo (recto, arqueado y circular) o de todos los dedos a la vez (golpe, aleteo, agarre y cierre).

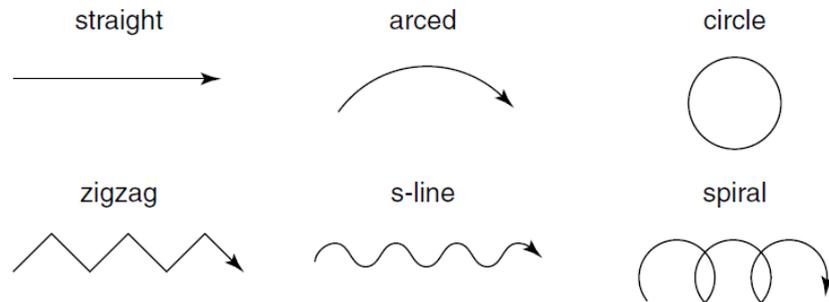


Fig. 11 Tipos básicos de movimientos (Bressem 2013b: 1088).

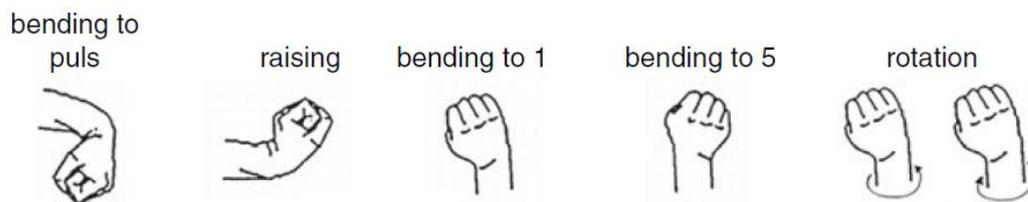


Fig. 12. Tipos de movimiento de muñeca (Prillwitz et al 1989 [Bressem 2013b:1089])

También se pueden especificar las cualidades de movimiento (Bressem 2013b: 1090): tamaño, velocidad y flujo de movimiento (es decir, si está acentuado); que se relacionan con la entonación verbal y la acentuación de segmentos.

Por último, para las posiciones del gesto en el espacio del hablante se toma la división en sectores de McNeill (1992: 89) usando cuadros concéntricos (centro-centro, centro, periferia y periferia extrema) y el eje vertical y horizontal (Figura 13).

Tanto la transcripción de la forma como la anotación de su motivación formal se lleva a cabo libre de contexto, sin escuchar los enunciados hablados para que no influyan en la interpretación. La motivación de la forma se detecta mediante principios cognitivo-semióticos (modos de representación) y la base del gesto en patrones motores, esquemas de imagen y acciones (Bressem et al. 2013: 110). Es un paso previo que ayudará a averiguar su significado y función en relación con el habla.

La base del significado gestual es la mimesis. Hay dos modos para vincular los medios de articulación y los objetos referidos: actuar, de manera que las manos cumplen la acción, y representar, cuando las manos representan eventos u objetos del mundo (Bre-

negrita para el *stroke*. La segunda línea tiene en cuenta la progresión lineal simultánea de las fases gestuales individuales según Kendon (1980): P = preparation, S = stroke (numerados), R = recovery, rp = rest-position, y otros estados intermedios. La tercera línea describe la forma de los *strokes* individuales.

4.4.3. Sistema de transcripción multimodal de Mondada

Finalmente, explicamos el sistema de Mondada (2018)⁸², que no establece unos códigos para cada movimiento corporal, sino que propone una manera de organizar la anotación en la transcripción en papel para poder relacionar los distintos modos que se sincronizan en la interacción. Es aplicable a cualquier comportamiento corporal (gesto, mirada, postura, manipulación de objetos, etc.) y sistema de codificación.

Esta convención se basa en tres principios: la identificación de los hablantes, la representación de la temporalidad y la caracterización de la acción visible. En cualquier transcripción dialógica debe estar claro a qué hablante pertenece cada emisión, pero esto es un reto en situaciones donde las acciones visibles de varios participantes confluyen. Es importante que la relación temporal entre las actividades multimodales quede marcada mediante símbolos idénticos para indicar el inicio y el final de la acción (Figura 15) y así analizar las relaciones que se establecen con la emisión verbal y entre las acciones. Asimismo, la descripción de la acción visible debe ser breve y coherente para que permitan que se entiendan fácilmente y de forma unívoca. Se pueden utilizar abreviaturas, capturas de pantalla o describir la acción con más detalle en el comentario analítico, si es necesario.

1 CUS +EH s'il vous plaît+ un bout d'gruyère
 eh please a piece of Gruyère
 +points at gruyère+

Fig. 15. Ejemplo de alineación de habla y gesto en la transcripción de Mondada.

El símbolo que marca el inicio y final de las acciones es diferente para cada hablante, de manera que no se confundan sus acciones. Si hay distintas acciones de un mismo hablante se anotan en distintas líneas y se crean más símbolos distinguirlos. Si el que realiza la acción es el mismo que habla, no se anota al participante que la realiza, pero

⁸² A esta publicación le han sucedido actualizaciones que la investigadora publica en su propia página web. La más reciente es de 2022:
https://www.lorenzamondada.net/files/ugd/ba0dbb_3978d2a34cf44376adb7a341975d23aa.pdf
 (Visitada: 18/05/2023).

si es otro participante, sí. En caso de turnos solapados, todas las anotaciones de las conductas visibles se colocan después del habla solapada.

Para que la acción se alinee con el texto, se utiliza línea discontinua (----) para que los símbolos delimitadores coincidan. Si la acción continúa en líneas siguientes, se indica con una flecha (--->) y en la siguiente línea se repite hasta alcanzar el símbolo (---->*). La doble flecha indica si la acción descrita comienza antes del comienzo del extracto (>>) o continúa después del final del extracto (--->>).

También se puede marcar en el texto la referencia a un fotograma que ilustre la acción, utilizando la almohadilla (#) para marcar su posición en el turno (Figura 16).

```

1  SEL      #eh::: piensa* que# eso que es* muy líquido#
          eh::: be aware of this that it's very liquid
          *takes w RH*2H around-----*palpates->>
fig    #fig.1          #fig.2          fig.3#
    
```



Fig. 16. Alineación de fotogramas con la transcripción.

Si una acción visible se produce en el silencio o comienza en medio de una pausa, debe indicarse el tiempo que dura la acción (Figura 17).

```

2      (0.5) ± (1.5)±
      cus      ±points±
3  CUS      ah probier ich so `n Camembert
          oh I try PRT a Camembert
    
```

Fig. 17. Alineación de gestos en fragmentos sin habla.

Es posible distinguir las fases del gesto, si es el objetivo, utilizando pequeños puntos para la fase de preparación (.....) y comas para la de recuperación (,,,,,).

4.5. Recapitulación

De esta breve revisión de la metodología aplicada al estudio de los gestos, concluimos que para nuestro trabajo es necesario reunir un corpus de conversaciones coloquiales que mantengan sus condiciones naturales y cotidianas para poder analizar la estructura dialógica real de este tipo de discurso.

Los corpus en vídeo en español cumplen gran parte de las características de una conversación coloquial como la finalidad interpersonal, la relación vivencial o la temática no especializada. El corpus de ELE de Columbia (Ruiz Fajardo 2016), el COGILA (Barros 2012) y el FST (Pfänder), además, se han grabado en espacios cotidianos que promueven la coloquialidad; al contrario que los corpus extranjeros, grabados en laboratorio, pero con equipos de monitoreo de mayor calidad.

Ningún corpus en español ha anotado los gestos con un sistema de anotación (frente a otros corpus extranjeros), por lo que es una tarea que debe llevar a cabo el propio investigador. Para el análisis es importante una anotación formal sistemática, que no recurra a perífrasis de significado. En §5.5.3. se adaptan los sistemas mostrados (MUMIN y LASG) a un programa de anotación y a la transcripción en papel de Val.Es.Co. utilizando estrategias de Mondada (2018).

BLOQUE II

CAPÍTULO 5. METODOLOGÍA

5.1. OBJETIVOS Y METODOLOGÍA.....	211
5.2. DISEÑO DEL CORPUS	214
5.2.1. PRESERVACIÓN DE LOS RANGOS SITUACIONALES DE LA CONVERSACIÓN Y CRITERIOS ÉTICOS	214
5.2.2. CARACTERÍSTICAS DE LAS CONVERSACIONES	216
5.2.3. EQUIPO DE GRABACIÓN.....	218
5.3. TRANSCRIPCIÓN Y SEGMENTACIÓN DE UNIDADES CONVERSACIONALES.....	220
5.3.1. TRANSCRIPCIÓN.....	221
5.3.2. SEGMENTACIÓN	226
5.4. ANÁLISIS DE CAMBIOS EN LA SEGMENTACIÓN	229
5.4.1. VARIABLE CAMBIO	230
5.4.2. VARIABLE CAUSA	232
5.4.3. CRITERIOS DE SEGMENTACIÓN DE GESTOS	240
5.4.4. ANOTACIÓN DE CAMBIOS EN LA SEGMENTACIÓN	249
5.5. ANÁLISIS DE LOS GESTOS.....	252
5.5.1. VARIABLE POSICIÓN	257
5.5.2. VARIABLE FORMA.....	258
5.5.3. ANOTACIÓN DE GESTOS: LA FORMA	259
5.5.4. VARIABLE FUNCIÓN	271
5.5.5. ANOTACIÓN DE GESTOS: LA FUNCIÓN	273

5.1. Objetivos y metodología

Nuestro objetivo es describir la manera en la que la multimodalidad afecta a la estructura dialógica y a la dimensión social de conversaciones coloquiales en español. El modelo de segmentación de la conversación de Val.Es.Co. (§2) es el que nos permite sistematizar los efectos de la multimodalidad en la conversación. Nuestro foco se encuentra en el nivel dialógico, en la unidad estructural *intervención* y en la unidad social *turno*, y en los elementos multimodales que sirven como criterio para segmentar estas unidades.

Para comprobar qué elementos multimodales afectan a la segmentación, nuestro análisis se basa en la comparativa entre la segmentación de conversaciones grabadas en audio y la de las mismas conversaciones grabadas en vídeo. La única diferencia entre ambos materiales es la información de cada grupo, la únicamente auditiva y la que cuenta con audio e imagen. De esta manera, podemos observar qué información visual es la que genera otra interpretación de lo que sucede en la conversación a nivel dialógico.

A partir de la segmentación del material, se realizan dos análisis diferenciados:

1. El análisis de los cambios estructurales y sociales que se producen al incorporar los elementos visuales: se comprueba que la serie de elementos visuales propuestos en este capítulo producen cambios en la segmentación dialógica; se evalúa cuáles tienen mayor efecto en la segmentación y qué tipo de efecto; se describen los efectos de los elementos multimodales en la segmentación estructural (creación, modificación o eliminación de intervenciones) y social (creación, modificación o eliminación de turnos); se pretende demostrar la pertinencia de los elementos visuales en la segmentación dialógica mediante la relación entre los cambios producidos entre la primera y la segunda segmentación (sin y con vídeo).
2. El análisis de los gestos que afectan a la segmentación dialógica: se describen las características funcionales de los gestos y se relacionan con su forma (parte del cuerpo y su configuración) y su posición en la estructura de la intervención; para ello, se propone una anotación formal de los gestos para el sistema de transcripción de Val.Es.Co. (Briz y Grupo Val.Es.Co 2002a), tanto en papel como en programas informáticos de anotación de vídeo.

A través de los resultados metodológicos y analíticos, este estudio exploratorio pretende dar pautas para futuros estudios multimodales mediante las unidades conversacionales de Val.Es.Co.

Para realizar este estudio, seguimos las siguientes fases de trabajo, que se corresponden con el orden de apartados de este capítulo.

1. Creación del corpus de conversaciones (§5.2)
2. Transcripción y segmentación (§5.3)
 - i. Segmentación oral (audio sin imagen)
 1. Transcripción verbal y de sonidos
 2. Segmentación en intervenciones y turnos
 - ii. Segmentación visual (audio con imagen)
 1. Transcripción verbal actualizada
 2. Segmentación en intervenciones y turnos, teniendo en cuenta todos los elementos visuales
3. Análisis 1: Causas de cambio entre la segmentación verbal y visual (§5.4)
 - i. Elementos visuales causantes de cambio
 - ii. Segmentación de gestos independientes y coverbales
 - iii. Anotación de cambios
 - iv. Análisis
4. Análisis 2: Forma y función de los gestos independientes y coverbales (§5.5)
 - i. Anotación de forma
 - ii. Anotación de macrofunción
 - iii. Anotación de función específica
 - iv. Análisis

Aunque en §4 se han mostrado algunos corpus de conversaciones en español grabadas en vídeo, en este estudio se han grabado cinco conversaciones, adaptando la metodología de obtención de conversaciones de Val.Es.Co. (Pons Bordería 2022) a la grabación en vídeo, para asegurar que cumplieran los criterios de Val.Es.Co. de una conversación coloquial natural (§5.2).

Tras la obtención del corpus, tanto la transcripción como la segmentación (§5.3) se realiza en el programa ELAN (<https://archive.mpi.nl/tla/elan/download>), producto del Instituto Max Planck de Sicolingüística en Nijmegen, utilizando únicamente la información auditiva (verbal y paralingüística). Después, se incorpora el vídeo para completar el análisis con los datos visuales. Se revisa la transcripción verbal, pues puede haber cambios en ella por la lectura de labios, por el sentido de los gestos o por la atribución de emisiones al hablante equivocado. Entonces, se vuelven a segmentar las categorías conversacionales, esta vez contando con todos los elementos visuales.

Para realizar el primer análisis (§5.4), en ELAN se anotan los elementos visuales que causan cambios entre la segmentación verbal y la visual. En el caso de los gestos, estos se segmentan temporalmente, se identifican como coverbales o independientes y los independientes se categorizan por el tipo de intervención que desempeñan. En el análisis de los datos, se tienen en cuenta los cambios producidos en la dimensión estructural y en la social y los elementos visuales que los producen.

En ELAN se anota la forma de los gestos, es decir, la parte del cuerpo con la que se realiza y su configuración, y se identifican las funciones que realizan, así en el segundo análisis (§5.5), se relaciona la posición del gesto con su función y su forma y es posible proponer una tipología de gestos que frecuentemente afecta a la segmentación dialógica.

Los datos que se utilizan en cada análisis se obtienen del archivo de ELAN, que deben anotarse de manera sistemática siguiendo un esquema. Mediante el esquema de anotación, primero, se identifica el tipo de información que cada dominio debe codificar y, segundo, se elige la forma de codificar la información de forma homogénea, de manera que permita la alineación de los dominios (Blache et al. 2017). Nuestras líneas de ELAN agrupan diferentes cometidos: 1) la transcripción, 2) la segmentación en unidades conversacionales, 3) la anotación de los cambios en la segmentación, 4) la anotación formal de los gestos, 5) las funciones pragmáticas de los gestos. Para cada dominio se requiere una serie de líneas (atributo) en la que se anotan los datos (valor), algunos codificados y otros no. En este capítulo se explican con más detalle las fases de trabajo y la incorporación de datos en ELAN, así que la Tabla 12, que presenta los dominios, pero no los atributos y los valores, se completa a lo largo de este.

DOMINIO	ATRIBUTO	VALOR
Transcripción oral		
Segmentación oral		
Transcripción visual		
Segmentación visual		
Análisis de cambios en seg. visual		
Análisis de gestos: anotación formal		
Análisis de gestos: anotación funcional		

Tabla 12. Esquema de anotación: dominios.

5.2. Diseño del corpus

En este apartado, se presentan las bases metodológicas para el diseño del corpus de trabajo. En primer lugar, se justifica la necesidad de crear un corpus de conversaciones coloquiales en vídeo, diseñado para este trabajo (§5.2.1). En segundo lugar, se presenta el diseño del corpus, el equipo técnico utilizado y los informantes (§5.2.2). Por último, se presentan las características de cada grabación en forma de tabla resumen (§5.3.2).

5.2.1. Preservación de los rasgos situacionales de la conversación y criterios éticos

En esta investigación, centrada en la estructura de la conversación coloquial, es importante que los datos sean lo más cercanos posibles a la conversación coloquial prototípica, descrita en §2.2. Algunos de los corpus en vídeo en español mostrados en el capítulo 4, grabados en espacios cotidianos, cumplen gran parte de las características de una conversación coloquial; sin embargo, el conocimiento de los informantes de que están siendo grabados es un factor que puede afectar a su comportamiento: es la Paradoja del Observador (Labov 1983: 265-266). Por tanto, “el lenguaje más informal es el más difícil de observar (...) se produce un cambio en su lenguaje hacia un estilo más formal” (Stubbs 1987: 218), que no podemos evitar. Resulta interesante la reflexión de Stubbs (1987: 218): puede que estemos investigando cómo hablan las personas cuando se sienten incómodas. Para preservar la naturalidad de los datos, la situación ideal es que los informantes desconozcan que se les graba, pero ocultar el material de grabación genera problemas éticos y afecta a la calidad de los datos porque no pueden utilizarse cámaras y micrófonos visibles. Frente a esta problemática, que es común en las investigaciones que requieren la grabación de material social auténtico, existe la idea consensuada de que los participantes no están constantemente pendientes de la presencia de la cámara (Stubbs 1987) y de que según se desarrolla la conversación pierden la inhibición que mostraban al inicio de la grabación (Holler 2013: 838). Por tanto, según esta perspectiva, el material grabado de esta forma es igual de rico analíticamente y de revelador de las normas sociales (Mondada 2013c). Una medida para favorecer ese olvido de la grabación y evitar la paradoja es lograr un ambiente natural para los participantes al interactuar con personas que ya conocen (Stubbs 1987: 219). Además, este tipo de grupos es el idóneo para que se dé una conversación informal, pues una de sus características es la familiaridad entre los participantes (Briz 1998: 41). El corpus NCCSp (Torreira y Ernestus 2012) es el único en

español que cuenta con una porción grabada de forma secreta, pero el espacio de laboratorio es descoloquializador.

Aunque generalmente se suele ceder a la practicidad de una grabación consciente durante su desarrollo, el Grupo Val.Es.Co. ha creado para sus corpus (Val.Es.Co. y Ameresco) un protocolo para grabar conversaciones en audio de manera secreta (Pons 2022). Para evitar el problema ético, el gancho informa a los hablantes de que en las próximas semanas pueden ser grabados, de manera que dan su consentimiento previo, pero en el momento de la grabación no son conscientes de esta. Los participantes firman, entonces, dos autorizaciones, una previa a la grabación y otra posterior para que validen la cesión del material tras haberlo escuchado (Anexo II).

Al no existir un corpus de conversaciones coloquiales reales secretas en español grabadas en vídeo, necesario para cumplir el objetivo de este trabajo, hemos creado uno propio. El mismo protocolo de grabación secreta de Val.Es.Co. se ha aplicado en esta investigación, pero informando de que la grabación es audiovisual.

La grabación en vídeo suma la complicación práctica de ocultar correctamente las cámaras para que los participantes no las vean y que aun así el ángulo de captura permita abarcar a todos ellos. Esto reduce las oportunidades de reunir de forma natural a los informantes en un lugar adecuado sin generar sospechas, pues para la grabación de audio basta con ocultar una grabadora o un teléfono inteligente para mantener el secreto porque se integra de forma natural en el ambiente⁸³. De las cinco conversaciones, dos se han grabado en la casa de la investigadora, así que en una de ellas se pudieron colocar dos cámaras antes de la llegada de los invitados, mientras que en la otra no fue necesario por la disposición de los hablantes. Las otras tres se han grabado en casas de familiares o un bar y son el resultado de varios intentos de colocar la cámara con disimulo y en un ángulo apropiado.

Debido a estas complicaciones prácticas, la investigadora es participante en todas las conversaciones⁸⁴. La existencia de un gancho, es decir, de una persona que inicia la grabación y se sabe grabada, compromete la espontaneidad de sus emisiones (Holler 2013: 848); más aún cuando esta persona es la investigadora, que conoce las hipótesis del

⁸³ Estos dispositivos no llaman la atención de los informantes y los audios resultantes son de buena calidad en condiciones ambientales adecuadas (Moore y Llompart 2017).

⁸⁴ Para obtener grabaciones de conversaciones coloquiales es común recurrir a interacciones de familiares y amigos del investigador; ya a principios de 1970 Charles y Candy Goodwin filmaron conversaciones de cenas y otros encuentros sociales (50 horas), que se describieron y analizaron en Goodwin (1981, 1986).

experimento y puede manipular consciente o inconscientemente su propio comportamiento y el de los demás participantes.

Se diseñó un protocolo para que conocidos de la investigadora pudieran grabar a personas de su círculo cercano (Anexo III), pero estas grabaciones no se consiguieron por rechazo a participar o la dificultad encontrada en ocultar las cámaras. La solución para evitar la contaminación de los datos debido a la participación de la investigadora es separar del análisis sus gestos y los cambios que estos han producido en la segmentación. Además, al aparecer en todas las conversaciones, restaría valor estadístico al estudio.

5.2.2. Características de las conversaciones

La relación familiar o de amistad entre los participantes ha facilitado los rasgos coloquializadores del encuentro: relación de igualdad entre los interlocutores, relación vivencial de saberes y experiencias compartidas, marco discursivo familiar y temática no especializada (Briz 1998: 41-42). Asimismo, las localizaciones son espacios conocidos en los que los participantes están cómodos: sus hogares o, en el caso de la conversación 4 (Conv4), el bar donde trabaja una de las informantes. Ni los temas, ni los roles, ni los turnos están predeterminados y las conversaciones se han desarrollado como una conversación coloquial prototípica más de estos familiares o amigos, pues se han grabado de manera secreta.

Si bien el número de hablantes ideal es tres para las conversaciones grabadas en audio del Corpus Val.Es.Co., este corpus cuenta con conversaciones de tres y cuatro participantes, pues, gracias al vídeo, los diálogos secundarios no dificultan la identificación de los hablantes y, además, permiten que, mientras el “hablante principal” emite su turno, los otros interlocutores se miren, se susurren o utilicen gestos. En secuencias de historia o broma, especialmente, el hablante busca ampliar el foco de interacción a más de un receptor, por lo que se requiere de más trabajo interaccional para mantener los marcos de participación (Stivers 2021). Una conversación de dos hablantes no da lugar a una gran complejidad de estructuras y, además, como la investigadora es el gancho, el hecho de que haya más de dos hablantes evita que influya demasiado en la estructura de la conversación.

Al crear un corpus propio, no es posible una muestra de hablantes extensa. En el corpus participan once informantes, además del gancho, repartidos en cinco conversaciones de entre 16 y 36 minutos (131 minutos en total). Restringir las variables en un grupo pequeño de informantes es imposible, pues sin una cantidad estadísticamente

significativa, no se pueden controlar. Es decir, no se sabe si un fenómeno ocurre porque el informante es hombre, porque es joven o porque siente algún malestar ese día. Además, si estas variables nos interesaran, habría que tener en cuenta que los hablantes se adaptan a los gestos de los interlocutores (Rasenberg et al. 2020). Esta alineación entre interlocutores afectaría a la forma del gesto o la frecuencia con la que se utiliza.

Por tanto, con este material no es posible realizar análisis comparativos entre los hablantes porque no son suficientes para ser representativos de su clase⁸⁵. Igualmente, para que no afectara ninguna variable social, se han buscado informantes heterogéneos: de distintas edades, sexos, nivel de estudios y con relaciones distintas entre sí (familiar, de amistad, de pareja). La única característica que se ha restringido es la variedad diatópica: todos los hablantes pertenecen al área del levante español, concretamente, a los municipios de Alicante y Villena. Los nombres que aparecen en los ejemplos de este trabajo, tomados del corpus, se han anonimizado.

El objetivo es descubrir las estructuras que aparecen en la conversación al incluir la información visual, por tanto, la característica más importante debe ser que sean conversaciones auténticas. Cada una se interpreta en su particularidad y los patrones encontrados podrían replicarse en corpus más extensos en el futuro.

El corpus ocupa un total de 131 minutos de grabación, alrededor de 28.300 palabras de transcripción y 453 gestos transcritos. Presentamos a continuación una tabla con las características básicas de cada una de las muestras:

	Lugar	Hablantes (sexo, edad, nivel de estudios)	Relación	Fecha	Duración
Conv1	Cocina de casa (Villena)	A- Hombre 24 medio B- Gancho C- Mujer 52 bajo D- Hombre 52 medio	Familia	9 de julio de 2021	30 min.
Conv2	Casa (Alicante)	A- Gancho B- Hombre 33 medio C- Hombre 32 medio D- Mujer 32 medio	Amigos (2 parejas)	25 de julio de 2021	36 min.
Conv3	Casa (Alicante)	A- Hombre 33 medio B- Gancho C- Mujer 63 bajo	Familia	12 de agosto de 2021	20 min.

⁸⁵ Poyatos (1994:147-151) denomina *sociokinésica* a la relación entre en el uso de un gesto y los factores de edad, sexo, estado emocional, ambiente, contexto y clase sociocultural.

Conv4	Casa (Villena)	A- Gancho B- Hombre 52 medio C- Mujer 85 bajo	Familia	15 de noviembre de 2021	16 min.
Conv5	Bar (Villena)	A- Gancho B- Mujer 29 alto C- Mujer 29 alto D- Mujer 29 medio	Amigas	3 de febrero de 2022	29 min.

Tabla 13. Características de las muestras

5.2.3. Equipo de grabación

Una característica de la grabación que consideramos importante es que capte el comportamiento no verbal de todos los participantes, incluso si su rol es predominantemente de oyente o es el gancho. Por ejemplo, una grabación en la que se enfoca a un hablante cada vez que emite un turno, pero ignora el comportamiento de los oyentes, no es válida, ya que esa perspectiva monologal no da cuenta de la colaboración que se produce en la construcción del discurso (Goodwin 1981, Bavelas et al. 2016). Lo ideal es poder incluir un plano amplio que permita abarcar toda la interacción, pero también otros cerrados en los que se aprecien detalles más pequeños como la expresión facial (Goodwin 1993).

Con el objetivo de asegurar que todos los elementos de la conversación queden grabados, se ha cuidado la posición de los hablantes y el ángulo que las cámaras pueden captar. La selección del equipo se ha restringido a las mejores opciones de calidad dentro de las cámaras que son lo suficientemente pequeñas o discretas para que los participantes no las perciban. La primera es una Xiaomi MiJia 360 Sphere, de 3456 x 1728 a 30fps, que graba en un ángulo de 360 grados y está escondida en una caja decorativa. La segunda es una cámara espía V4, de 2x2 cm de tamaño, y una resolución de 1920x1080P. La duración de batería de la primera cámara es menor, de 1 hora, frente a las 2 horas de la cámara pequeña. Ambas cámaras generan varios archivos de vídeo en cada grabación a intervalos de 5 minutos (cámara pequeña) y de 10 minutos (cámara 360), por lo que en la transición entre vídeos se puede perder la imagen durante medio segundo.

En las ocasiones en las que ha sido posible, se han utilizado ambas cámaras para asegurar la visibilidad del comportamiento no verbal de todos los participantes (Imagen 2), uniendo ambas grabaciones en un solo vídeo y seleccionando el audio de mejor calidad. En otras, solo se ha podido colocar una cámara, pero esta ha cumplido el mismo

propósito gracias a la disposición espacial de los participantes. La cámara 360 ha sido prioritaria por su ángulo, pero, aun escondida en una caja, solo se ha podido utilizar en el domicilio de la investigadora o desde su bolso (Imagen 2, 3 y 4), pues los propietarios de otros espacios sospecharían de ella (Imagen 1 y 5). Generalmente, hay una visión frontal de los hablantes, pero algunos se ven de perfil. Afortunadamente, es la investigadora la que suele estar en esta posición de perfil y su comportamiento no figurará en el análisis.



Imagen 1. Conv. 1, grabación con una cámara espía.

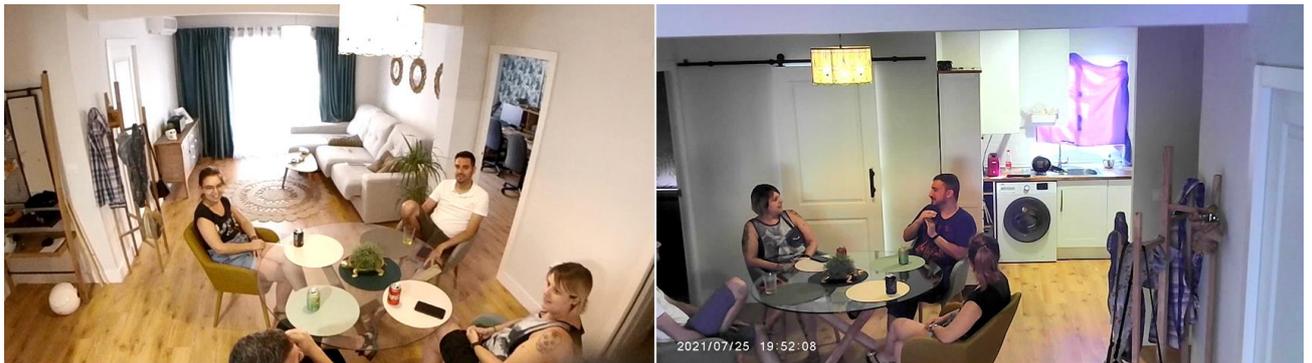


Imagen 2. Conv. 2, grabación con dos cámaras (360 y espía).



Imagen 3. Conv. 3, grabación con una cámara 360.



Imagen 4. Conv. 4, grabación con una cámara 360.



Imagen 5. Conv. 5, con una cámara espía.

El objetivo de nuestro análisis no es fonológico, por tanto, la calidad de audio es lo suficientemente buena para que el habla sea transcribible. Se recomiendan micrófonos independientes de la grabadora a una distancia adecuada de los hablantes, siendo los más eficaces los micrófonos de solapa, que, sin embargo, son intrusivos (Ten Have 1999: 71) e imposibles en esta recolección. Goodwin (1993) recomienda utilizar un micrófono dirigido hacia los hablantes, por ejemplo, en el centro de la mesa alrededor de la que se conversa, pero admite que también es intrusivo. En nuestro caso particular, el micrófono no debe ser reconocible, así que se ha optado, en las ocasiones en las que ha sido posible, por grabar el audio con un teléfono inteligente (Conv4 y Conv5).

5.3. Transcripción y segmentación de unidades conversacionales

En la siguiente fase, las conversaciones se transcriben y segmentan en unidades dialógicas utilizando el programa ELAN. Este trabajo se realiza dos veces en cada conversación, primero a partir de la grabación de audio y después a partir de la grabación de vídeo.

El sistema de transcripción verbal y el de segmentación son una adaptación de la transcripción de Val.Es.Co. (Briz y Grupo Val.Es.Co. 2002a) y su sistema de unidades (Briz y Grupo Val.Es.Co. 2003) al programa informático de anotación, siguiendo los pasos del Corpus Val.Es.Co. 3.0 (Pons Bordería 2022). A continuación, explicamos esta adaptación.

5.3.1. Transcripción

El formato vertical del papel a veces complica la lectura de un texto cuando hay varios elementos que se dicen o suceden simultáneamente. Normalmente, estas partes solapadas se escriben paralelamente y se indican con algún símbolo (el más extendido es el corchete), pero la continuación de las emisiones de los hablantes no puede continuar más allá del margen del papel, por consiguiente, deben continuar en la línea inferior y a veces se puede generar confusión.

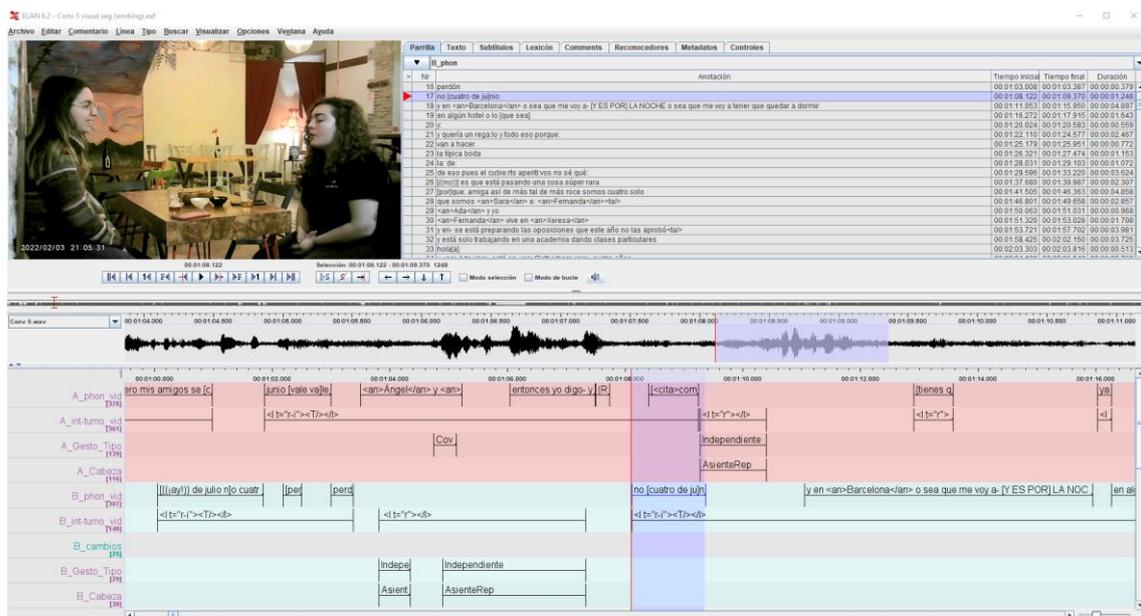


Imagen 6. Interfaz de ELAN con algunas líneas de los hablantes A y B.

El formato más adecuado para representar eventos simultáneos es el horizontal o de partitura que nos ofrece el software de anotación de ELAN (Imagen 6). En el eje vertical se utilizan varias líneas y cada una puede corresponder a un hablante o a distintos modos de comunicación; en el eje horizontal, la transcripción se alinea temporalmente con el archivo sonoro o de vídeo. Esto permite registrar de forma precisa el contenido y cómo se coordinan y se relacionan temporalmente los hablantes; así como efectuar una lectura paralela con la reproducción del archivo de sonido. Véase la diferencia entre la

transcripción en papel (2a) y en ELAN (2b) de un mismo fragmento de conversación (Conv5, 0:21:18-0:21:27).

(2a)
 B: per[o]
 C: [ya ((a ver))]
 B: [no ha- no ha sío] caro el tatuaje en [verdad eh porque hostia cien e]uros
 D: [((ha sío)) caro]
 C: [ha sío caro (RISAS)]
 B: cien euros ocho centímetros
 D: () lo que pasa es que se cree[e-

(2b)



En ELAN, la segmentación temporal de cada emisión está marcada por la extensión de cada comentario en la línea. Si dos comentarios de dos líneas diferentes se solapan temporalmente en algún momento, podemos ver la duración en la que ambos comentarios coinciden, pero no qué contenido transcrito se solapa, salvo si utilizamos los corchetes. No obstante, en la transcripción en papel, a falta de una alineación temporal, es necesario encajar las emisiones justo debajo de la parte de texto con la que coincide.

El nivel de transcripción verbal necesario para nuestro análisis de las unidades dialógicas no requiere información fonética o fonológica. Tampoco una transliteración ortográfica es adecuada porque se pierde información sobre las pausas y la entonación, que afectan a la interpretación pragmática. En inglés, existen transcripciones verbales especializadas en la interacción como la de Jefferson (1992, 2004) o GAT2 (Selting et al. 2011). Para este trabajo en español, tomamos el sistema de transcripción de Val.Es.Co. (Briz y Grupo Val.Es.Co 2002a), ya que estamos utilizando también su sistema de unidades y es la transcripción más extendida en los estudios de la oralidad en español. Esta transcripción se caracteriza por aprovechar los recursos tipográficos y los signos ortográficos convencionales para integrar los elementos propios de la oralidad en el texto y asegurar así una legibilidad fácil (Hidalgo y Sanmartín 2005: 28-30⁸⁶).

⁸⁶ Hidalgo Navarro, A. y Sanmartín Sáez, J. (2005). Los sistemas de transcripción de la lengua hablada, *Oralia: Análisis del discurso oral*, 8, 13-36. <https://ojs.ual.es/ojs/index.php/ORALIA/article/view/8325>

Los ejemplos mostrados durante este trabajo van a seguir la transcripción oral original de Val.Es.Co., pero en nuestra transcripción en ELAN, algunos signos desaparecen porque no son necesarios en el programa (Tabla 14). Especialmente, los relacionados con la extensión temporal de las emisiones verbales y las pausas entre emisiones no son necesarios porque en ELAN la extensión de las anotaciones se alinea con el eje temporal horizontal. En la Imagen 7, el espacio entre la primera y la segunda anotación del hablante C indica una pausa de 337 ms.

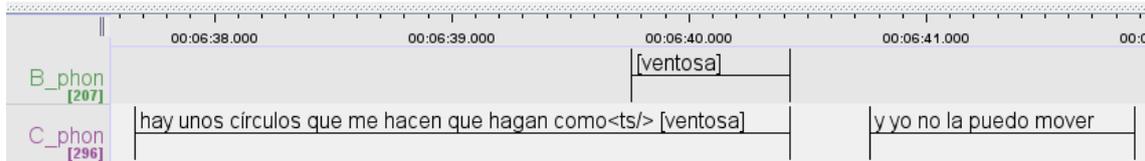


Imagen 7. Ejemplo de transcripción verbal en ELAN.

	Transcripción de Val.Es.Co. (Briz y Grupo Val.Es.Co 2002a)	Etiquetas TEI de Val.Es.Co. en ELAN (Pons Bordería 2022)	Transcripción en ELAN en este trabajo
Emisión de un interlocutor	:		
Interlocutor no reconocido.	¿?:		
Sucesión inmediata, sin pausa apreciable, entre dos emisiones de distintos interlocutores.	§		
Mantenimiento del turno de un participante en un solapamiento.	=		
Lugar donde se inicia un solapamiento superposición.	[[[
Final del habla simultánea.]]]
Reinicios y autointerrupciones sin pausa.	-	-	-
Pausa corta, inferior al medio segundo.	/		
Pausa entre medio segundo y un segundo.	//		
Pausa de un segundo o más.	///		
Silencio (lapso o intervalo) de 5 segundos; se indica el nº de segundos en las pausas de más de un segundo, cuando sea especialmente significativo.	(5")		
Entonación ascendente.	↑	<ta/>	<ta/>
Entonación descendente.	↓	<td/>	<td/>
Entonación mantenida o suspendida.	→	<ts/>	<ts/>

Entonación circunfleja		<tc/>	<tc/>
Los nombres propios, apodos, siglas y marcas, excepto las convertidas en “palabras-marca” de uso general, aparecen con la letra inicial en mayúscula.	Cou	Cou	Cou
Pronunciación marcada o enfática (dos o m letras mayúsculas).	PESADO	<enf t="prm">pesado</enf>	PESADO
Pronunciación silabeada.	pe sa do	<enf t="sil">pesado</enf>	pe sa do
Fragmento indescifrable.	(())	<in/>	(())
Transcripción dudosa.	((siempre))	((siempre))	((siempre))
Interrupciones de la grabación o de transcripción.	((...))		
Reconstrucción de una unidad léxica que se ha pronunciado incompleta, cuando pueda perturbar la comprensión.	(en)tonces	<fsr t="tonces">entonces</fsr>	
Fenómenos de fonética sintáctica entre palabras, especialmente marcados.	p'l	<fsr t="pa'l">para el</fsr>	p'l
Fragmento pronunciado con una intensidad baja o próxima al susurro.	°()°	<s> </s>	°()°
Aspiración de “s” implosiva.	h	<fsr t="h">s</fsr>	h
Aparecen al margen de los enunciados. En el caso de las risas, si son simultáneas a lo dicho, se transcribe el enunciado y en nota al pie se indica “entre risa”.	(RISAS, TOSES, GRITOS...)	<risas/> <e_risas> </e_risas> <tos/> <gritos/>	(RISAS, TOSES, GRITOS...)
Alargamientos vocálicos.	aa	<al/>	aa
Alargamientos consonánticos.	nn	<al/>	nn
Interrogaciones exclamativas.	¿i i!?	¿i i!?	¿i i!?
Interrogaciones. También para los apéndices del tipo “¿no?, ¿eh?, ¿sabes?”	¿ ?	¿ ?	¿ ?
Exclamaciones.	i !	i !	i !
Fragmento de conversación en valenciano. Se acompaña de una nota donde se traduce su contenido al castellano.	és que se pareix a mosatros	<obs t="valenciano: es que se parece a nosotros">és que se pareix a mosatros </obs>	
Reproducción e imitación de emisiones.	<i>Letra cursiva</i>	<cita> </cita>	<cita> </cita>

Estilo directo, característico de los denominados relatos conversacionales.			
Anotaciones pragmáticas que ofrecen información sobre las circunstancias de la enunciación. Rasgos complementarios del canal verbal. Añaden informaciones necesarias para la correcta interpretación de determinadas palabras (la correspondencia extranjera de la palabra transcrita en el texto de acuerdo con la pronunciación real, siglas, marcas, etc.), enunciados o secuencias del texto (p. e., los irónicos), de algunas onomatopeyas, etc.	Notas a pie de página	<code><obs t=""></obs></code> <code><ext t=""> </ext></code> <code><siglas t=""></code> <code></siglas></code>	
Escisiones conversacionales	Sangrados a la derecha		

Tabla 14. Sistema de transcripción de Val.Es.Co. (Briz y Grupo Val.Es.Co 2002a, Pons Bordería 2022) y adaptación a ELAN en este trabajo.

Como otros corpus⁸⁷, los corpus en línea de Val.Es.Co. (Ameresco, Val.Es.Co. 2.0, 2.1 y 3.0.) han creado versiones TEI (Text Encoding Initiative) de estos símbolos. Las etiquetas TEI codifican de manera estandarizada los datos externos de la conversación y sus hablantes, así como información sobre partes del texto relativa a características segmentales, suprasegmentales, paralingüísticas, extraverbales, léxicas y discursivas con el objetivo de facilitar la búsqueda informáticamente (Pons Bordería 2022). Sin embargo, en este trabajo no han sido utilizados para la transcripción porque la lectura mediante el sistema de transcripción original era suficiente para realizar el trabajo de segmentación de las emisiones verbales en intervenciones. Las únicas etiquetas utilizadas han sido la de cita y las de entonación porque en ELAN no se puede escribir en cursiva ni se puede usar el símbolo de flecha. Pero la inclusión de las etiquetas TEI sería necesaria si este corpus se compartiera en línea.

La transcripción oral se realiza en una línea por cada hablante (A_phon, B_phon, etc.); junto a una sola línea común para Observaciones, que ya existe en el Corpus Val.Es.Co. 3.0., para explicar hechos que no son elementos comunicativos propiamente dichos, pero interfieren en la interacción (por ejemplo, que uno de los informantes realiza alguna actividad que le impide participar en la interacción o que abandona el

⁸⁷ Algunos corpus que utilizan etiquetas TEI en sus sistemas de transcripción oral son el corpus COREC (Corpus Oral de Referencia del Español Contemporáneo), el ALCORE (Alicante, Corpus oral del español), el VUM (Vernáculo Urbano Malagueño) y el PRESEEA (Proyecto de estudio sociolingüístico del español de España y de América) (Hidalgo y Sanmartín 2005).

lugar). Además, se han añadido un par de líneas de observación para anotar dudas que han surgido durante la segmentación (Obs_segmentación) y la anotación del vídeo (Obs_vídeo).

5.3.2. Segmentación

La segmentación de las emisiones verbales en unidades conversacionales se ha realizado como en el Corpus Val.Es.Co. 3.0 (Pons Bordería 2022, 2023). Se requieren tres líneas para cada hablante: una de transcripción, otra para *tokenizar* las palabras (una anotación para cada palabra) y la propia línea de segmentación. Las anotaciones de la línea de segmentación pueden hacerse de la misma extensión que las anotaciones de la transcripción, pero partiendo de la línea *tokenizada*, que es dependiente de la transcrita (por eso es de color amarillo), se evitan desalineaciones entre las líneas de segmentación y de transcripción y, además, es posible indicar que dos partes de una transcripción hecha en un comentario de ELAN pertenecen a dos intervenciones distintas (Imagen 8).

B_phon [198]	ya está no no no [na- no solo se hacía-] solo se hacía la prueba e[sa a ver si sabías aparcar]
B_palabras [1191]	ya está no no no [na- no solo se hacía solo se hacía la prue e[sa a ver si sabía apar
B_int-turno [84]	<I t="r-i"><T/></I>

Imagen 8. Ejemplo de segmentación de intervención en ELAN.

Los tipos de intervenciones que contempla el modelo Val.Es.Co. (§2.3) son, desde el punto de vista estructural: intervenciones iniciativas (Ii), intervenciones reactivas (Ir), intervenciones reactivo-iniciativas (Ir-i) e intervenciones independientes (Iind). Desde el punto de vista social, estas intervenciones pueden ser turno (T) o no, simplemente intervención (I). El Corpus Val.Es.Co. 3.0 utiliza para la segmentación de intervenciones las etiquetas TEI de la siguiente manera (Pons Bordería 2022): se marcan con la etiqueta doble <I>...</I> las intervenciones y se especifica el tipo de intervención en el atributo de dicha etiqueta (t= “i”, t= “r”, t= “r/i”, t= “ind”); también se incluye la etiqueta de turno (<T/>) dentro de las intervenciones que lo sean. Por ejemplo, una intervención reactivo-iniciativa turno se vería así: <I t= “r/i”><T/></I>.

Las etiquetas utilizadas en la línea de segmentación de este estudio corresponden a las mismas utilizadas para la web del Corpus Val.Es.Co. 3.0. Tan solo se ha incorporado un nuevo atributo para las intervenciones ininteligibles (t= “int”), que no es necesaria en la web porque los fragmentos de habla ininteligible no se segmentaron como

intervención. Pero en este trabajo resulta útil para realizar la comparativa entre la segmentación verbal y visual, pues se identifica rápidamente que una nueva intervención (Ir, Ir-i, etc.) en la segmentación visual no ha aparecido directamente de la información visual, sino que en el análisis verbal ya se identificaba un fragmento ininteligible que, gracias al resto de la información, es segmentable.

Las intervenciones se anotan en una línea propia para cada hablante (A_int- turno, B_int- turno, etc.). Las líneas de transcripción y de observación no tienen restricciones de escritura, pero para las de segmentación se utiliza un vocabulario predeterminado desplegable con el objetivo de evitar errores de escritura (Imagen 9).

<l t="i"><T/...	Intervención iniciativa
<l t="r"><T/...	Intervención reactiva (turno)
<l t="r"></l>	Intervención reactiva (no turno)
<l t="r-i"><T/...	Intervención reactivo-iniciativa
<l t="ind"></l>	Intervención independiente
<l t="int"/></l>	Intervención Ininteligible

Imagen 9. Desplegable para segmentar las intervenciones en ELAN.

Como se compara la segmentación con vídeo y sin vídeo, las líneas de transcripción y segmentación se duplican, añadiendo *_vid* en el nombre de la línea que tiene en cuenta la información visual (A_phon_vid, A_int- turno_vid).

En resumen, en esta fase de transcripción y segmentación las líneas que vamos a utilizar son las que aparecen en la Tabla 15, con el nombre que aparece en la columna de atributo.

DOMINIO	ATRIBUTO	VALOR
Transcripción oral	A_phon	(escritura libre)
	A_palabras	(transcripción <i>tokenizada</i>)
	Observaciones	(escritura libre)
Segmentación oral	A_int- turno	Intervención iniciativa, Intervención reactiva (turno), Intervención reactiva (no- turno), Intervención reactivo- iniciativa, Intervención independiente, Intervención ininteligible.
Transcripción visual	A_phon_vid	(escritura libre)
	A_palabras_vid	(transcripción <i>tokenizada</i>)
	Obs_vídeo	(escritura libre)
Segmentación visual	A_int- turno_vid	Intervención iniciativa, Intervención reactiva (turno), Intervención reactiva (no- turno), Intervención reactivo- iniciativa, Intervención independiente.
	Obs_segmentación	(escritura libre)

Tabla 15. Esquema de anotación: líneas de transcripción y segmentación.

En cuanto a la visualización de la segmentación dialógica en la transcripción en papel, Val.Es.Co. codifica a los hablantes alfabéticamente en orden de aparición (A, B, C...) y la intervención verbal de cada hablante está acompañada por números previos y posteriores para indicar el conteo de intervenciones y turnos (Briz y Grupo Val.Es.Co. 2002a; Briz 2006: 51). El número de a la izquierda indica el turno y el número a la derecha, la intervención. Por ejemplo, B1 es la primera intervención de B en la conversación, pero no es turno porque no tiene ningún número a la izquierda; 1B2 es su segunda intervención, pero esta sí es turno. Si aparece la letra del hablante sin número, esto indica que esa emisión forma parte de la intervención indicada previamente de ese hablante; es necesario cuando la intervención se solapa con la de otro hablante. Véase el ejemplo 3a.

(3a)

1C1: he aprendió más que en toda la carrera/ porque son todo apli[ca]
 B1: [(joer)]
 C: =ciones ee programas de cómo hacer cosas

Conv5, 0:17:04 (sin gestos)

En ciertos ejemplos de este trabajo también nos serviremos de la herramienta de visualización Dientes de sierra (Espinosa Guerri 2016, Espinosa-Guerri y García-Ramón 2019) para dejar más clara la relación iniciativa y reactiva de los intercambios, especialmente en casos de solapamiento donde sea necesario. Esta herramienta se basa en líneas de colores para cada hablante que vincula el código de cada intervención generando picos entre intercambios. Véase su aplicación en el ejemplo 3b. La intervención 1C1 está vinculada con una intervención previa al ejemplo y con una intervención posterior de B no visible, creando picos continuados (figura de cremallera). B1 es una intervención reactiva, así que está unida a 1C1 por arriba, pero no por abajo porque no genera reacción. La línea intermitente en C indica que esa transcripción es continuación de 1C1.

(3b)

1C1: he aprendió más que en toda la carrera/ porque son todo apli[ca]
 B1: [(joer)]
 C: =ciones ee programas de cómo hacer cosas

El diagrama muestra tres líneas de colores que conectan las intervenciones: una línea azul que forma un pico continuo entre 1C1 y una intervención anterior no visible; una línea roja que conecta B1 con 1C1 por arriba; y una línea azul intermitente que conecta C con 1C1.

Conv5, 0:17:04 (sin gestos)

5.4. Análisis de cambios en la segmentación

El principal objetivo del análisis de los cambios estructurales y sociales que se producen al incorporar los elementos visuales es confirmar que los cambios producidos por la segmentación multimodal son suficientes para justificar la incorporación de los elementos visuales en la descripción de la estructura dialógica y su dimensión social. De ahí derivan los objetivos de descripción de los cambios producidos entre la segmentación oral y la segmentación visual:

- 1) comprobar que la serie de elementos visuales propuestos en este capítulo producen cambios en la segmentación dialógica, tanto en el nivel estructural como en el social;
- 2) describir los efectos de los elementos multimodales en la segmentación estructural (creación, modificación o eliminación de intervenciones) y social (creación, modificación o eliminación de turnos);
- 3) evaluar cuáles tienen mayor efecto en la segmentación y en qué tipo de cambios.

A partir del conocimiento que tenemos sobre las unidades conversacionales (§2) y la multimodalidad (§3) y, sobre todo, de los resultados de un análisis preliminar (Cabanes Pérez 2020), podemos anticipar algunos resultados esperados que explicamos a continuación.

En el estudio preliminar se utilizó la misma metodología comparando la segmentación de una conversación de 13 minutos mediante las unidades dialógicas de Val.Es.Co. a partir de la grabación de audio y de la grabación de vídeo de esta misma conversación. De un total de 44 cambios en la segmentación dialógica, un 57% fueron provocados por gestos que no forman parte de una intervención verbal, que no acompañan a actos verbales, sino que por sí mismos forman el único acto de una intervención gestual; mientras un 43 % de los cambios fueron provocados por el resto de las causas. Nuestra hipótesis es que podemos encontrar unas proporciones similares en un corpus más amplio, pero las nuevas variables que detallan el resto de las causas nos ayudarán a explicar mejor los cambios producidos. Respecto a la dimensión social, la mayoría de estos gestos ocupaban posición de Ir (88%), así que esperamos que muchas de las intervenciones gestuales nuevas no sean turnos; además, solo el 36% afectaba socialmente a otras intervenciones cercanas, es decir, que solo el 36% validaba a una intervención que en la segmentación oral se consideraba intervención no turno. Por consiguiente, nuestra

hipótesis es que en este estudio la mayoría de estos gestos tampoco afectará a la dimensión social de otras intervenciones. En cambio, esperamos que las otras variables generen menos intervenciones nuevas y afecten en mayor medida al cambio estructural y social de intervenciones verbales ya existentes, como en el análisis preliminar.

Para realizar este análisis, establecemos los tipos de cambio que se pueden producir entre ambas segmentaciones. Siguiendo el modelo Val.Es.Co., observamos los cambios que se producen en la dimensión estructural, que incumbe al tipo de intervención (Ii, Ir, Ir-i, Iind) o en la dimensión social (turno/intervención); además, consideramos el alcance de estos cambios en intervenciones colindantes. Estos cambios son las variables dependientes del análisis estadístico, pues son los hechos que encontramos en las segmentaciones y son producidos por las variables independientes.

También establecemos qué causas visuales son las consideradas para modificar la segmentación de la segmentación. La tipología de estos elementos se basa en la bibliografía y en un análisis preliminar. Son las variables independientes del análisis estadístico, pues son las causas que produce los cambios en la variable dependiente.

Mediante la relación de estas variables, podemos averiguar cómo y en qué medida cada una de las causas visuales modifican la segmentación estructural y social de la conversación.

5.4.1. Variable cambio

En el estudio preliminar (Cabanes Pérez 2020), el foco estaba en los gestos que no acompañaban al habla y que por sí mismos forman intervenciones. Se contabilizó cuántas de estas intervenciones funcionaban como Ir que daban retroalimentación a un hablante (la mayoría), las veces en las que esta retroalimentación causaba cambios de validación en la intervención verbal previa o esta se mantenía igual y los gestos que ocupaban la posición de un turno que es parte de un par adyacente. Asimismo, aparecieron otros cambios en la segmentación no causados por estos gestos, sino por otros elementos visuales que aparecían en el vídeo que se agruparon como “mejora de la transcripción”.

Estos resultados hicieron notar la necesidad de preestablecer unas causas visuales de cambio más específicas, así como mayor detalle en los cambios producidos. Por consiguiente, las variables dependientes del primer análisis de este estudio son el cambio en la dimensión estructural, el cambio en la dimensión social y el alcance del cambio.

El cambio estructural comprende la aparición de nuevas intervenciones y el cambio de tipo de una intervención verbal ya existente en la segmentación del audio, es decir, si la intervención verbal cambia a ser iniciativa, reactivo-iniciativa, reactiva o independiente. El cambio en la dimensión social se relaciona con la validación de la intervención, o sea, una intervención resulta validada en la segmentación visual y se convierte en turno o, por el contrario, una intervención considerada turno ya no lo es. Sin embargo, puede haber intervenciones que cambien de tipo sin que cambie su validación; por ejemplo, si una *Iind* cambia a *Ir*, continúa sin estar validada. Asimismo, las nuevas intervenciones que aparecen en la segmentación también forman parte de la dimensión social y pueden estar validadas o no, en consecuencia, pueden suponer nuevos turnos para la conversación. Por tanto, los dos valores de la variable de cambio estructural (nueva intervención y modificación de intervención), se divide en los seis valores de la variable de cambio social, tal como se observa en la Tabla 16.

Variables	Valores					
Variable cambio estructural	Nueva intervención			Modificación de intervención		
Variable cambio social	Nueva intervención (no validada)	Nuevo turno	Modificación de validación		Sin modificación de validación.	
			De intervención a turno	De turno a intervención	Se mantiene como turno	Se mantiene como intervención

Tabla 16. Variables de cambio estructural y de cambio social y sus valores

Respecto a la variable del alcance, las intervenciones no son entes aislados, las acciones realizadas en una intervención pueden modificar la segmentación de otra intervención con la que forman un intercambio. Entonces, un elemento visual puede causar cambios en la propia intervención en la que aparece, en una intervención ajena o en ambas. A cada valor de las variables de cambio estructural y social se le aplica la variable alcance para crear nuevos valores (Tabla 17).

Valores del cambio estructural y social	Variable alcance	
	En intervención propia	En intervención ajena
Nueva intervención	Nueva intervención propia	Nueva intervención ajena
Nueva intervención (no validada)	Nueva intervención (no validada) propia	Nueva intervención (no validada) ajena

Nuevo turno	Nuevo turno propio	Nuevo turno ajeno
Modificación de intervención	Modificación de intervención propia	Modificación de intervención ajena
Modificación de validación: de intervención a turno	Modificación de validación propia: de intervención a turno	Modificación de validación ajena: de intervención a turno
Modificación de validación: de turno a intervención	Modificación de validación propia: de turno a intervención	Modificación de validación ajena: de turno a intervención
Sin modificación de validación: se mantiene como turno	Sin modificación de validación propia: se mantiene como turno	Sin modificación de validación ajena: se mantiene como turno
Sin modificación de validación: se mantiene como intervención	Sin modificación de validación propia: se mantiene como intervención	Sin modificación de validación ajena: se mantiene como intervención

Tabla 17. Variable alcance y valores de cambio estructural y de cambio social.

Las variables dependientes se ven afectadas por las variables independientes, es decir, los cambios en la segmentación son causados por los elementos visuales.

5.4.2. Variable causa

No se pretende en nuestro estudio identificar, describir y cuantificar todos los elementos de cualquier modalidad visual que aparecen en la conversación. Para cumplir nuestros objetivos, buscamos únicamente los que generan cambios en la segmentación dialógica. En el estudio preliminar tan solo se tuvieron en cuenta dos variables independientes para explicar los cambios en el nivel de la intervención: 1) los gestos que no acompañaban al habla y que son la única información emitida por un hablante como *intervenciones turno* o como *intervenciones no turno*, 2) otros comportamientos físicos visibles. Estos últimos podían incluir cualquier información extraída de la información visual: articulación labial, las posturas y las miradas para identificar el emisor y el receptor de las intervenciones verbales o los gestos que acompañaban a las intervenciones verbales. El foco estaba únicamente en los gestos que no acompañan al habla, pero el hecho de que otros datos modificaran la interpretación de la intervención hizo necesario crear esta variable. El estudio actual extiende la cantidad de material para analizar y

pretende ser más riguroso estableciendo unas variables más específicas de las causas del cambio.

Las *variables independientes* de este estudio son una serie de causas que crean cambios estructurales y sociales, tanto en la propia intervención en la que aparecen como en otras intervenciones verbales contiguas. Considerando el estudio preliminar, la bibliografía y el contraste entre la segmentación oral y visual de nuestras propias conversaciones, se establecen como variables los gestos independientes, los gestos coverbales, la orientación corporal en las intervenciones verbales, las acciones extralingüísticas y la lectura de labios. También puede haber una combinación varias causas.

Variables independientes	Gestos independientes
	Gestos coverbales
	Orientación corporal en las intervenciones verbales
	Acciones extralingüísticas
	Lectura de labios

Tabla 18. Variable causa.

Gestos independientes y gestos coverbales

El modelo Val.Es.Co ha analizado unidades del discurso, principalmente marcadores del discurso, a través de la posición que ocupan en él (Briz, Pons y Portolés 2008, Briz y Pons Bordería 2010, Briz 2012b, Salameh Jiménez et al. 2018, Pons Bordería 2018, Llopis Cardona 2020). Si las unidades del sistema son *subacto, acto, intervención, diálogo* y *discurso*, el elemento estudiado puede constituir una de ellas y operar en una posición *inicial, medial, final* o *independiente* de una unidad inmediatamente superior. Estas posiciones se vinculan con las funciones que el marcador cumple en cada contexto estructural. Mediante la rejilla de unidades x posiciones (Pons Bordería 2022, Pons et al. en prensa) se contempla esquemáticamente esa relación (Imagen 10) y mediante el análisis de varios marcadores se pueden encontrar celdas especializadas en ciertas funciones.

¿EH? ^{1, 2, 3}	SUBACTO		Acto	INTERVENCIÓN		DIÁLOGO	DISCURSO
	SSS	SSD		Ir	Ii		
Inicial	(1) FORMULACIÓN (SAT)	(4) FORMULACIÓN (SAT)		(6) APELATIVO (SAI) (7) DESACUERDO (SAI/SAI) (8) PETICIÓN DE ACLARACIÓN (SAI)			
Inicial relativa	(2) FORMULACIÓN (SAT)						
Media			(9) APELATIVO (SAI / SAM) (10) INTENSIFICACIÓN (SAM)				
Final	(3) INTENSIFICACIÓN (SAM)	(5) INTENSIFICACIÓN (SAM)	(12) INTENSIFICACIÓN (SAM) (13) INTENSIFICACIÓN (SAM/SAI)		(11) APELATIVO (SAI)		
Independiente				(14) PETICIÓN DE ACLARACIÓN (SAI)	(15) DESACUERDO (SAM / SAI)		

Imagen 10. Posición (avanzado) de ¿eh? en DPDE <<http://www.dpde.es/#/entry/eh1>>

La partícula ¿eh?⁸⁸ es un subacto adyacente que cumple una función textual de formulación cuando antecede a un suabacto sustantivo, pero la función es modal de intensificación después del subacto o del acto o a mitad del acto. También a mitad de acto la función puede ser interpersonal apelativa, igual que en posición inicial de intervención reactiva o a final de intervención iniciativa. A inicio de Ir también puede tener función de desacuerdo o petición de aclaración, las cuales también pueden aparecer de manera independiente de intervención, sin acompañar a ningún acto.

Este mismo tipo de análisis se podría aplicar a los gestos para analizar cómo afectan a la estructura de la conversación o a su dimensión social (Tabla 19). Hay gestos que no acompañan a ningún tipo de contenido verbal del mismo interlocutor (§3.4.2.4), por lo que ellos mismos son el único acto de la intervención. Es acto porque contiene una fuerza ilocutiva vinculada a la forma del gesto y a su uso en el contexto y es intervención porque responde a la dinámica interactiva de ser o provocar reacción. Para analizar este gesto en posición independiente de intervención, hemos creado los términos *gesto independiente*, en contraste con el gesto coverbal, o *intervención gestual*, en contraste con la intervención verbal.

⁸⁸ En el *Diccionario de partículas del español (DPDE)* (Briz, Pons y Portolés 2008) se indica la posición de gran parte de las partículas discursivas (sección *Posición avanzado* de cada entrada).

		de acto	de intervención
Posición independiente			#acto gestual#
Posición coverbal	inicial	# {subacto gestual} {subacto verbal} #	#acto gestual# #acto verbal#
	medial	# {subacto verbal} {subacto gestual} {subacto verbal} #	#acto verbal# #acto gestual# #acto verbal#
	final	# {subacto verbal} {subacto gestual} #	#acto verbal# #acto gestual#
	simultanea	# {subacto verbal} # {subactogestual}	

Tabla 19. Posiciones estructurales del gesto

El *gesto coverbal* designa al gesto que sí forma parte de una intervención que incluye contenido verbal. Observando el esquema de posiciones de Val.Es.Co. un gesto podría ser un acto que ocupara la posición inicial, medial o final de intervención o un subacto en posición inicial, medial o final de acto. La diferencia entre el acto y el subacto estriba en la interpretación del elemento como poseedor de una fuerza ilocutiva propia o como constituyente informativo menor que requiere de otros para conformar el acto. Sin embargo, aunque los gestos y los subactos puedan cumplir funciones similares que completan el contenido del acto a nivel semántico o pragmático (*subacto sustantivo – gesto referencial; subacto adyacente textual – gesto discursivo; subacto adyacente modal – gesto modal; subacto adyacente interpersonal – gesto interactivo*), no se puede afirmar que son dos modalidades de una misma categoría porque la relación que se produce entre la información verbal y entre la información verbal y la gestual, por la naturaleza de cada modalidad, es distinta. Por ejemplo, un gesto referencial normalmente dependerá de un SS verbal; sin él, el gesto no se entendería por sí mismo porque no tiene una forma convencionalizada. En cambio, un subacto verbal sustantivo subordinado (SSS), aun desconociendo el subacto sustantivo director (SSD) al que complementa, se entiende a nivel semántico porque está formado por palabras. Asimismo, elementos paraverbales como la pausa o el tonema al final de la intervención no se han contemplado como subactos en el análisis oral de la conversación de Val.Es.Co, sino como parte de las unidades verbales, aun cuando las modifican pragmáticamente. Los gestos comparten con los elementos paraverbales el hecho de que pueden realizarse de forma simultánea al habla.

La comunicación multimodal hace más complejo este esquema de posiciones, pues, si bien los marcadores del discurso se pueden ubicar en posición inicial, medial o final de otro elemento lingüístico, el gesto puede realizarse también, y a menudo es así, de forma simultánea al habla, al igual que la entonación. Al ser dos modos de

comunicación distintos, pueden realizarse al mismo tiempo, quebrando la limitación de la linealidad del habla. Por tanto, podemos plantear una nueva posición: la simultánea.

Aunque hemos dicho que la clasificación de un gesto coverbal como subacto está abierta a debate, en este estudio lo consideramos subacto gestual para simplificar el esquema (Tabla 19). En posición simultánea, puede acompañar a otro subacto verbal, quizás director del acto, o a varios, si bien la sincronía no tiene por qué ser perfecta: es posible que el subacto gestual no coincida exactamente con la emisión verbal y el gesto se produzca un poco antes o un poco después y durante el subacto verbal. No se ha incluido porque no debe ser muy común, pero también puede haber simultaneidad de habla y gesto en posición independiente de intervención. Se daría en el caso de que un hablante realizara cada emisión, la verbal y la visual, hacia interlocutores distintos.

Esta propuesta de posiciones del gesto no se aplica a este análisis de manera completa. Para simplificar el análisis, tomamos únicamente la distinción entre *gestos independientes*, es decir, los actos gestuales en posición independiente de intervención, y *gestos coverbales*, es decir, los gestos en posición coverbal de intervención o de acto, ya se produzcan durante, antes, en medio o después de otros subactos o actos verbales. Las posiciones internas de los constituyentes de la intervención verbal no son una prioridad para analizar los cambios en el nivel dialógico; basta con anotar que son coverbales.

Puesto que en nuestro análisis solamente se segmentan los gestos que afectan a la segmentación de intervenciones de la conversación, todos los gestos independientes se segmentan porque por sí mismos son nuevas intervenciones, pero solo algunos gestos coverbales se segmentan, los que modifican la interpretación estructural (como Ii, Ir, Ir-i o Iind) o social (I o T) de la propia intervención a la que pertenece o la de una intervención previa o posterior.

Orientación corporal

Hemos visto en §3.3 que la mirada y la orientación del cuerpo, al contrario que otros movimientos corporales, solo cumplen funciones interactivas relacionadas con el compromiso de los interlocutores en la conversación y la gestión del turno. Los movimientos de estos modos visuales son binarios, “hay” o “no hay”: mirada o no mirada hacia el interlocutor, postura o no postura hacia el interlocutor. En cambio, los gestos manuales y faciales, por su variedad articulatoria, pueden contener significado además de

cumplir funciones interactivas. La mirada y la postura están implicadas en la muestra de atención y compromiso hacia el interlocutor, lo que deriva en funciones reguladoras del turno por parte del emisor; incluso elicitan la respuesta de otra persona. Esta función procedimental exclusiva no es por sí misma argumento suficiente para separar estos movimientos de los del resto de partes del cuerpo, pues ya se ha visto que el asentimiento de cabeza también puede tener un uso predominantemente procedimental confirmando solamente el seguimiento (otras veces transmite entendimiento o acuerdo también).

Sin embargo, es discutible que sin acompañar a una intervención verbal la orientación de la mirada o la postura tenga la fuerza ilocutiva que la convierta en acto, como sí sucede con el resto de los gestos en posición independiente de intervención. La simple mirada de un interlocutor puede demostrar atención, que el emisor está siendo escuchado, pero no creemos que valide la contribución del emisor como contenido relevante en la conversación. En cambio, otras reacciones más explícitas, como un contenido lingüístico o un gesto explícito con la cabeza, la cara, los hombros o las manos dirigidos al primer emisor al terminar o cerca de la terminación de la intervención, sí lo validan. Estos elementos sí reaccionan al turno y lo validan porque los atienden “de manera manifiesta y simultánea” (Briz 2000, Briz y Grupo Val.Es.Co. 2002a, Grupo Val.Es.Co. 2014). Quizás la orientación de cuerpo y mirada supongan otro nivel de validación menos fuerte, no contemplado en la validación verbal. Al fin y al cabo, si ni siquiera recibe una mirada, el emisor deja de hablar o solicita de alguna manera la atención del oyente. Pero una escucha pasiva no puede tener el mismo peso que una reacción explícita. Tal como explica Val.Es.Co., “cabe considerar que tal intervención no aceptada ha sido percibida fisiológicamente (oída) por el/los otro/s interlocutor/es, esto es, ha habido receptor/es, pero al no aceptarla, no se les atribuye el carácter de oyente/s” (Grupo Val.Es.Co. 2014: 28). Es decir, que el hecho de haber dejado el espacio a la intervención para que se emita y se escuche no la valida. La mirada, entonces, podemos considerarla como una herramienta de esa escucha que por sí misma no valida.

Como hace notar Schegloff (1982: 78), aparecen señales fáticas en forma de *uh-huh* y asentimientos aunque el oyente mantenga la mirada; si esta última se considerara una muestra de atención suficiente, no haría falta añadir nada más. Y es que hacia el posible final de un turno son relevantes estas muestras de entendimiento y de postura ante lo dicho hasta ahora (Schegloff 1982: 81), no realizables con la mirada.

La mirada y la postura no encajan como otros movimientos en las unidades de Val.Es.Co. porque no funcionan como reacciones, son modalidades que funcionan de manera distinta y se aprecia en su segmentación temporal. La mirada de un oyente puede mantenerse dirigida hacia el emisor durante varias intervenciones de este y mientras él mismo reacciona verbalmente a la narración (Rossano 2012); de igual modo, la mirada del hablante que se dirigía hacia el oyente puede mantenerse sin cambios cuando cambian los roles y ahora es el oyente de su interlocutor (Oertel *et al.* 2012). Por tanto, las miradas coverbales no se vinculan estrictamente a unas intervenciones u otras. Están relacionadas con la gestión de los turnos, muestran y monitorizan la atención, pero su segmentación es distinta a otros gestos más breves que reaccionan al contenido. Para el modelo de segmentación de la Coenunciación (§2.4) la mirada y la dirección de la cabeza delimitan una unidad que agrupa dos o más párrafos: el hiperpárrafo (Bouvet y Morel 2002: 112).

La mirada también funciona como un canal de comunicación abierto. Dos personas pueden mirarse y decidir no iniciar una conversación o un nuevo diálogo. Efectivamente, hay una comunicación dentro de esa modalidad, pero no de contenido. Esas miradas sirven de andamiaje para empezar a construir la interacción lingüística.

Esta diferencia no es problemática si asumimos el carácter multimodal de la comunicación. Al igual que la prosodia no tiene un significado segmentable, sino dependiente de lo verbal, la mirada y la postura también trabajan de manera paralela y vinculadas a la interacción. Téngase en cuenta, además, que, al contrario que los movimientos de otras partes del cuerpo, los ojos son el medio de percepción con el que se observa el comportamiento del otro.

Entiéndase que no negamos que la mirada y la postura tengan funciones comunicativas en la interacción. Nuestra distinción del resto de movimientos es una cuestión estructural. Mientras al resto de gestos se les puede aplicar el esquema de posiciones de Val.Es.Co (visto en la Tabla 19), la mirada y la postura son elementos con una temporalidad demasiado amplia. Otros autores no entran en este problema categorial porque no necesitan situar la mirada o la postura como *acto* u otra unidad jerárquica para explicar su funcionamiento en la interacción. Por tanto, cuando hablemos de *gestos* en nuestro análisis, nos referiremos a los movimientos de cabeza, cara, hombros y manos que por sus efectos comunicativos tienen cabida en el sistema de unidades de Val.Es.Co.

Esta distinción no implica que se haya descartado la información de la orientación de los interlocutores en el análisis. Si bien el tipo de grabación no permite segmentar

adecuadamente las miradas del corpus, la orientación de la cabeza durante el gesto sí se va a incluir como una característica formal más de este. Asimismo, en el primer análisis de este trabajo, la orientación corporal durante una intervención verbal es una variable más que también puede cambiar la interpretación de esa u otra intervención como turno.

La orientación corporal es una modalidad que establece el espacio de interacción, los interlocutores que participan en la conversación, pero también en cada turno, y su grado de compromiso con la actividad (§3.3.1), de manera que, para nuestra segmentación, es un indicador de la dirección de las intervenciones entre emisor y receptor. Mientras hablamos, orientamos nuestro cuerpo hacia los receptores, por consiguiente, en nuestra segmentación, a veces la orientación del hablante puede cambiar la interpretación de la intervención verbal, de entre quienes se produce el intercambio, pues solamente con el contexto auditivo se podía creer que el destinatario era otro o era difícil de identificar. Este valor compete a la orientación de los interlocutores durante una intervención verbal, pues la orientación durante la realización de un gesto es un parámetro de forma del propio gesto (§5.5.2.), que no se valora por separado como causa de cambios en la segmentación.

Acciones extracomunicativas

No todas las acciones corporales que realizan los hablantes en el espacio de interacción son gestos, es decir, movimientos más o menos convencionales o espontáneos con significado semántico o procedimental que forman parte de la conversación, sino que son parte de otras actividades físicas, como comer o coger un objeto.

En el modelo modular de Roulet et al. (2001), se contemplan acciones que influyen directamente en la interacción. El módulo referencial (las actividades sociales) y el módulo jerárquico (las estructuras textuales) se relacionan dentro del marco de la estructura operacional para comprender la complejidad de las acciones comunicativas (Roulet et al. 2001: 211), porque en la interacción también intervienen actividades materiales y prácticas sociales no lingüísticas que conllevan conductas ostensivas (Roulet et al. 2001: 205-206). Por ejemplo, dar el dinero en una compraventa influye en el desarrollo de la interacción, pero es una acción comunicativa no lingüística. Estas acciones no suelen aparecer tanto en la conversación familiar como en otros géneros dialógicos transaccionales (Roulet et al. 2001: 222), pero podrían encontrarse en nuestro corpus afectando a la interpretación de las intervenciones, por ejemplo, cuando un

interlocutor pide a otro que realice una acción con algún objeto o cuando un objeto causa una reacción en alguien.

Lectura de labios

Estrictamente, la lectura de labios no es una modalidad distinta a la oral, pues la configuración de los labios al pronunciar palabras es un efecto secundario de la comunicación verbal, pero los interlocutores pueden servirse del canal visual para confirmar lo que escuchan si las condiciones de escucha no son buenas, así como el transcriptor de las grabaciones.

La lectura de labios fue de gran ayuda en el análisis preliminar para identificar a los emisores en situaciones en las que la voz de dos hablantes es similar, hablan con volumen bajo o emiten sonidos paraverbales (risa, bufido, chasquido, etc.) que no se identifican claramente. Esto produce cambios en la estructura de la interacción porque en una intervención se identifica al hablante incorrectamente o porque se descubre una intervención completamente nueva. Además, aunque el emisor estuviera bien identificado, la lectura de labios puede ayudar a comprender un fragmento ininteligible auditivamente o a corregir una mala interpretación. Todos estos cambios afectan a la segmentación de las intervenciones verbales y a su relación estructural y social con otras intervenciones.

5.4.3. Criterios de segmentación de gestos

Como la definición teórica de los gestos independientes y los gestos coverbales ha generado dudas en su aplicación en la anotación del corpus, dedicamos este apartado a explicar las decisiones tomadas en algunos casos en los que la identificación necesita más criterios que los planteados hasta ahora. Los ejemplos transcritos utilizan el sistema de transcripción de gestos explicado posteriormente en §5.5.4.

Es necesario tener en cuenta la información auditiva para identificar los gestos que forman parte de una intervención verbal o son independientes. La segmentación de los gestos independientes, *a priori*, es clara porque se trata del único componente de la intervención, no acompaña a una intervención verbal de manera simultánea ni en una posición previa o posterior al habla que pueda considerarse parte de la misma intervención. El gesto coverbal está más claro cuando es simultáneo al habla, pero si aparece antes o después del habla con la suficiente distancia, podemos plantearnos si son

intervenciones distintas. También es complicado distinguir a veces si dos gestos forman parte de la misma intervención o si el segundo es coverbal del siguiente acto verbal del mismo hablante.

Criterio de reacción

El principal criterio para distinguir un gesto independiente de uno coverbal es el hecho de que reaccione a un acto distinto de un hablante. En actos narrativos largos hemos considerado también la reacción a un subacto distinto, pues se reacciona a una información nueva concreta. Algunos parámetros que pueden ayudar a diferenciar reacciones son el tiempo entre el acto y la reacción y la relación entre la función del gesto y la del acto.

En el ejemplo 4 se observa la división de la retroalimentación del oyente, que reacciona a distintos actos del turno del hablante. Se han dividido las reacciones de A en 3 intervenciones: primero A1, con risas y un asentimiento repetido, que reaccionan a la descripción de C de la tarea con el gesto de manos; después, el asentimiento repetido no ha parado, pero se intensifica con la elevación de cejas y reacciona a la explicación de C “[((o sea))] me ha dao un montón de trabajo”; por último, una nueva elevación de cejas seguida de “mm” mientras asiente reaccionan a la explicación “porque hemos hecho una- hasta una página web y to”.

(4)

1C1: sí y ha sío como Mano⁸⁹

A1: (RISAS) [(RISAS)] [AsienteRep-----]

C: [(o sea)] [me ha dao un montón de trabajo] [() eso]

A2: [CejEleva]
[AsienteRep--]

C: porque hemos hecho una- hasta una]

A: -----]

C: página web y to

A3: CejEleva [Asiente]
[mm]

(Conv5, 0:16:22)

⁸⁹ gesto de que le explota la cabeza

En cambio, en el ejemplo 5, D reacciona con dos asentimientos al mismo acto de A según se desarrolla. El de A es un acto breve, no podemos pensar que sean dos actos: “((digamos que)) va con lo de la entrada negra y el resto gris”. El primer asentimiento de D demuestra que valida la explicación con la que A va a continuar tras haber intentado D contribuir a ella para que C entienda que el personaje del que hablan no va siempre con la misma peluca; el segundo asentimiento confirma o muestra acuerdo con la explicación dada por A. Es uno de los casos encontrados en los que una intervención está formada por dos gestos:

(5)

C1: [ay yo creía-]

1A1: [a ver pero es- pero el mismo] [tono ((digamos que))]

D1: [el estilo]

A: =va [con lo de la entrada ne]

D2: [Asiente-----]

A: =gra y el resto gris [¿no?]

D: [Asiente]

(Conv2, 00:14:22)

En el ejemplo 6, un acto de asentimiento verbal se descubre gracias al gesto coverbal. Antes del acto de preguntar, ambos reaccionan a la misma información “el domingo me dijo porque tú trabajabas el sábado me dijo Noa” (Conv 3 0:15:12), así que no es una intervención distinta a la de la pregunta.

(6)

1C1: el domingo me dijo porque tú [trabajabas el [sá]bado]

2A1: [AsienteRep-----]
[mm]

C: = me dijo Noa

A: ¿pero este- es este ya?

Por lo visto en estos ejemplos, hay gestos que no suceden a la vez que el habla sino justo antes o después, pero forman parte de la misma intervención. En consecuencia, no se anotan como intervenciones independientes, sino coverbales. Como se ha visto en A1 y A3 del ejemplo 4 y 2A1 del ejemplo 6. La diferencia entre dos intervenciones, una gestual y otra verbal, y dos actos de una misma intervención, uno gestual y otro verbal, es si reaccionan a distintas partes del discurso del otro interlocutor.

Se ha observado que es común una retroalimentación gestual seguida de retroalimentación verbal, formando ambas partes de la misma intervención. Es decir, el oyente opta por asentir repetidamente a lo largo del turno del hablante y al final del turno el oyente cierre su retroalimentación con una expresión verbal como “mm” o “ya” (similar a los asentimientos del Ej. 5). En el Ejemplo 7, C asiente ligera, pero repetidamente durante 4,3 segundos y, tras el tonema ascendente de A, C hace “mm”:

(7)

1A1: y porque no has visto los cables por detrás están todos escondidos por la pata
[/ o sea está- tiene un panel- panel secrete- detrás en- meto los cables por unos

C1:[AsienteRep-----]

A: =canale]

C: -----]

A: =tas↑

C: mm

(Conv3, 0:12:12- 0:12:17)

En estos casos, el asentimiento no se ha contabilizado como gesto independiente, porque forma parte del mismo acto que “mm”; no de manera solapada, pero seguida. Tampoco se ha contabilizado en el análisis como coverbal que modifica la segmentación porque el “mm” verbal ya validaba el turno en el análisis oral.

Combinaciones similares también suceden con elevamiento de cejas y “mm”, que reaccionan a un mismo acto (Véase Ej. 8). Las cejas pueden utilizarse para marcar que se está prestando atención al inicio.

(8)

1C1: [pues:] aparte de nosotros dos y Luis↑ ee Jaime y Ricky // y ya está / el Sovrino
(RISAS) ahora os ense[ño la-]

2D1: [te[nía o]tro cumple]

3A1: [CejEleva-----]
⇒D

D: =ños

A: mm // pero ¿que le avisó Devi?

(Conv. 2, 0:04:00-0:04:01)

Un caso muy particular en el que ha sido difícil decidir si el gesto es independiente o coverbal es el Ejemplo 9, en el que 1C1 está hablando con A, pero al mismo tiempo hace un gesto modal solo dirigido a B, un guiño, en relación con lo que le cuenta a A,

creando complicidad con B. 2B1 reacciona y C2 termina ese diálogo secundario entre C y B del que A no se da cuenta.

(9)

1C1: [ahí a las últimas de la]s- las últimas de los casos lo pones encima y les guardas las patas y cuando te cambies de casa OjoCierra
BocaOtro⁹⁰
⇒B

A1: mm
⇒∅

2B1: Asiente
CejEleva
⇒C

C2: Sonrisa
⇒B

A: [pero vamos]

C: [te pones] las patas

(Conv. 3, 0:18:04-0:18:05).

Aunque el guiño solo esté dirigido a B, podríamos considerarlo coverbal porque está relacionado pragmáticamente con la intervención hacia A, con una función modal. Sin embargo, el guiño abre una secuencia en la que 2B1 corresponde a esa complicidad, C2 termina ese intercambio y después concluye la oración condicional de la intervención 1C1. No podemos considerar todas las emisiones de C una sola intervención porque C2 es una reacción a B, entonces lo consideramos un caso especial en el que C continúa con la intervención 1C1 dirigida a A (este ejemplo da cuenta de la importancia de la orientación corporal en la segmentación §6.1.3.). Respecto al guiño, por ese diálogo interno se ha planteado que pudiera ser un gesto independiente, pero finalmente la función pragmática ha primado.

Criterio de duración y criterio de intensidad

También han surgido dudas con gestos que se mantienen de forma muy extendida en el tiempo o con los que tienen poca amplitud de movimiento y son casi imperceptibles. A veces, las expresiones faciales, especialmente la sonrisa, se extienden más allá de la reacción a un turno concreto; es decir, durante los siguientes turnos se mantiene sin

⁹⁰ Una comisura hacia abajo al realizar un guiño de ojo exagerado.

cambios. Podemos considerar que el momento álgido se produce al inicio y luego se queda la fase de mantenimiento del gesto hasta que cambie a otra expresión o se relaje a una expresión neutra de reposo. En los casos en los que el gesto facial se ha extendido durante muchos turnos, hemos decidido anotar solamente la primera fase e ignorar su mantenimiento porque entendemos que se hace reacción a un primer turno y no ratifica los siguientes, aunque se pueda deducir que está prestando atención, como sucede con la mirada y la postura.

Respecto a los movimientos casi imperceptibles, como un asentimiento en el que la inclinación de la cabeza hacia abajo tiene muy poca amplitud, se ha decidido anotarlos aunque no tengan la misma intensidad que en otros contextos. La amplitud/intensidad podría ser una variable para estudiar los gestos que nosotros no vamos a tomar en cuenta para este análisis, pero, sin importar el grado, son movimientos que en la grabación se perciben y que posiblemente los interlocutores también perciban y tengan en cuenta en la interacción.

La intervención gestual con dos gestos independientes

Siguiendo la segmentación temporal de los gestos de Kendon (2004) (§3.4.2.1.), la unidad superior del gesto, *unidad gesto*, puede estar formada por una o más *frases gesto*. Cada frase gesto incluye un solo punto álgido que expresa una acción significativa. La sucesión de frases que forman parte de una unidad u otra se distingue por el movimiento de recuperación. Varios puntos álgidos sucesivos sin que se recupere la posición de reposo de la parte del cuerpo forman una unidad gesto.

Generalmente, en el corpus, los gestos en posición independiente de intervención son la única frase gesto de una unidad gesto. Pero también hay una minoría de unidades gesto en posición independiente formada por una combinación de unidades gesto o de frases gesto. Han sido 14 intervenciones en las que, generalmente, se han combinado dos gestos independientes del habla. Pero también hay un caso de una intervención especialmente compleja que combina 6 gestos.

En el ejemplo 10, el asentimiento de A2 tiene función de seguimiento de la justificación de C, mientras que la sacudida de cabeza de la misma intervención muestra acuerdo con la información de C de que él dijo que no valían esos canelones. Aunque reaccione a dos partes de la misma intervención, no consideramos que “porque una vez

compré conelones de ahí y dijiste no valen” sean dos actos, así que A da dos muestras de retroalimentación a lo largo de un acto largo.

(10)

1C1: [((igual) que)) los canelones que- porque una vez compré ca[nelones de ahí y]=

A2: [Asiente-----]

C: =((dijiste)) *no valen* [/ y [la] musa]=

A: [CabSacude]

B: [[m]m]

C: =ka es una cosa más buena

(Conv1. 00:05:52, 1A.Cab.1 y 1A.Cab.2)

Dado que dos gestos pueden reaccionar a un solo acto, esto hace que ambos formen parte de una misma intervención. Sin embargo, dentro de la estructura de intervención, la cuestión es si forman parte de un mismo acto según Val.Es.Co. Nosotros hemos catalogado cada gesto como un acto porque, generalmente, cada gesto tiene una función distinta (en §6.2.2. se muestran las funciones que combinan estos gestos), pero no nos atrevemos a decir que uno esté subordinado a otro, pues sin la dependencia del habla todos los gestos son en cierta medida performativos. También se enfrentó a este problema el Grupo Val.Es.Co. (Pons Bordería 2022) con la combinación de partículas con funciones procedimentales en posición independiente, que típicamente son subactos de un acto sustantivo director. El criterio del grupo entonativo fue el que utilizaron para segmentar los subactos como parte de actos distintos.

También hay dos casos (ejs. 11 y 12) en los que los dos gestos tienen la misma función y casi la misma forma, pero se ha tomado la decisión de separarlos en dos actos porque entre ambos hay una pausa en la que se recupera la posición de reposo y porque reaccionan a dos partes del mismo turno.

En el ejemplo 11, B repite el elevamiento de cejas, bien porque cree que C no le ha visto, bien porque C se ríe de lo que va a contar y B lo anima así a que continúe con el turno. Son dos gestos con diferente *stroke*, pues hay una pausa entre ambos de 1397ms.

(11)

D1: yo- yo dije- YO PEN [SA BA]

1C2: [inci][so]

B3: [CejEleva]

C: (RISAS) [(RISAS)]

B: [Ce]Eleva]

C: con lo de Grindr

(Conv2, 0:08:05)

En el ejemplo 12, aunque ambos gestos sean asentimientos, tienen diferente *stroke*, hay una pausa entre uno y otro de 982 ms. Pero sigue siendo la misma intervención porque reacciona a dos partes del mismo turno, el segundo gesto reafirma la función de seguimiento del primero ante las dificultades de A para explicarse: autoreparación con *o sea*.

(12)

1A1: caro es que digo tienes profesor[es es]=

C1: [Asiente]

A: =pecializados en- o sea [los- los profesores de educación especial]=

C: [AsienteRep-----]

A: que ya- que estén como metidos en el- en el mundo del colegio del instituto y todo eso

(Conv5, 0:15:23-0:15:29)

La multimodalidad y la interacción entre más de dos hablantes

Los elementos visuales, y especialmente la orientación corporal, han dado muestra de estructuras particulares en las conversaciones de más de dos hablantes que no son posibles con la segmentación oral. En el ejemplo 13, 1B1 lanza una pregunta y C se sincroniza con la respuesta de D, mirándola, repitiendo “ay” y asintiendo a su evaluación, reaccionando a D realmente. 3C2 ya es otro turno porque contesta genuinamente a la pregunta desde su punto de vista.

(13)

1B1: [bueno vamos a ver] tenemos- venga vamos haciendo resumen ¿qué tal- qué te- qué- qué opinión te merece Hugáceo Crujiente?

2D1: ¡ay [me encanta!] [(RISAS)]

C1: [ay (RISAS)] [Asiente-]
⇒D

D: es geni[al (RISAS)]

3C2: [yo no lo he visto to]do todo o [sea ella s- ella sí]

D: [es genial] e[s genial]

C: [yo he visto a ca]chicos
pero // no sé me- me cayó bien

(Conv2 0:05:49-0:05:52)

Esta intervención coral (Lerner 2002) puede deberse a que C y D son pareja. Esto tiene implicaciones durante la conversación porque su marco epistémico es compartido. Entonces, aunque una pregunta o una valoración se dirija a D, a menudo C reacciona también porque conoce la respuesta previamente. En esta situación, la figura del individuo se diluye en la pareja; no se está hablando con una persona, sino con dos. Aunque la orientación corporal solo vaya hacia uno de ellos, ambos se sienten libres de reaccionar por su pareja⁹¹.

En el Ejemplo 14, la orientación de 8A2 demuestra que no expresa cambio de estado epistémico por la respuesta dada por 7B4, sino que valida la opinión de C sobre el precio del tatuaje. Aun así, 7B4 se mantiene como turno porque la respuesta permite la continuación de la secuencia.

- (14)
- 1C1: ¿y qué os ha cobrado?
- 2B1: cien// [¿qué te parece?]
- 3C2: [CabInclina-----]
 [OjosAbre-----]
- 4B2: [es ca]ro ((¿verdad?))
- 5C3: sí
- B3: ya
- 6A1: ¿el qué? ¿el tatuaje?
- 7B4: Asiente
- 8A2: [mm]**
 [Asiente]
 ⇒C
- 9C4: hos[tia es caro]
 ⇒B
- A3: [Asiente--]

(Conv5 0:19:53)

Por otro lado, hay que considerar que un hablante puede dirigirse a varios oyentes a la vez sin elegir a uno como el principal. Entonces, un cambio en el eje de orientación no tiene por qué indicar una intervención distinta que se dirige a otro interlocutor distinto.

⁹¹ Véase Goodwin (1981: 156-159) sobre los comportamientos que se producen al hablar en presencia de un oyente que conoce la información.

En el ejemplo 15, B hablaba con C principalmente, pero gira hacia A; no es un turno distinto porque forma parte del mismo razonamiento, que puede ir hacia cualquiera de los dos.

(15)

1C1: [dos] y con su marido seis (1,5) [el t-]

2B1: [jy] el papá se lleva cuatro! // ah claro cuatro y [dos seis]
 ⇒C ⇒A

3A1: cuatro] con eel- con el primo [claro

(Conv. 4 0:12:39-0:12:40)

5.4.4. Anotación de cambios en la segmentación

Aunque dividamos las fases de segmentación y la de análisis para explicar de manera organizada los procesos, realmente, la segmentación visual se realiza al mismo tiempo que la anotación de los cambios. Es decir, a medida que una intervención cambia su tipología (por ejemplo, de Ir a Ir-i), en la línea de cambios, se crea una anotación alineada a la intervención modificada y que indica la causa del cambio. Para cada hablante hay una línea de cambios (A_Cambios, B_Cambios, etc.). El comportamiento de cada hablante se observado por separado con el objetivo de no perder de vista ningún movimiento.

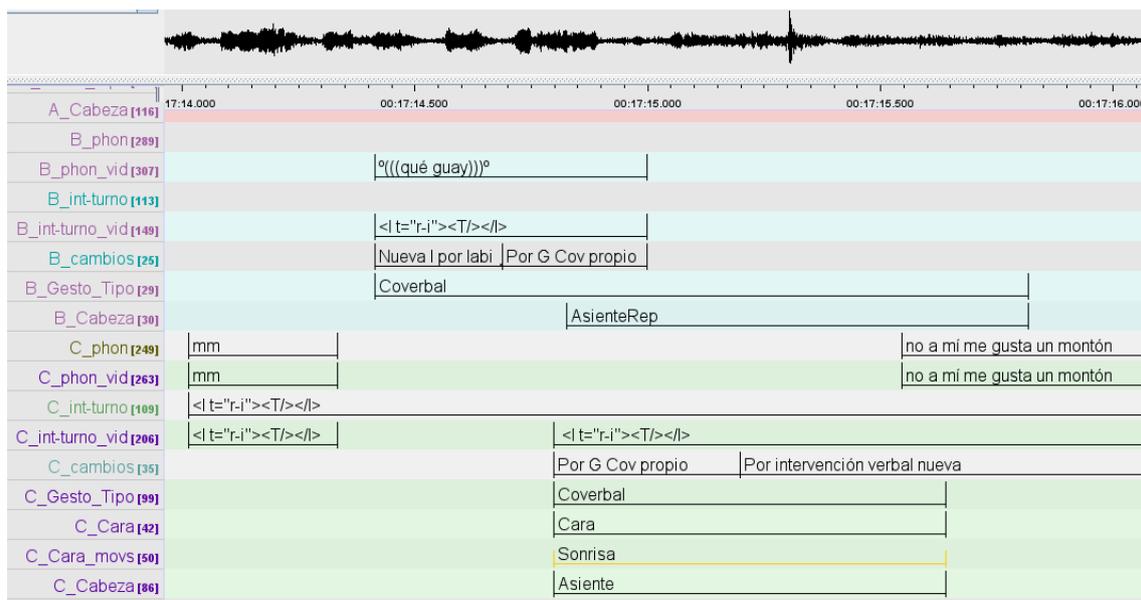


Imagen 11. Ejemplo de anotación de cambios de segmentación en ELAN.

En la imagen 11, la intervención de B no se percibe en la segmentación oral, en cambio, está transcrita en B_phon_vid y segmentada en B_int-turno_vid. En B_cambios se anotan las dos causas que generan esta intervención: la lectura de labios y el gesto coverbal propio, un asentimiento repetido. Esta nueva intervención repercute en la segmentación de C, cuya intervención se divide en dos. La segunda intervención tiene como causa la intervención verbal nueva de B y su propio gesto coverbal, la sonrisa y el asentimiento, que reacciona a la evaluación de B.

En la Tabla 20 se muestra el vocabulario cerrado utilizado para anotar las causas. Hay más anotaciones para los efectos de la lectura de labios porque al principio se iban a contar como variables distintas, pero finalmente, se decidió aunarlas todas en la lectura de labios porque realmente es la misma causa con distintos efectos.

	Cambio en la propia intervención	Cambio en otra intervención
Lectura de labios	Nueva I por labios	Por intervención verbal nueva
	Cambio Hablante por Labios	Por Cambio Hablante ajeno
	Por Transcripción	Por Transcripción ajena
Orientación corporal	Por Orient propia	Por Orient ajena
Gesto coverbal	Por G Cov propio	Por G Cov ajeno
Gesto independiente		Por G ind
Acción		Por Acción

Tabla 20. Anotación de cambios en la segmentación en ELAN.

Todos los cambios producidos por el gancho se han anotado con el término “Gancho” precediéndolos para poder separar estos casos del resto.

Hay una distinción en el tratamiento de las causas: mientras que la lectura de labios y la orientación corporal no se segmentan en sí temporalmente, tan solo sus efectos; las acciones se segmentan en la línea de observaciones Obs_vídeo y los gestos se segmentan en la línea Gestos_tipo de cada hablante, ya que los consideramos elementos que forman parte de las unidades conversacionales. En esta línea de gestos, se distinguen los dos tipos: gestos independientes y gestos coverbales. Los gestos coverbales se segmentan y modifican la segmentación estructural o social de la intervención a la que afectan; los gestos independientes generan nuevas intervenciones en la línea de segmentación visual, con las mismas etiquetas de intervención y turno (Imagen 12). Debido a esto, no es necesario repetir en la línea de cambios “Por G Ind propio”; a través de las anotaciones de Independiente de la línea Gesto_Tipo, podemos extraer todos los

gestos independientes y su segmentación para los análisis (excluyendo los del gancho). La etiqueta “Por G ind” en la líneas de cambios significa que el gesto independiente, además de crear su propia intervención, también modifica a otra intervención verbal existente.

A_Cabeza [116]	500	00:19:52.000
B_phon [289]		
B_phon_vid [307]		
B_int-turno [113]		
B_int-turno_vid [149]		< t="r-i"><T/></ >
B_cambios [25]		
B_Gesto_Tipo [29]		Independiente
B_Cabeza [30]		AsienteRep

Imagen 12. Segmentación de gesto independiente.

En la exportación de los datos a Excel, se utilizan la línea de segmentación oral, la de segmentación visual y la de cambios de cada hablante por separado para observar los cambios estructurales y sociales y hacer el recuento. En el ejemplo de la Imagen 13, un gesto independiente que reacciona a la pregunta “¿qué os vais al campo aquí?” convierte la Ir en Ir-i turno. En Excel hemos creado una nueva columna (Cambios) para especificar el tipo de cambio y analizar la variable cambio (Tabla 21) en relación con la variable causa (A_cambios). En futuros estudios esta anotación podría hacerse directamente en ELAN y sustituir el nombre de la línea A_cambios por A_causas.

Tiempo	Tiempo	A_int-turno	A_phon_vid	A_int-turno_vid	Cambios	A_cambios
0:21:54	0:21:55	< t="r"></ >	¿qué os vais al campo aquí?	< t="r-i"><T/></ >	I>T	Por G ind

Imagen 13. Ejemplo de exportación de cambios entre segmentaciones.

Nueva intervención (no validada)	Nueva I
Nuevo turno	Nueva T
Modificación de validación: de intervención a turno	I>T
Modificación de validación: de turno a intervención	T>I
Sin modificación de validación: se mantiene como turno	T>T
Sin modificación de validación: se mantiene como intervención	I>I

Tabla 21. Anotación de cambios dialógicos.

En la anotación la variable alcance se aplica a la causa y no al cambio. Esto es algo que hemos corregido en el análisis, pero los datos de los que partimos están organizados así y en un análisis futuro cambiaríamos las anotaciones para facilitar la tarea.

En resumen, el esquema de anotación en ELAN del análisis de cambios en la segmentación es el de la Tabla 22.

DOMINIO	ATRIBUTO	VALOR
Transcripción oral	A_phon	(transcripción oral)
	A_palabras	(transcripción <i>tokenizada</i>)
	Observaciones	(escritura libre)
Segmentación oral	A_int-turno	Intervención iniciativa, Intervención reactiva (turno), Intervención reactiva (no-turno), Intervención reactivo-iniciativa, Intervención independiente, Intervención ininteligible.
Transcripción visual	A_phon_vid	(transcripción oral)
	A_palabras_vid	(transcripción <i>tokenizada</i>)
	Obs vídeo	(escritura libre)
Segmentación visual	A_int-turno_vid	Intervención iniciativa, Intervención reactiva (turno), Intervención reactiva (no-turno), Intervención reactivo-iniciativa, Intervención independiente.
	Obs_segmentación	(escritura libre)
Análisis de cambios en seg. visual	A_cambios	Nueva I por labios, Cambio Hablante por Labios, Por Transcripción, Por Orient propia, Por G Cov propio, Por intervención verbal nueva, Por Cambio Hablante ajeno, Por Transcripción ajena, Por Orient ajena, Por G Cov ajeno, Por G ind, Por Acción
	A_Gesto_Tipo	Independiente, Coverbal

Tabla 22. Esquema de anotación en ELAN del análisis de cambios en la segmentación

5.5. Análisis de los gestos

En el primer análisis se consideran varias causas visuales que explican los cambios estructurales y sociales en la segmentación de la conversación. Algunas de estas causas visuales no son unidades de la conversación (lectura de labios, orientación corporal durante intervención verbal y acciones extralingüísticas); en cambio, los gestos independientes y los gestos coverbales sí pueden formar parte de unidades conversacionales: los gestos independientes como actos e intervenciones y los gestos coverbales como actos o subactos (§5.4.2). Sin embargo, esta característica de la posición no explica por sí sola la manera en la que estos gestos modifican la segmentación. La función pragmática que cumplen y la forma corporal elegida para llevarla a cabo son

variables que nos ayudan a caracterizar los gestos que afectan a la segmentación dialógica. De la misma manera que la función del marcador discursivo se vincula a una cierta posición en un tipo de unidad determinado y esto permite establecer límites a sus posibilidades combinatorias y distintos paradigmas funcionales (Briz y Pons 2010), la misma relación podremos encontrar en el análisis de los gestos. Por consiguiente, en este segundo análisis se relacionan la posición del gesto en la intervención, la función y la forma de los gestos y, en el caso de los independientes, qué tipo de intervención crean.

El punto de partida por el que seleccionamos los gestos que vamos a analizar es la posición. No partimos de un gesto con una forma concreta y nos preguntamos qué función desempeña en el discurso; ni siquiera buscamos distintas formas que cumplen una función concreta; sino posiciones en la secuencia. Es un análisis que podríamos relacionar con la *gramática sensible a la posición* (Schegloff 1996, Fox et al. 2013: 739): para una posición secuencial hay un conjunto restringido de formas que se pueden utilizar para llevar a cabo la acción relevante en esa ubicación. La selección de estas formas depende del contexto secuencial y, a su vez, una misma forma puede tener significados distintos en distintas posiciones secuenciales. No podemos entrar en un gran nivel de detalle en el análisis de cada forma o función, pero esperamos encontrar algunos patrones con los que plantear hipótesis sobre qué funciones desempeñan qué formas afectando a la dimensión estructural y social.

Para averiguar las funciones que cumplen nuestros gestos, vamos a utilizar *Methods of Gesture Analysis*, un método basado en la forma del gesto para reconstruir el significado del gesto sistemáticamente (Bressemer, Ladewig y Müller 2013, Müller 2010b, Müller, Bressemer y Ladewig 2013, Müller, Ladewig y Bressemer 2013). Cada aspecto de la forma del gesto se trata como potencialmente significativa y no aleatoria. Este método es especialmente útil para el análisis de los gestos recurrentes (§3.4.1.2.), que en su mayoría son pragmáticos. Se divide en cuatro fases (Bressemer, Ladewig y Müller 2013, Ladewig y Bressemer, 2013):

- 1) análisis de la forma,
- 2) análisis de la estructura secuencial de los gestos en relación con el habla y otros gestos,
- 3) análisis del contexto local de uso de los gestos, es decir, su relación con aspectos sintácticos, semánticos y pragmáticos del habla, y
- 4) distribución de los gestos en diferentes contextos de uso.

La primera fase se trata de la determinación de las unidades gestuales como signos motivados por sus características semántico-cognitivas, sus límites temporales, la conexión entre gestos y una descripción detallada de la forma para detectar patrones, sin tener en cuenta el habla (Bresse 2021: 14-15). En nuestro caso, si bien todos los gestos independientes se van a analizar, no todos los gestos coverbales entran en el análisis. Nuestra selección se basa en que la información visual modifique la segmentación de unidades conversacionales efectuada con la información auditiva, así que necesitamos escuchar las emisiones verbales que acompañan a los gestos. La forma de los gestos manuales se describe a partir del sistema de anotación que utiliza MGA; Linguistic Annotation System for Gestures (LASG) (Bresse 2013b; Bresse, Ladewig y Müller 2013) (§4.4.2) adaptado (§5.5.4); el resto de partes del cuerpo mediante MUMIN (§4.4.1) adaptado (§5.5.3), pues es una anotación ya creada con el objetivo de describir gestos interactivos.

En la fase 2, el análisis de los gestos es sensible al contexto. Se observa cómo los gestos se relacionan temporalmente con el habla. En nuestro caso, aprovechamos que ya se ha anotado la posición de los gestos como *independientes* o *coverbales* y el tipo de intervención que el gesto independiente genera estructuralmente (Ii, Ir, Ir-i, Iind) y socialmente (turno/intervención). Luego, en la fase 3, se analizan las funciones semánticas, pragmáticas o sintácticas que los gestos tienen en el enunciado en cada caso particular (selección de funciones en §5.5.4); para después, en la fase 4, comparar la función de estos gestos en diferentes contextos de uso y distinguir así el núcleo semántico del gesto de su significado local y de sus variantes contextuales (Bresse y Müller 2014a: 1577-8).

Por último, una vez establecidas las posiciones, formas y funciones más comunes, podemos relacionar estas variables: las funciones generales y específicas más comunes en posición independiente o coverbal, las funciones más comunes según su validación (turnos o intervenciones) y las formas más comunes para estas funciones.

Respecto a las funciones que realizan los gestos en las posiciones de nuestro interés, nuestra hipótesis es que los gestos independientes pueden ser gestos referenciales, interactivos y performativos, mientras que los rítmicos, los discursivos y los modales siempre están modificando mediante diferentes valores pragmáticos a un enunciado verbal.

Se podría decir que el gesto propio de la posición independiente es el performativo, pues, además de contribuir a la identificación del acto verbal (posición coverbal de acto), puede contener en sí mismo la fuerza ilocutiva que le permite ser un acto gestual, bien como parte de una intervención verbal, bien como único acto de una intervención. Entonces, mediante un gesto, el interlocutor podrá emitir un turno en el que pregunte, responda, evalúe, ofrezca, exprese sorpresa, ignorancia, desacuerdo, etc.

También Kendon (2004) explicaba que un gesto referencial (icónico o deíctico) en posición independiente de intervención se utiliza para colaborar en el turno de otro hablante o para solicitar u ofrecer un objeto; de manera que no solo puede contribuir semánticamente a un acto verbal, sino que por sí mismo realiza un acto comunicativo que tiene sentido en una secuencia. La performatividad del gesto en este sentido es una cuestión de posición en la secuencia, ya tenga un valor semántico o pragmático.

Lo mismo ocurre con los gestos interactivos. Estos suelen acompañar a un turno para indicar la gestión del mismo (si se toma, si se mantiene, si se cede, etc.), pero también el emisor puede utilizarlos independientemente, precisamente cuando tiene el rol de oyente y decide no tomar el turno. Por consiguiente, mediante gestos de retroalimentación o de cesión del turno el oyente emite intervenciones gestuales que dejan clara su pretensión de no tomar el turno de habla. Igual que el gesto referencial, el gesto interactivo en posición independiente de intervención es también performativo por cargar con toda la fuerza ilocutiva del acto y la intervención. Esta función interactivo-performativa es una cuestión que también se ha estudiado en Va.Es.Co. con las partículas discursivas que funcionan habitualmente como subactos adyacentes, interpersonales o de otro tipo, que afectan a un subacto sustantivo. Pueden aparecer también en posición independiente de intervención, por lo que no modifican al contenido proposicional de su acto, sino que por sí mismas forman un acto procedimental que depende de la intervención de otro hablante con la que interactúa (Pons Bordería 2022).

En la bibliografía el único tipo de gesto que se contempla en posición independiente es el performativo. Ya que en los casos descritos esa posición convierte al gesto en acto, se podrían clasificar todos como performativos simplemente; pero consideramos que es mejor utilizar una etiqueta doble, como referencial-performativo o interactivo-performativo, porque resulta más descriptiva del tipo de gesto.

Además, otra de nuestras hipótesis es que la mayoría de los gestos independientes van a ser gestos interactivos porque en el estudio preliminar (Cabanés Pérez 2020) la

mayoría fueron intervenciones de apoyo emitidas por el oyente, 22 de 25 intervenciones gestuales. Cumplían una función interactiva de retroalimentación o expresaban algún tipo de evaluación; aunque hubo más de las primeras que de las segundas, no se contabilizó esta distinción. En el estudio actual, con una mayor muestra de gestos, queremos distinguir la función concreta que desempeña la intervención reactiva del hablante en el contexto. Respecto a las restantes 3 intervenciones gestuales de 25 del estudio piloto, eran información o evaluación solicitadas por un primer par adyacente verbal. En la muestra del estudio actual, aun con una proporción similar, esperamos encontrar más ejemplos de gestos que sean la primera o la segunda parte de un par adyacente, los cuales cumplen más variedad de funciones que las intervenciones de apoyo.

Los gestos que se clasificaron como turno eran respuestas a preguntas explícitas o enunciados declarativos, por lo que ocupaban la segunda parte de un par adyacente. Hubo tres turnos gestuales en esa conversación: una pregunta fue respondida con movimiento de la cabeza afirmativo; una declaración recibió como reacción un *shrug* simple encogiendo de hombros para expresa un acuerdo poco comprometido; y otra declaración recibió un *shrug* escéptico más complejo (elevamiento de cejas, comisuras de labios hacia abajo y encogimiento de hombros). Aun así, en el análisis social más estricto que vamos a realizar en este estudio (§2.3.4), solo se aceptarían como turno estas evaluaciones mediante *shrugs* si recibieran una reacción posterior. Actualmente las clasificaríamos como Ir no turno, así que la respuesta a la pregunta sería el único turno gestual. Por otro lado, en aquella conversación no se encontraron gestos con una función que encajara como primera parte del par adyacente, validada o no validada, pero esperamos encontrarlos en un corpus más amplio.

Por tanto, nuestra hipótesis es que la mayoría de los gestos independientes van a cumplir funciones típicas de las Ir del oyente (seguimiento, evaluación) y la minoría, funciones que aparecen en pares adyacentes. Nos interesa encontrar funciones exclusivas de cada posición (independiente y coverbal) y comunes a ambas; posiblemente, la función interactiva sea común en ambas posiciones. Los gestos coverbales pueden cumplir todas las funciones de la tipología: se vinculan con el significado de las palabras, con la relación entre los enunciados, con la entonación, con el tipo de acto o con la gestión de la interacción. Puesto que nuestro interés está en los gestos coverbales que sean la única muestra de validación de un turno previo (el contenido verbal es insuficiente para ello), nuestra hipótesis es que su función será interactiva, es decir, centrada en la gestión

interactiva de los turnos. No obstante, no estamos cerrados a la posibilidad de que gestos con otra función puedan validar un turno verbal previo.

Respecto a la forma de los gestos, así como los gestos independientes más comunes en el estudio preliminar expresaban seguimiento del oyente hacia el hablante, el movimiento utilizado para esta función fue la inclinación de cabeza de arriba abajo, es decir, el asentimiento. Pero también hubo sonrisas, sacudidas de cabeza y *shrugs* más y menos complejos. Esperamos encontrar esta dominancia del asentimiento en un corpus más amplio, pero también en menor frecuencia variedad de otras partes del cuerpo y de configuración de movimientos. Las expresiones faciales y los movimientos de cabeza están más relacionados con aspectos interaccionales, mientras los gestos manuales son a menudo representacionales (Tabacaru 2019: 5-6). Puesto que esperamos una mayor frecuencia de funciones interactivas en los gestos coverbales, presumiblemente, también la cabeza podría ser la parte del cuerpo más utilizada en esta posición. Según la bibliografía, la orientación de la cabeza hacia el interlocutor es más frecuente durante el rol de oyente (Kendon 1967) (§3.3.2), así que buscaremos su relación con la Ir.

En definitiva, esperamos que el uso del modelo Val.Es.Co. nos permita observar patrones de forma y función no vistos anteriormente en el estudio dialógico de los gestos.

5.5.1. Variable posición

En el primer análisis de este estudio (§5.4) la posición de los gestos forma dos causas para el cambio en la segmentación: los gestos independientes y los gestos coverbales. Puesto que en este segundo análisis buscamos la función y la forma de estos gestos, la posición se convierte en la variable dependiente.

La posición de los gestos ya se ha anotado en la línea Gesto_Tipo de cada hablante con los valores *Independiente* y *Coverbal*. Esta anotación sirve tanto para las causas del primer análisis como para la posición del gesto en el segundo análisis. Asimismo, cada gesto independiente se segmenta en ELAN como una intervención más de la conversación y en la línea Int-turno_vid se anota el tipo de intervención que es (Ii, Ir-i, IrT, Ir, Iind). En el primer análisis, esta anotación servía para las variables de cambio estructural y cambio social, generaba los valores: Nueva intervención, Nueva intervención (no validada) y Nuevo turno. En cambio, en el segundo análisis, gracias a esta anotación podemos ver

qué tipo de unidad intervención ocupa el gesto independiente con más frecuencia y relacionar el tipo de intervención con la función del gesto.

		Variable tipo de intervención	
		Turno	Intervención
Variable posición	Gesto independiente	Intervención iniciativa (Ii)	Intervención independiente (Iind)
		Intervención reactivo-iniciativa (Ir-i)	
		Intervención reactiva turno (IrT)	Intervención reactiva no-turno (Ir)
	Gesto coverbal		

Tabla 23. Variable posición.

5.5.2. Variable forma

Respecto a las modalidades visuales utilizadas para desempeñar las distintas funciones de nuestra tipología, en §3 se han mostrado ejemplos de gestos realizados con la cabeza, la cara, los hombros y las manos. Estas cuatro partes del cuerpo van a ser consideradas variables formales de los gestos, así como la orientación del gesto (Tabla 24). No se consideran partes del cuerpo inferiores, como las piernas, las rodillas o los pies, porque los participantes de nuestro corpus permanecen sentados.

Variable forma	Partes del cuerpo	Cabeza
		Cara
		Hombros
		Manos
	Orientación	

Tabla 24. Variables de forma de los gestos.

Algunas partes del cuerpo permiten más variedad articulatoria que otras y, además, estas variantes se combinan entre sí. Las configuraciones de articulación se detallan en el sistema de anotación de la forma (§5.5.3.). Para simplificar el análisis de estas variables, ya sea para identificar las funciones de los gestos más frecuentes o para identificar las formas más frecuentes de cada función, primero se tendrá en cuenta la parte del cuerpo, luego, las formas de articulación más frecuentes y, después, el resto de formas más particulares.

5.5.3. Anotación de gestos: la forma

La anotación de la forma del gesto requiere dos parámetros principales: la temporalidad y la configuración. Respecto a la temporalidad, se indica de forma completamente distinta en ELAN y en papel. En ELAN, los límites del gesto se sitúan temporalmente desde el tiempo de inicio del gesto hasta su fin siguiendo los cambios en la imagen. La reproducción del fragmento y los comentarios coincidentes en otras líneas permiten observar la sincronía entre distintos modos de comunicación. Sin embargo, en papel no existe un marcador temporal que indique el momento en el que empiezan y terminan las intervenciones o los gestos. Son los símbolos de solapamiento los que indican la sincronía entre intervenciones o entre modos, al margen de la medición temporal.

Tanto en un formato como en otro, se ha decidido no fragmentar el gesto en sus fases (§3.4.2.1.) porque nuestro objetivo no está relacionado con la sincronía entre habla y gesto (la relación semántica entre gesto y habla o la relación entre gesto o prosodia), basta con identificar que el gesto está. La delimitación del gesto se ha establecido desde su fase de preparación hasta el fin del punto álgido (*stroke*), pues esto nos da información sobre el momento en el que el participante reacciona, y consideramos que en la fase de recuperación, posterior al punto álgido, el gesto ya ha terminado. Estas dos fases constituyen la unidad *gesture phrase*, que es suficiente para identificar los gestos y después valorar su función en la secuencia, tal como hace el sistema de codificación MUMIN⁹². La transcripción de la configuración del gesto se basa en la articulación y el movimiento que se produce en la fase del punto álgido.

Mediante ELAN en el eje vertical podemos utilizar varias líneas que correspondan a un modo de comunicación. Esto permite registrar cómo se coordinan y se relacionan temporalmente los modos de comunicación en cada momento de la interacción. Téngase en cuenta que mientras en la transcripción verbal Val.Es.Co. utiliza corchetes “[]” para indicar las palabras que se solapan, para la segmentación de gestos no hemos indicado con ningún signo con qué palabras se solapa el gesto. Esto se debe a que no hemos considerado la sincronía entre gesto y habla como variable para analizar los gestos⁹³. Sin

⁹² Visitada 30/08/2021: <<https://www.cst.dk/mumin/resources.html>>

⁹³ Bresse et al. (2013: 1098), en cambio, además de la línea de transcripción verbal general, utilizan una segunda línea de transcripción que se alinea con las fases del gesto, para saber qué palabras coinciden con cada fase.

embargo, en la transcripción en papel, a falta de una alineación temporal, sí es necesario indicar la coincidencia entre los elementos verbales y visuales.

Durante la transcripción de un gesto, se utilizan tres tipos de líneas de ELAN: una para clasificar su posición (Gesto_tipo), otra para describir su configuración y otra para indicar hacia qué interlocutor está orientado (Orientación). Para la configuración hay una o dos líneas para cada parte del cuerpo utilizada: cabeza (A_Cabeza), hombros (A_Hombros), cara (A_Cara y A_Cara_movs), manos (A_Manos_Conf y A_Manos). La orientación del gesto se aplica como parámetro común a todos los gestos (A_Gesto_Orient).

Vistos los sistemas de anotación más y menos detallados en §4, consideramos que para nuestro objetivo los atributos y valores que propone MUMIN (Allwood et al. 2007) son los más adecuados y proponemos la siguiente adaptación (Tabla 25).

MUMIN		Adaptación	
Atributo	Código de anotación de valor	Atributo	
			Vectores
HeadMovement	Nod, Jerk, BackUp, Tilt, SideTurn, Waggle, HeadBackward, HeadForward, Down, HeadOther	Cabeza	
HeadRepetition	Single, Repeated	Asiente, AsienteArriba, CabInclina, CabSacude, CabOscila, CabAtrás, CabAdelante, CabAbajo, CabOtro, AsienteRep	
Shoulders	Shrug, ShouldersOther	Hombros	
Eyes	X-open, CloseOne, CloseRepeat, EyesOther	Cara	Ojos
Eyebrows	Frown, Raise, BrowsOther		Cejas
MouthOpenness	OpenMouth, CloseMouth		Boca
MouthLips	CornersUp, CornersDown, Protruded, Retracted, LipsOther		
Face	Smile, Laughter, Scowl, FaceOther	Cara	Sonrisa, Risa, CaraOtro
BodyInterlocutor	BodyToInterlocutor, BodyAwayInterlocutor	Orientación	
		aA, aB, aC, aD, RetiradaDesdeA, RetiradaDesdeB, RetiradaDesdeC, RetiradaDesdeD, aNadie	

Tabla 25. Adaptación de MUMIN para la anotación formal de los gestos.

Además de la traducción de la codificación, se han modificado algunos valores. Para los valores de la cabeza, Jerk se ha traducido como AsienteArriba, a falta de una traducción literal en español. SideTurn se tradujo inicialmente como CabGira, pero finalmente se ha utilizado CabSacude porque no estaba claro si se referían a un giro de cabeza que desvía la orientación de la cabeza o a una sacudida de derecha a izquierda. Como la orientación del gesto tiene su propio atributo y la sacudida ha aparecido varias veces en el corpus, se ha optado por el valor CabSacude. Down/CabAbajo es el único valor que se ha tomado de la versión de MUMIN de Alwood et al. (2007), donde se utilizan más valores de cabeza que especifican detalles como la velocidad o la repetición. El atributo de repetición del movimiento de cabeza se ha añadido como un valor más para la cabeza, pero solamente para el asentimiento (AsienteRep), ya que es el gesto de cabeza más frecuente y quizás pueda marcar alguna distinción funcional. Como la fase de recuperación de un asentimiento vehemente puede confundirse con una repetición, se ha marcado como AsienteRep a los gestos de asentimiento realizados con tres o más cabezadas. Se ha incluido el encogimiento de hombros como gesto corporal, como en Navarretta y Paggio (2013), porque ya lo observamos en Cabanes Pérez (2020).

Los atributos Eyes, Eyebrows, MouthOpenness, MouthLips y Face de MUMIN se han unido en el atributo de Cara y cada parte de la cara es un vector. Dependiendo del número de líneas que queramos crear podemos usar una línea por cada atributo o factorizar todos los atributos en un vector, creando una sola línea (Blache et al. 2017). Para no crear tantas líneas por hablante, A_Cara indica la parte de la cara implicada en cada momento del gesto (o si son varias a la vez) y A_Cara_movs, que es una línea dependiente de la anterior, especifica los movimientos concretos. De esta manera, si anotamos Varios en A_Cara, en la línea A_Cara_movs se pueden incluir dos o más anotaciones que dependen temporalmente de la anotación Varios. En el ejemplo de la Imagen 14, el elevamiento de cejas se mantiene, pero el hablante añade las comisuras hacia abajo, que se solapan con el movimiento de cejas.

C_Gesto_Tipo	Independiente		
C_Cara [42]	Cejas	Varios	
C_Cara_movs	CejEleva	CejEleva	ComisurAbajo

Imagen 14. Ejemplo de anotación en líneas de cara en ELAN.

Todos los movimientos faciales de MUMIN se han traducido e incorporado excepto CornersUp y Scowl. El primero porque no hemos encontrado cuál es la distinción entre CornersUp y Smile. Se puede interpretar que la elevación de las comisuras de los

labios no siempre tiene que formar parte de una sonrisa, sino que se puede combinar con otro movimiento facial que no forme una sonrisa “con toda la cara”. Sin embargo, en nuestro corpus, no se ha encontrado ningún caso así; entonces, a posteriori, eliminamos esta forma de la codificación. Respecto al Scowl, es una cara de enfado o molestia que también se traduce a veces como *ceño fruncido*. Ya que en español no hay una equivalencia que describa la forma de esta cara, sino el sentimiento que se interpreta de ella, y ya tenemos CejFrunce para describir las cejas, optamos por eliminar Scowl y, si las cejas fruncidas se combinan con algún movimiento de boca, describir también este con los valores que tenemos para la boca. Hacemos notar que, si bien la risa se ha segmentado en la línea de transcripción oral con la convención de Val.Es.Co., (*RISAS*), en la línea de los gestos faciales solamente se ha transcrito si la única manera de captarla era mediante la imagen porque en el audio no se oye.

El corpus NOMCO danés, que también utiliza la codificación MUMIN, descartó utilizar los atributos de ojos y mirada debido a la calidad de las grabaciones y por el ángulo de los sujetos (Navarretta y Paggio 2017). Creemos que en nuestros vídeos se puede apreciar de manera suficiente los gestos realizados con los ojos, pero no la dirección de la mirada. En vez de esta, la orientación del gesto hacia el interlocutor es un atributo que se codifica en todos los gestos porque puede afectar al análisis funcional. Este atributo se inspira en BodyInterlocutor, pero en vez de anotar si la orientación es hacia el interlocutor o no, detallamos al interlocutor concreto y si el cuerpo se dirigía a él justo antes de desviarse.

MUMIN cuenta con varios atributos para describir los gestos manuales, pero su acotación de valores responde a que el sistema está pensado para gestos interactivos. Preferimos el sistema LASG (Bressemer et al. 2013) porque es más detallado y puede servir para gestos referenciales y pragmáticos. También LASG se ha adaptado (Tabla 26).

Respecto a los vectores de los gestos manuales, no es que pueda haber varios movimientos simultáneos, sino que la doble línea sirve para crear dependencia entre los seis parámetros que pueden describir la forma de la mano. A_Manos_Conf sirve como línea madre y en ella se anota la configuración de los dedos; en la línea A_Manos se anotan el resto de los parámetros (Imagen 15). Así reducimos 6 líneas en dos.

LASG		Adaptación		
Atributo	Código de anotación	Atributo		Código de anotación
			Vector	
Hand Shape	Fist, flat hand, 1-5 (fingers), stretched, bent, crooked, flapped down, connected, touching	Configuración de los dedos		Puño, mano plana, índice, corazón, anular, meñique, pulgar, estirado, laxo, curvado, torcido, perpendicular, conectado, tocando
Orientation	PU (Palm Up), PD (Palm Down), PL (Palm Lateral), PV (Palm Vertical), TC (Towards Center), AC (Away Center), TB (Towards Body), AB (Away Body) di (Diagonal)	Otros	Orientación de la palma	PalmaArriba, PalmaAbajo, PalmaLateral, PalmaVertical, HaciaCentro, HaciaFuera, HaciaCuerpo, HaciaFrente, Di
Position	Center-center, center, upper, upper right, upper left, right, left, lower, lower right, lower left		Posición	Centro-centro, Centro, Encima, EncimaDerecha, EncimaIzquierda, Debajo, DebajoDerecha, DebajoIzquierda.
Movement Type	Straight, arced, circle, zigzag, s-line, spiral...		Tipo de movimiento	Recto, Arqueado, Círculo, Zigzag, Eses, Espiral
Movement Direction	Right, left, up, down, away from body, towards body		Dirección de movimiento	ADerecha, AIzquierda, Arriba, Abajo, ACuerpo, AFrente
Movement Quality	Reduced, enlarged, decelerated, accelerated, accentuated		Cualidad de movimiento	Reducido, Alargado, Decelerado, Acelerado, Acentuado

Tabla 26. Adaptación de LASG para la anotación de manos.

C_Manos_Conf [12]	Mano Plana
C_Manos [39]	PalmaArriba Diagonal ADerecha

Imagen 15. Ejemplo de anotación en líneas de manos en ELAN.

Puesto que los gestos manuales son minoría en el corpus, optamos por transcribir la configuración del dedo con su nombre (pulgarc, índice, etc.), porque no es necesario que el lector tenga que recordar a qué hace referencia el número para un elemento que no es tan común (véase la Tabla 27 con representaciones de algunas configuraciones de la mano).

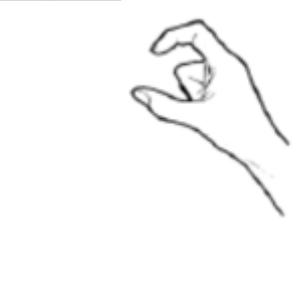
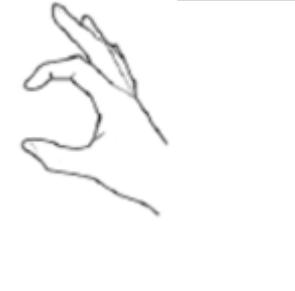
ManoPlana	ManoPlanaLaxa	ManoPlanaEstirada	Puño
			
ÍndiceEstirado	PulgarÍndiceCurvado	PulgarÍndiceTorcido	
			
Índice-MeñiquePerpendicular		PulgarÍndiceTocando	PulgarÍndiceConectado
			

Tabla 27. Forma de la mano. Selección de imágenes a partir de Bressemer (2013b)

Los valores de la orientación de la palma se combinan combinables entre sí: la PalmaLateral, la PalmaVertical, la PalmaArriba y la PalmaAbajo se orientan HaciaCentro, HaciaFuera, HaciaCuerpo o HaciaFrente (Tabla 28). Aunque los valores sean una traducción de LASG, no indican la orientación de la misma manera. La PalmaLateral y la PalmaVertical indican que la mano está orientada con los dedos en horizontal o en vertical, y los Hacias especifican la orientación de la palma en el eje X; mientras que PalmaArriba y PalmaAbajo indican la orientación de la palma y los Hacias la dirección de los dedos en el eje X. A todas las orientaciones de base se les puede añadir la orientación diagonal (Di). Para no confundir la orientación de la palma con la dirección del movimiento, en la traducción se ha seleccionado una preposición distinta (*a*).

	HaciaCentro	HaciaFrente	HaciaCuerpo	HaciaFuera
PalmaLateral				
PalmaVertical				
PalmaArriba				
PalmaAbajo				

Tabla 28. Valores de la orientación de la palma⁹⁴.

Igualmente, no todos los gestos manuales deben describirse mediante todos los parámetros mostrados, sino los que sean significantes en su contexto. Por ejemplo, si se mantiene la configuración de los dedos de un gesto previo, no se anota porque realmente esa configuración no es importante para transmitir la función, sino el movimiento que se realiza con él.

Las líneas de anotación de la forma tienen la anotación restringida a un vocabulario predeterminado desplegable (el que aparece en el esquema), con el objetivo

⁹⁴ Imágenes tomadas de los *queirotopemas* de la lengua de signos española (Rodríguez González 2003) (https://www.cervantesvirtual.com/obra-visor/lenguaje-de-signos--0/html/ffbeaf86-82b1-11df-acc7-002185ce6064_9.html)

de evitar errores de escritura. En la Tabla 29 se resumen las líneas de anotación de la forma.

DOMINIO	ATRIBUTO	VALOR
Análisis de gestos: anotación formal	A_Cabeza	Asiente, AsienteArriba, AsienteRep CabArriba, CabInclina, CabSacude, CabOscila, CabOtro
	A_Hombros	Encoge, HombrosOtro
	A_Cara	Cejas, Ojos, Boca, Cara, Varios
	A_Cara_movs	OjosAbre, OjoCierra, Parpadeo, OjosOtro, CejFruce, CejEleva, CejOtro, BocaAbre, BocaCierra, ComisurAbajo, LabiosFruce, LabiosRetrae, LabiosOtro, Sonrisa, Risa, CaraOtro
	A_Manos_Conf	Puño, ManoPlana, Indice, Corazón, Anular, Meñique, Pulgar, Estirado, Laxo, Curvado, Torcido, Perpendicular, Conectado, Tocando
	A_Manos	PalmaArriba, PalmaAbajo, PalmaLateral, PalmaVertical, HaciaCentro, HaciaFuera, HaciaCuerpo, HaciaFrente, Diagonal, Centro-centro, Centro, Encima, EncimaDerecha, Encimalzquierda, Debajo, DebajoDerecha, Debajolzquierda, Recto, Arqueado, Círculo, Zigzag, Eses, Espiral, ADerecha, Alzquierda, Arriba, Abajo, ACuerpo, Afrente, Reducido, Alargado, Decelerado, Acelerado, Acentuado
	A_Gesto_Orient	aA, aB, aC, aD, RetiradaDesdeA, RetiradaDesdeB, RetiradaDesdeC, RetiradaDesdeD, aNadie

Tabla 29. Líneas de anotación de la forma en ELAN.

Esta codificación de la forma de los gestos se ha diseñado para ser utilizada tanto en un programa de segmentación como en la transcripción leída en papel de manera clara. La incorporación de los gestos a la transcripción en papel requiere una serie de decisiones para distinguir entre gesto independiente y coverbal evitando la dificultad en la lectura.

El sistema de transcripción de gestos utilizado en este estudio se inspira en el de Mondada (2018), quien plantea una metodología para la relación temporal de los gestos con el habla, pero no unos códigos cerrados para los propios gestos, de manera que su sistema es adaptable a la transcripción de cualquier investigador —en nuestro caso, a la transcripción de Val.Es.Co. (Briz y Grupo Val.Es.Co 2002a)—. A continuación, enumeramos algunas decisiones tomadas para adaptar nuestros gestos y la alineación temporal a la transcripción de Val.Es.Co., sensible a la segmentación de intervenciones.

1) *Separación entre la información oral y visual.* A la transcripción del ejemplo 3 vamos a añadirle los gestos en el ejemplo 16. Los gestos están en color gris para que el lector distinga bien qué palabras se deben leer como transcripción verbal y qué palabras forman parte del código de anotación de los gestos. Mondada (2018) utiliza la negrita para la parte verbal y una fuente distinta para la gestual. Nosotros no queríamos perder la negrita para resaltar información al explicar un ejemplo, pero ambas son buenas opciones para quien quiera utilizar este sistema de transcripción.

(16a)

1C1: he aprendió más que en toda la carrera/ porque son todo apli[ca]

B1: [(joer)]

C: =ciones ee programas de cómo hacer cosas

(Conv. 5, 0:17:04, sin gestos)

(16b)

1C1: he aprendió más que en toda la carrera/ [porque]

B1: [AsienteArriba]

C: =son todo apli[[ca]

B: [(joer)]
[AsienteRep-----]

C: =ciones] ee programas [de cómo hacer cosas]

B2: [AsienteRep-----]

(Conv. 5, 0:17:04)

2) *Símbolos para indicar el solapamiento de habla y gesto.* El solapamiento del gesto con el habla se indica igual que el solapamiento verbal, con corchetes, pero de color gris. Esta elección se basa en la no duplicidad de signos para una misma función (solapamiento) y en que otros signos que podrían enmarcar una extensión de texto ya tienen otras funciones en Val.Es.Co.: el paréntesis para texto ininteligible y elementos paraverbales; la llave para la segmentación de subactos. En el ejemplo 16b, la emisión verbal de B1, ((joer)), está enmarcada por corchetes negros porque se solapa con la emisión verbal de C y también por corchetes grises porque se solapa con su propio gesto. Preferimos el corchete que el sistema de leyendas de Mondada (2018) para dar un símbolo a cada parte del cuerpo de cada hablante, lo cual no es económico ni claro para el lector.

3) *Diferencia entre gestos coverbales e independientes.* En el ejemplo 16b, se observan tres posiciones del gesto: dos asentimientos coverbales en B1⁹⁵, uno que sucede antes del contenido verbal y otro que sucede al mismo tiempo; y un asentimiento independiente en B2. El espaciado de líneas es importante para indicar que el gesto se produce solo o que acompaña a la emisión verbal del mismo hablante. El gesto que acompaña a ((*joer*)) tiene espaciado 0 con esta emisión, mientras que los otros gestos tienen 8 puntos de espaciado, al igual que el resto de las intervenciones verbales. En Mondada (2018) se repite el código del hablante en el margen izquierdo para indicar a quién pertenece el gesto solapado y los gestos independientes se solapan con un silencio del hablante en segundos en la línea verbal. Sin embargo, la codificación de intervenciones de Val.Es.Co. y este espaciado evita tanto la indicación temporal del silencio (que no es una pausa real, sino algo a lo que vincular el gesto) como que se tenga que indicar nuevamente al hablante y que así se confunda el gesto con una intervención nueva.

4) *La sucesión de emisión verbal y gestual en un mismo turno.* El primer gesto coverbal de B1 (ejemplo 17) está separado del resto de la intervención debido a la interrupción de C, pero si esta no existiera, aparecería en la misma línea que el habla, como ocurre en el ejemplo 5:

(17)

C: Asiente mm (1) pero ((el otro igual))

(Conv5, 0:05:18)

5) *Alineación temporal de habla y gesto.* En el ejemplo 16b, se comprueba que la extensión de la codificación del gesto no siempre va a coincidir con el tamaño del habla solapada, puede ser menor o mayor⁹⁶. El gesto AsienteArriba es más extenso que el *porque* de 1C1, pero los corchetes ayudan a distinguir la sincronía. Ambos AsienteRep se extienden durante más periodo de habla, así que agrandamos la anotación del gesto con guiones para que los corchetes se alineen, como hace Mondada (2018).

⁹⁵ La numeración de turnos e intervenciones que aparecen en los ejemplos de este trabajo no se corresponden con la conversación completa porque esta no se ha numerado. En cada ejemplo, el recuento empieza desde 1.

⁹⁶ El código prioriza el verbo porque es más breve: Asiente vs. Asentimiento, CejFrunce vs. CejFruñidas

Hemos intentado que los códigos de los gestos sean breves para que se entienda el alineamiento con el habla. Sin embargo, las manos suelen requerir más parámetros descriptivos que otras partes del cuerpo, en consecuencia, generalmente, se anota “Mano” y en una nota al pie se especifican sus valores; como en el ejemplo 18.

(18)

1C1: [¿y esos muebles qué vai]s a hacer? ¿moverlos pa'ca? [dices]

2B1: [no () esos] muebles los
sacaré aquí [y si nadie me dice]

C2: [Mano⁹⁷-----]

B: =algo esta noche se ((van a ir)) fuera

(Conv.3, 0:10:45)

6) *Alineación temporal de gestos de un mismo hablante.* En los ejemplos 4 y 5 el gesto estaba formado por un solo movimiento, en los ejemplos 19, 20 y 21 se muestra la sincronía entre los movimientos de un mismo hablante. El ejemplo 19 muestra dos gestos independientes complejos de distintos hablantes, A y D; ambos asienten mientras elevan las cejas solapándose entre ellos y con la intervención C. El espaciado deja claro que forman un mismo gesto/intervención.

(19)

1C1: [eso se lo hacen a las] mujeres mayores de treinta y cinco años

2B1: es que ella tiene treinta y ocho

A1: [AsienteRep-----]
[CejEleva-----]

D1: [AsienteRep-----]
[CejEleva-----]

C: [[se lo hicieron a mi hermana] Paula o] Chabel o no sé a quién [se lo hicieron]
(Conv. 1, 0:13:13)

Ciertos gestos más largos, como AsienteRep, no se sincronizan con otros, como ocurre en 3B1 del ejemplo 20. El código del asentimiento se repite tres veces: antes, durante y después del gesto de ojos, pero no son tres gestos distintos. Esto es algo que ya sucedía en el Corpus Val.Es.Co. con la sincronización de elementos paraverbales, como (RISAS).

⁹⁷ ManoPlana PalmaAbajoDi HaciaDentro AFrente

(20)

1B1: qué guay bueno si te vas en septiembre podré ir a verte

A1: mm

B: seguro

2A2: ya habrá ahorroh

3B1: AsienteRep [AsienteRep]
[OjosOtro⁹⁸-]

A3: [Sonrisa-----]

B: AsienteRep

A: en fin a ver si [me voy a algún lao]

B: [((sí porque)) es]te año lo tengo jodío

(Conv 5. 0:00:35)

Mondada (2018) utiliza una línea para cada parte del cuerpo del hablante; nosotros solamente utilizamos una línea distinta si los movimientos son simultáneos. En A2 del ejemplo 21, el movimiento de la cabeza y de la cara se suceden y están en la misma línea para que la relación temporal esté clara.

(21)

1A1: [bueno ahora] vas a empezar en una espiral de que [cada mes vas a tener]

2B1: [ya no ya]

A: =[uno]

A2: [AsienteRep] [Sonrisa--]

B: [ya me he] visto otro

A3: [(RISAS)]

C1: [Sonrisa--] [LabRetrae]
⇒B [CejEleva--]

A: ahora que has visto que no duele ven[ga pa'lante]

(Conv. 5, 0:26:21)

7) *Orientación corporal*. El símbolo ⇒, visible en C1 del ejemplo 8, lo tomamos de Gallardo y Sanmartín (2005: 49). Ellas lo utilizan para la mirada en el Corpus PerLa en casos en los que es relevante; por consiguiente, lo hemos adoptado para la orientación del gesto o de la intervención verbal cuando sea necesario para indicar hacia quién se

⁹⁸ Ojos hacia arriba

dirige el hablante porque el contexto es ambiguo y no se deduce de la contigüidad de las intervenciones. $\Rightarrow A$, $A \Rightarrow \emptyset$, $\Rightarrow \emptyset$ sustituyen en la transcripción en papel lo que en ELAN es *aA*, *RetiradaDesdeA* y *ANadie* y se colocan al inicio de la intervención debajo de la transcripción. Las flechas no deben confundirse con las utilizadas en Val.Es.Co. para la entonación (\uparrow , \downarrow , \rightarrow).

5.5.4. Variable función

En §3.4.3. se han descrito una serie de funciones de los gestos que combinan varias tipologías, principalmente, la tipología de Kendon (2004) y la de Müller (1998, 2013, en prep.): primero, la distinción entre gesto referencial y pragmático y, dentro del pragmático, discursivo, modal, performativo e interactivo (Tabla 29).

Variable función	Referencial
	Pragmática performativa
	Pragmática interactiva
	Pragmática modal
	Pragmática discursiva

Tabla 29. Variable función de los gestos.

En resumen, los gestos referenciales tienen un significado referencial y contribuyen al contenido proposicional del acto verbal, repitiendo el significado de las palabras, restringiéndolo, añadiendo otros significados o representando un objeto (Kendon 2004: 176-185). Los gestos pragmáticos no pertenecen al contenido proposicional del discurso del hablante. Entre ellos, los gestos performativos pueden acompañar a un acto verbal para contribuir en la interpretación de su fuerza ilocutiva (Kendon 2004: 159) o ser el propio acto si no acompaña a un acto verbal (Kendon 2004: 225). Los gestos interactivos están relacionados con la gestión del turno (autoselección, mantenimiento del turno, cesión del turno, interrupción, elicitación del hablante al oyente, gestos de retroalimentación del oyente), no se incluyen en las clasificaciones de Kendon y Müller, pero son descritos por otros autores (Ekman y Friesen 1969, Bavelas et al. 1992, Streeck 2005, Wehling 2017). Los gestos modales contribuyen a la interpretación de lo dicho como hipótesis, aserción, cita (Kendon 2004: 159) o a cualificarlo como obvio (Müller 2004), irrelevante, preciso (Müller en prep.), humorístico, equivocado (Bavelas y Chovil 2018), incierto (McClave 2000). Los gestos discursivos muestran cómo el hablante

organiza el contenido de su discurso y acompañan al ritmo prosódico⁹⁹ (Müller 1998, Kendon 2004).

Esta tipología resulta adecuada para el objetivo de nuestro estudio porque encontramos un paralelismo con unas unidades conversacionales de Val.Es.Co. (Tabla 30).

Funciones pragmáticas de los gestos	Unidades conversacionales de Val.Es.Co. (Briz y Grupo Val.Es.Co. 2003; Pons Bordería 2022)
Gestos discursivos (Kendon 2004; Müller 1998, 2013, en prep.)	Subacto adyacente textual (SAT)
Gestos modales (Kendon 2004)	Subacto adyacente modal (SAM)
Gestos interactivos (Ekman y Friesen 1969, Bavelas et al. 1992, Streeck 2005, Bohle 2013, Wehling 2017)	Subacto adyacente interpersonal (SAI)
Gestos performativos (Kendon 2004)	Acto

Tabla 30. Paralelismos entre las funciones de los gestos y de los subactos adyacentes.

Los gestos discursivos, al igual que los SAT de Val.Es.Co., funcionan “organizando y distribuyendo el flujo del habla, sin representar por sí mismos aporte semántico” (Briz y Grupo Val.Es.Co. 2003: 45). Los SAM vinculan al hablante con el texto que emite, matizan el subacto sustantivo al que se adhieren aportando atenuación, una actitud del hablante ante lo dicho (Briz y Grupo Val.Es.Co. 2003: 45), intensificación (Briz y Pons 2010), adhesión o distanciamiento, un estado psicológico (Pons Bordería 2022), aproximación, autorreafirmación, epistemicidad, evidencialidad (Pons Bordería et al. en prensa). En definitiva, los elementos modales generan efectos pragmáticos vinculados a los fenómenos sociales de cortesía y descortesía: estrategias de cortesía positiva y negativa (Brown y Levinson 1987[1978]), cortesía mitigadora o valorizante (Kerbrat- Orecchioni 1992, 1996) de autonomía y de afiliación (Bravo 1999).

Los SAI son similares a los gestos interactivos, vinculan a hablante y oyente (Pons Bordería 2022), incluyen interrogaciones fáticas apelativas del hablante hacia el oyente, muestras de atención del oyente hacia su interlocutor (Briz y Grupo Val.Es.Co. 2003: 45), muestras de acuerdo y desacuerdo y elementos que marcan la gestión del turno (cesión, mantenimiento, etc.) (Pons Bordería 2022, Pons et al. en prensa). Por último, no hay un

⁹⁹ Para simplificar, incluimos los gestos rítmicos en los discursivos, pues no esperamos que ninguno sea frecuente en este estudio.

tipo de subacto similar a la función performativa, pues es el acto el que tiene la fuerza ilocutiva y los subactos aportan significado de otro tipo, pero no sobre el tipo de acto que es. En este sentido, esta función gestual no es equiparable a un subacto verbal, pero es un elemento que contribuye junto a la información verbal a crear el acto comunicativo, de la misma manera que los gestos referenciales contribuyen al contenido semántico del acto. A pesar de ello, en posición independiente, un marcador discursivo (sobre todo SAM o SAI) puede formar un acto por sí mismo (Briz y Pons 2010) y ser, por tanto, performativo.

También hay que tener en cuenta la multifuncionalidad de los gestos o la ambigüedad de la propia tipología funcional (Kendon 2004, Streek 2009, Bressem y Müller 2014a). Por ejemplo, hemos visto clasificaciones del encogimiento de hombros como gesto modal y como gesto performativo, bien porque modaliza el enunciado verbal con el sentido epistémico de la incertidumbre, bien porque realiza el acto de mostrar incertidumbre. Lo mismo con el levantamiento de cejas para preguntar, bien marca la modalidad oracional, bien es performativo porque realiza una pregunta. Quizás sean ambos o quizás el límite determinante está en si son coverbales o no.

Aunque partimos de estas funciones generales, creemos en la utilidad descriptiva de identificar las funciones concretas de los gestos en cada contexto específico. Una vez clasificados los gestos funcionalmente, los agrupamos, buscamos patrones en los entornos secuenciales particulares (para esta tarea, ha sido útil la extracción del clip de vídeo fuera de ELAN para ver varios gestos rápidamente) y homogeneizamos una terminología para ellos basada en la bibliografía. En consecuencia, estas funciones específicas no se preestablecen ahora, sino que se descubren a través del análisis. A semejanza del AC, se han observado los casos sin unas categorías previas para no constreñir la interpretación a unos moldes preestablecidos.

Este nivel de detalle puede ser beneficioso para justificar los valores pragmáticos o semánticos de los gestos y observar las posiciones estructurales que ocupan las distintas funciones y cómo afectan a la dimensión social.

5.5.5. Anotación de gestos: la función

Tras la anotación formal, se anota la función de cada gesto en su contexto. Primero se clasifica según su macrofunción en la línea A_Func mediante un vocabulario controlado desplegable: pragmático performativo, pragmático interactivo, pragmático modal, pragmático discursivo, pragmático deíctico, referencial deíctico, referencial

representativo. Esta línea sirve para todos los gestos de un mismo hablante, pues entendemos que dos gestos no se van a solapar, sino que formarán parte de una misma unidad o uno se realizará detrás de otro.

Respecto a las funciones específicas que los gestos extraídos realizan en la interacción, no se ha partido de un listado específico de funciones, así que la anotación del gesto en la línea A_SubFun es de escritura libre.

En la Tabla 31, mostramos ya todas las líneas utilizadas en la anotación del corpus en ELAN.

DOMINIO	ATRIBUTO	VALOR
Transcripción oral	A_phon	(transcripción oral)
	A_palabras	(transcripción <i>tokenizada</i>)
	Observaciones	(escritura libre)
Segmentación oral	A_int-turno	Intervención iniciativa, Intervención reactiva (turno), Intervención reactiva (no-turno), Intervención reactivo-iniciativa, Intervención independiente, Intervención ininteligible.
Transcripción visual	A_phon_vid	(transcripción oral)
	A_palabras_vid	(transcripción <i>tokenizada</i>)
	Obs_vídeo	(escritura libre)
Segmentación visual	A_int-turno_vid	Intervención iniciativa, Intervención reactiva (turno), Intervención reactiva (no-turno), Intervención reactivo-iniciativa, Intervención independiente.
	Obs_segmentación	(escritura libre)
Análisis de cambios en seg. visual	A_cambios	
	A_Gesto_Tipo	Independiente, Coverbal
Análisis de gestos: anotación formal	A_Cabeza	Asiente, AsienteArriba, AsienteRep CabArriba, CabInclina, CabSacude, CabOscila, CabOtro
	A_Hombros	Encoge, HombrosOtro
	A_Cara	Cejas, Ojos, Boca, Cara, Varios
	A_Cara_movs	OjosAbre, OjoCierra, Parpadeo, OjosOtro, CejFruce, CejEleva, CejOtro, BocaAbre, BocaCierra, ComisurAbajo, LabiosFruce, LabiosRetrae, LabiosOtro, Sonrisa, Risa, CaraOtro
	A_Manos_Conf	Puño, ManoPlana, Indice, Corazón, Anular, Meñique, Pulgar, Estirado, Laxo, Curvado, Torcido, Perpendicular, Conectado, Tocando
	A_Manos	PalmaArriba, PalmaAbajo, PalmaLateral, PalmaVertical, HaciaCentro, HaciaFuera, HaciaCuerpo, HaciaFrente, Diagonal, Centro-centro, Centro, Encima,

		EncimaDerecha, Encimalzquierda, Debajo, DebajoDerecha, Debajolzquierda, Recto, Arqueado, Círculo, Zigzag, Eses, Espiral, ADerecha, Alzquierda, Arriba, Abajo, ACuerpo, Afrente, Reducido, Alargado, Decelerado, Acelerado, Acentuado
	A_Gesto_Orient	aA, aB, aC, aD, RetiradaDesdeA, RetiradaDesdeB, RetiradaDesdeC, RetiradaDesdeD, aNadie
Análisis de gestos: anotación funcional	A_Func	Prag Performativo, Prag Interactivo, Prag Modal, Prag Discursivo, Prag Deíctico, Ref Deíctico, Ref Representativo
	A_SubFun	(escritura libre)

Tabla 31. Esquema de anotación del corpus para un hablante. En naranja, la transcripción oral; en azul, la transcripción de gestos; en amarillo, las funciones de los gestos; en verde, la segmentación en unidades conversacionales; en gris, las observaciones.

CAPÍTULO 6. RESULTADOS

6.1. CAMBIOS EN LA DIMENSIÓN ESTRUCTURAL Y EN LA DIMENSIÓN SOCIAL.....	279
6.1.1. LA INFORMACIÓN VISUAL ES NECESARIA EN LA SEGMENTACIÓN DIALÓGICA DE LA CONVERSACIÓN	279
6.1.2. LA INFORMACIÓN VISUAL DA GANANCIA EN LA DIMENSIÓN SOCIAL DE LAS INTERVENCIONES	281
6.1.3. CAUSAS VISUALES DE CAMBIOS EN LA SEGMENTACIÓN ESTRUCTURAL.....	287
6.1.4. CAUSAS VISUALES DE CAMBIOS EN LA SEGMENTACIÓN SOCIAL	291
6.1.5. RESUMEN.....	295
6.2. ANÁLISIS DE GESTOS INDEPENDIENTES Y COVERBALES.....	297
6.2.1. FUNCIÓN	298
6.2.2. FUNCIÓN Y POSICIÓN	323
6.2.3. FORMA Y FUNCIÓN.....	330
6.2.4. POSICIÓN Y FORMA	334
6.2.5. RESUMEN.....	334

6.1. Cambios en la dimensión estructural y en la dimensión social

Los primeros resultados que vamos a mostrar son los del análisis entre la segmentación de las intervenciones con y sin elementos visuales, teniendo en cuenta los cambios en la dimensión estructural y en la social y la causa de estos.

No se incluye en ninguna parte del análisis los gestos independientes del gancho ni los cambios en las intervenciones verbales que se hayan producido por un gesto independiente o coverbal del gancho. El número de gestos independientes del gancho es muy superior al resto: 375 frente a 360 de todos los demás. Es una gran proporción, puesto que el gancho aparece en 5 conversaciones en las que participan 13 interlocutores más. Esto puede deberse a que el gancho intentaba participar menos en la conversación, pero esto le hacía producir más gestos reactivos, o a que es una persona que da mucha retroalimentación. Téngase en cuenta que el uso de gestos independientes no es homogéneo entre los individuos. Por ejemplo, en la conversación 2 (Conv2), B hizo 29, frente a C (66) y D (70); y no porque hablara poco —sus intervenciones verbales fueron: B 252, C 248, D 338— o tuviera un rol de oyente —179 intervenciones validadas y 76 no validadas—. Aun así, excluimos al gancho del análisis para solamente tener en cuenta los datos de los informantes que no eran conscientes de la grabación.

6.1.1. La información visual es necesaria en la segmentación dialógica de la conversación.

En la nueva segmentación que tiene en cuenta la información visual (en adelante, segmentación visual), podemos observar (Gráfico 1) que 2773 intervenciones se han mantenido como en el análisis oral, pero 803 han sufrido un cambio, bien porque el tipo de intervención o su validación ha cambiado, bien porque son intervenciones nuevas que surgen gracias a la información visual.

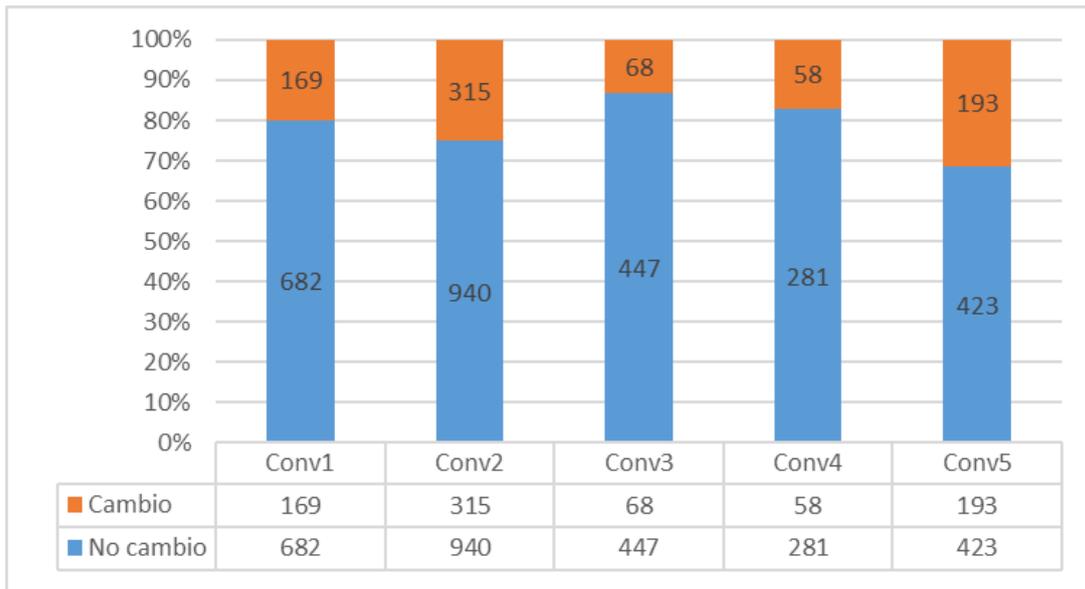


Gráfico 1. Número de cambios producidos en la segmentación visual.

El valor-p de la X^2 es bajo, así que estos resultados son representativos de la población.

X^2 (Valor observado)	67,082
X^2 (Valor crítico)	9,547
valor-p	<0,0001

Como las conversaciones tienen distintas duraciones y número de participantes, el número total de intervenciones es muy variado entre cada conversación, pero en el Gráfico 1 los cambios suponen un porcentaje entre 13 y 31% dentro de su conversación.

Una posible razón para esta variación entre el 13 y el 31% puede ser el número de participantes. En Conv1, Conv2 y Conv5 había 4 hablantes, mientras que en Conv3 y Conv4 había 3. A través de los residuos de la X^2 , observamos que especialmente en la Conv2 y la Conv5 hay más cambios y en la Conv3 y Conv4, menos. No forma parte de este análisis la variable del número de hablantes o del comportamiento particular de cada hablante, pero estos datos nos sirven para plantear una hipótesis para futuros trabajos: cuantos más hablantes en una conversación, mayor ganancia en la segmentación visual.

	No cambio	Cambio
Conv1	0.8600785	-1.598289
Conv2	-1.0638091	1.976882
Conv3	2.3841505	-4.430480
Conv4	1.1177940	-2.077203
Conv5	-2.5016534	4.648837

Tabla 32. Residuos de la X^2

Ya sin dividir los datos por conversaciones, los 803 cambios que se producen en todo el corpus corresponden a nuevas intervenciones que en la segmentación oral no existían o se deben al cambio en la segmentación del tipo de intervención con el que se clasifica las intervenciones (Ii, Ir-i, Ir, Iind, Iint). También ha habido 6 casos en los que la intervención se ha eliminado porque la lectura de labios ha mostrado que ese ruido no era parte de la intervención de ningún hablante (4), porque formaba parte de la intervención de otro hablante (1), o porque la orientación corporal ha mostrado que formaba parte de otra intervención (1). Como son casos marginales, los hemos contabilizado dentro de los cambios por nueva intervención, de manera que debe interpretarse esta variable como que afecta a la aparición/desaparición de intervenciones completas.

Nuestra hipótesis de más cambios por nuevas intervenciones que por modificación de las originales se cumple: hay 539 nuevas intervenciones (NI) y se han modificado 264 (MI).

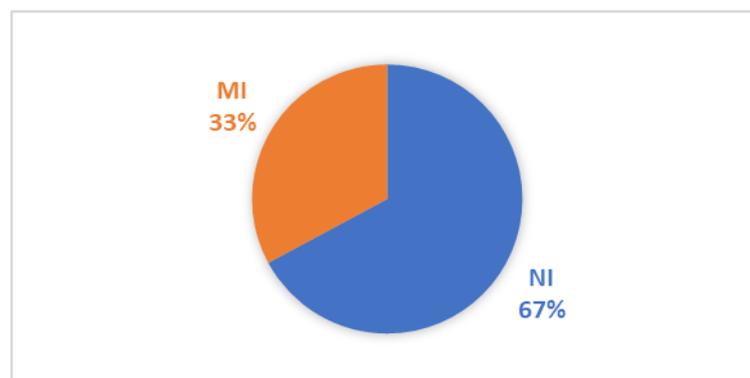


Gráfico 2. Cambios estructurales en la segmentación en porcentajes: nuevas intervenciones y modificación de intervenciones

6.1.2. La información visual enriquece la dimensión social de las intervenciones.

A través de las pruebas vistas en §6.1.1., confirmamos que la información visual genera cambios significativos en la segmentación dialógica. Veamos ahora cuántos cambios afectan a la dimensión social de las intervenciones.

En la dimensión social de las intervenciones, se distingue entre las intervenciones validadas (turno) y no validadas (intervenciones). En los Gráficos 3 y 4, observamos la ganancia en el número de intervenciones y turnos que aparecen en la segmentación oral y visual. Utilizando solamente la información auditiva, encontramos 3096 intervenciones,

de las cuales 1783 eran turno y 1313 no estaban validadas. En cambio, en la segmentación a partir de la información visual, hay 1971 turnos y 1590 intervenciones. Esto es un incremento del 21% en las intervenciones y del 10,5% en los turnos.

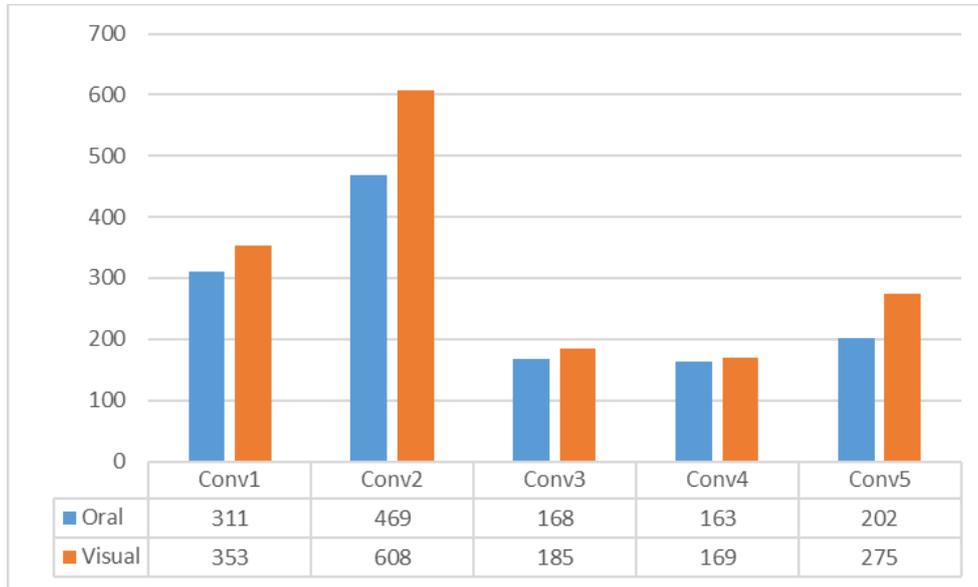


Gráfico 3. Número de intervenciones no validas en la segmentación oral y en la visual.

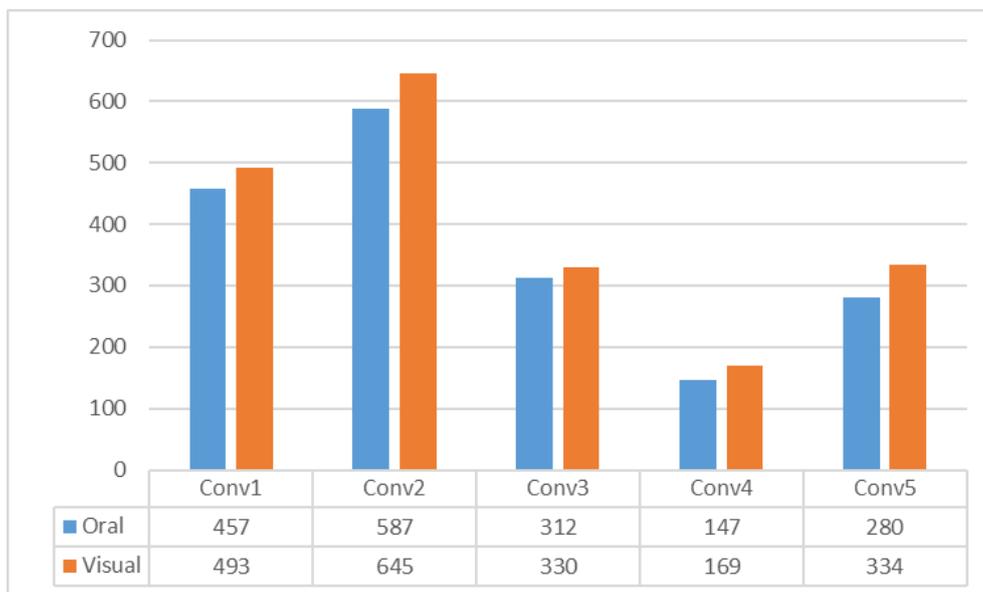


Gráfico 4. Número de turnos validados en la segmentación oral y en la visual.

Como la duración entre las cinco conversaciones varía de 16 a 36 minutos (16, 20, 29, 30, 36) hemos normalizado los datos a 20 minutos para poder contrastar las medias de intervenciones y turno de las conversaciones. El incremento en las medias es suficientemente elevado para justificar una ganancia en la segmentación (Tabla 33).

Intervenciones		Turnos	
Diferencia	195,790	Diferencia	263,926
valor-p (bilateral)	0,001	valor-p (bilateral)	<0,0001

Tabla 33. Diferencia de medias normalizadas de intervenciones y turnos entre segmentación oral y visual.

Como ya hemos visto en el Gráfico 2, la nueva segmentación incluye modificaciones de las intervenciones de la segmentación oral y nuevas intervenciones. Socialmente (Gráfico 5), el incremento de los turnos en la segmentación visual puede venir de que las nuevas intervenciones (NI) estén validadas (T) o de que la modificación de las intervenciones suponga una validación (MV) de las intervenciones ya conocidas (I>T).

En el ejemplo 22, la lectura de labios y la acción de llevar una cerveza a C permite distinguir ruidos que se creían de ambiente como nuevas intervenciones que, además, son turno porque son un par adyacente de agradecimiento que forman un diálogo paralelo a la conversación principal:

(22)

1B1: tenemos contacto o sea que [todas-]
 (D LLEVA CERVEZA A C)
 A1: [ya]
 1C1: [gracias cariño]
 B: no todos [los dí]=
 2D1: [((nada))]
 B: =as pero todas las semanas hablamos en plan

(Conv5, 0:06:20-0:06:25)

En el ejemplo 23b, la intervención 2A1 resulta estar validada por el gesto coverbal de 3C2, así que es una Ir-i, no una Ir sin validar (A1), como se podía interpretar de la información auditiva (23a):

(23a)

1C1: hoy eso he entregado la primera parte del TFM tenía que entregar la introducción [el índi]=
 A1: [ah qué guay]
 C: =ce y todo eso (1,1) yy/ ya esperar

(23b)

1C1: hoy eso he entregado la primera parte del TFM tenía que entregar la introducción
[el índi]=
2A1: [ah qué guay]
3C2: =[ce] y todo eso (1,1) yy/ ya esperar
[Asiente]

(Conv5, 0:14:12-0:14:25)

En cambio, aumentan las intervenciones no válidas cuando las NI encontradas son I o la nueva segmentación social desvalida turnos (T>I). En el ejemplo 24, el chasquido que se le atribuía a C resultar ser una reacción a B acompañada de un gesto; C2 y C3 son nuevas Ir no validadas.

(24)

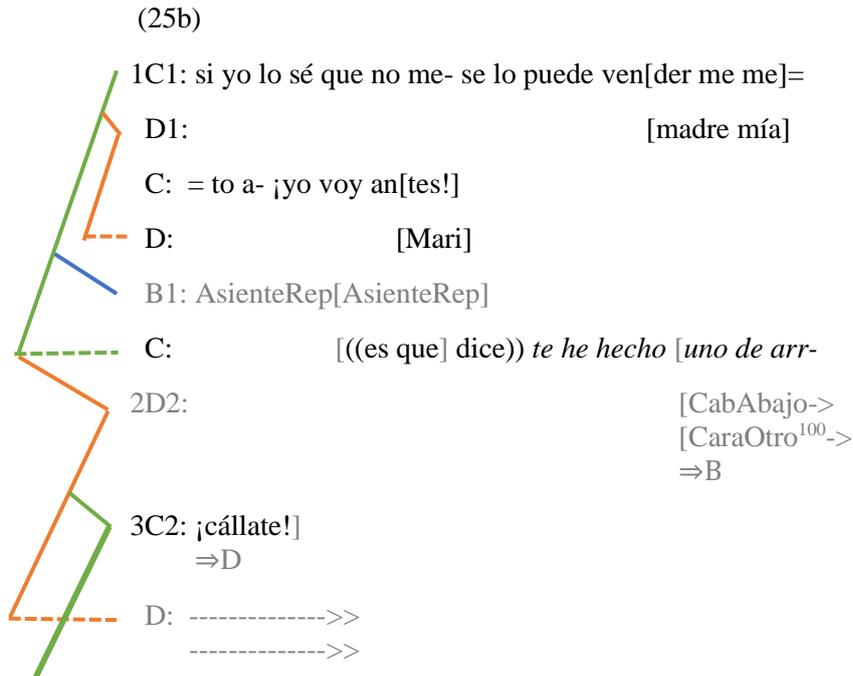
1C1: lo valientes que son ahora
2B1: no↑
C2: (CHASQUIDO) Asiente
B: túú que te hubieras enseñao [lo (BALBUCEO)- lo hubieras llevado]
C3: [AsienteRep-----]AsienteRep

(Conv4, 0:02:00-0:02:05)

En 25a, 1C1 está narrando una historia, pero 2D1 está cansado de escucharla y 3C2 no le contesta inmediatamente, continúa parte del relato, pero se interrumpe para contestarle. En cambio, en 25b vemos que D1, al no recibir reacción de C, le transmite su hastío por la situación a B con el gesto 2D2. Esa segunda queja sí recibe la reacción de 3C2, por tanto, la primera no es Ir-i turno (2D1), sino Ir no validada (D1). La intervención modifica su segmentación de turno a intervención (T>I).

(25a)

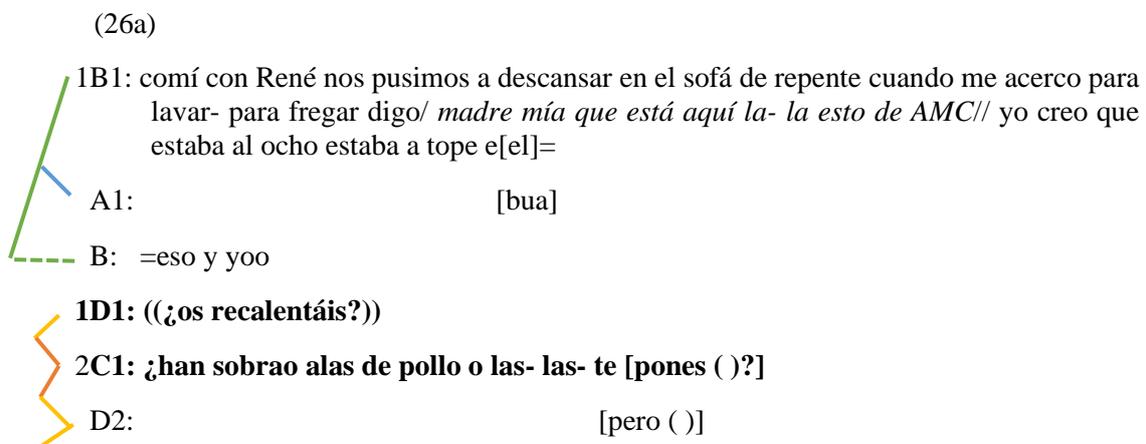
1C1: si yo lo sé que no me- se lo puede ven[der me me]=
2D1: [madre mía]
C: to a- ¡yo voy an[tes!]
D: [Mari]
C: ((es que dice)) *te he hecho uno de arr-*
3C2: ¡cállate!



(Conv 1, 0:07:50-0:07:56)

No se produce cambio social en las intervenciones que cambian de tipo de intervención, pero no modifican la validación (NMV); por ejemplo, si los elementos visuales cambian la segmentación de una Ir-i a Ii, sigue siendo un turno (T>T), o de una Ir a Ind, sigue sin estar validada (I>I), aunque estructuralmente la posición sea distinta.

En el ejemplo 26, el vídeo ayuda a segmentar el límite entre dos diálogos. En 26a, por contexto, se puede entender que 1D1 no reacciona a 1B1, sino que inicia un nuevo diálogo con una Ii y 2C1 le contesta. En cambio, en 26b, por la orientación corporal de los interlocutores, se ve que 2D1 es una Ir-i que reacciona a 1B1 y quien inicia el diálogo (Ii) es 1C1. No hay cambio social (T>T), pero sí estructural, de Ii a Ir-i y de Ir-i a Ii.



¹⁰⁰ Resopla

- (26b)
- 1B1: comí con René nos pusimos a descansar en el sofá de repente cuando me acerco para lavar- para fregar digo/ *madre mía que está aquí la- la esto de AMC// yo creo que estaba al ocho estaba a tope e[el]=*
 - A1: [bua]
 - B: =eso y yoo
 - 2D1: ((¿os recalentáis?))
 - 3B2: AsienteRep
 - 1C1: ¿han sobrao alas de pollo o las- las- te [pones ()?]
 - 2D2: [pero com-]

(Conv 1, 04:03-04:21)

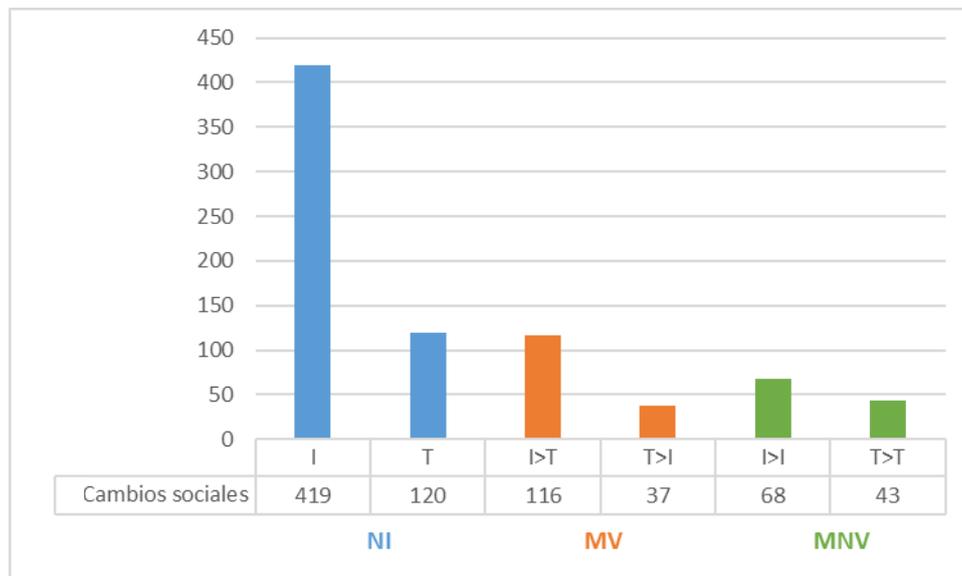


Gráfico 5. Cambios sociales por nueva intervención (intervención o turno), por modificaciones en la validación (de intervención a turno o viceversa) o sin cambios.

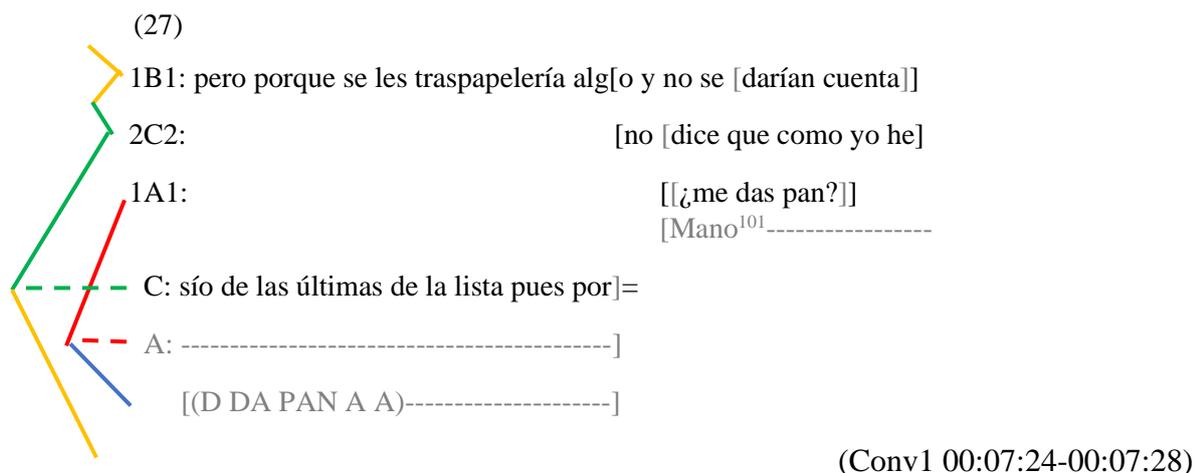
Se cumple la hipótesis de que mayoritariamente las nuevas intervenciones no están validadas (Gráfico 5). Respecto a las intervenciones modificadas, era esperable que los elementos visuales validaran intervenciones en vez de “desvalidarlas” porque es más sencillo que un nuevo elemento demuestre validación sobre una intervención que se creía ignorada; en cambio, para desvalidar, los cambios visuales deben haber modificado de alguna manera la interpretación de la secuencia en la que se creía que una intervención estaba validada. En el análisis de las causas (§6.1.4.) vemos qué elementos han causado estos 37 cambios de T a I.

Comprobamos en este análisis de las variables sociales que las intervenciones que modifican su tipo de intervención generalmente también cambian socialmente su estado a turno o no turno (un 58% de las MI son MV), mientras que menos cambian solamente el tipo de intervención estructural (un 42% de las MI son NMV), pero sin una diferencia muy amplia.

6.1.3. Causas visuales de cambios en la segmentación estructural

Las causas de cambio contempladas en el análisis son: los gestos independientes (Gind en los gráficos), los gestos coverbales (Gcov), la orientación corporal durante la intervención verbal (Orient), la lectura de labios (Lab) y la acción extralingüística (Acc). Para explicar el origen de los cambios, vamos a centrarnos, primero, en los cambios que aparecen en la estructura de la conversación, es decir, en el tipo de intervención según su lugar en la interacción. Distinguiremos, además, si el cambio resultante ha sido un cambio en una intervención ya conocida en la segmentación auditiva o la creación de una nueva intervención; así como si el cambio afecta al propio hablante que realiza el gesto/orientación/movimiento de labios/acción o a otro interlocutor que se ve afectado.

Hay 64 cambios multicausales, es decir, producidos por dos causas o, incluso, tres a la vez. En el ejemplo 27 la intervención 1A1 es una Ii que por su ininteligibilidad auditiva se había segmentado como Iint, pero la lectura de labios, el gesto coverbal de A y la acción resultante de D nos ayuda a comprender que A decía “¿me das pan?”. Es un diálogo paralelo al de B y C.



¹⁰¹ Mano: ÍndiceEstirado PalmaLateral Diagonal HaciaCuerpo

Sin embargo, para no complicar el análisis de las causas visuales para los cambios, cada causa se contabiliza como una unidad de cambio; en consecuencia, el total de causas de los próximos gráficos no coinciden con el total de cambios mostrado en el Gráfico 2.

Respecto a la dimensión estructural, en este estudio nos interesa saber qué elementos visuales crean nuevas intervenciones y cuáles modifican las ya existentes en la segmentación oral.

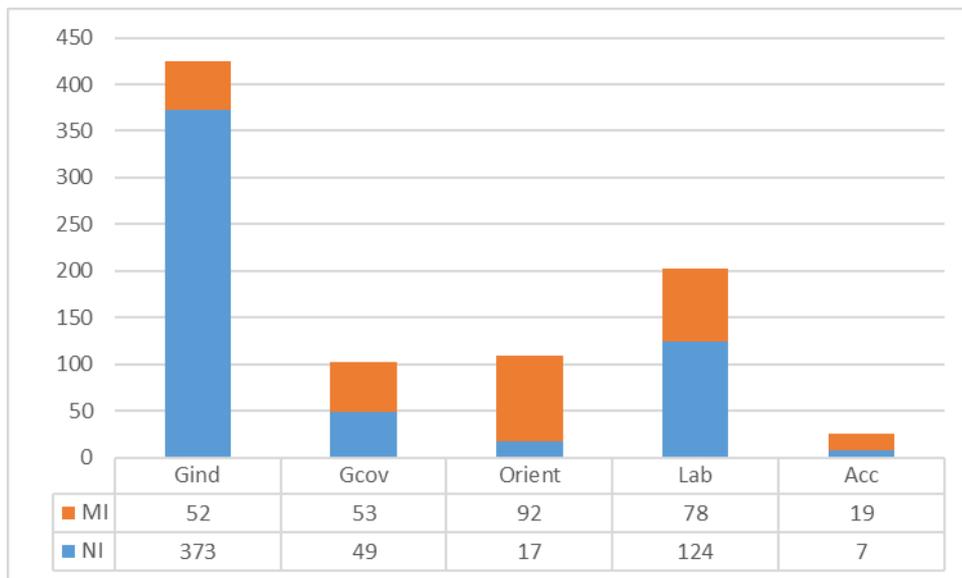


Gráfico 6. Causas visuales de los cambios estructurales.

X ² (Valor observado)	247,294
X ² (Valor crítico)	9,538
valor-p	<0,000
	1

Tabla 34. X² de las causas visuales de los cambios estructurales.

	NI	MI
Gind	5.5312162	-7.701661
Gcov	-2.2298345	3.104820
Orient	-6.4752336	9.016110
Lab	-0.8024853	1.117380
Acc	-2.4514196	3.413355

Tabla 35. Residuos de la X² de las causas visuales de los cambios estructurales.

En el análisis preliminar había más cambios provocados por gestos independientes (57%) que por el resto de causas (43%); en este corpus más grande los puestos se han invertido, pero también acercado (GInd 49%, resto 51%). Por tanto, se cumple la hipótesis de que es la causa que más cambios provoca, pero por sí misma, en conjunto el resto de las causas suman algunos cambios más. También se mantiene el hecho de que las nuevas

intervenciones normalmente las generan los gestos independientes (373), frente a las 197 debidas al resto de causas. Este contraste también se aprecia en el residuo de la chi-cuadrado. Especialmente, la lectura de labios es la causa que se posiciona en segundo lugar (124).

También se cumple que los gestos independientes modifican en menor cantidad las intervenciones verbales ya existentes, únicamente un 17% de las MI (-7.701661 en el residuo de la X^2). Esta proporción es menor que la encontrada en el estudio preliminar (37%), en la que solo se contabilizaban las modificaciones en la validación social. Del resto de causas, la orientación corporal¹⁰², seguida de la lectura de labios¹⁰³, es la causa de mayor peso en las MI.

Destaca el bajo número de acciones extralingüísticas (21) que afecta a la segmentación durante las cinco conversaciones. Queda claro de esta forma que generalmente en una conversación las acciones que realizan los hablantes no son determinantes en la segmentación de la conversación; aunque todo dependerá de la situación concreta de cada conversación, por ejemplo, si se conversa mientras se realiza una actividad, como cocinar. Pero en este corpus los interlocutores estaban enfocados en la actividad social de conversar.

Estos cambios estructurales pueden ser causados porque en la propia intervención nueva o modificada el emisor es el que realiza el gesto o la acción o es el sujeto en el que observamos un movimiento de labios o una orientación corporal (NI-P y MI-P en el Gráfico 7) o porque estos cambios en una intervención afectan estructuralmente a la intervención de otro emisor (NI-A y MI-A).

¹⁰² 17 de estas orientaciones se combinan con otras causas para las MI.

¹⁰³ 23 de estas lecturas de labios se combinan con otras causas para las MI.

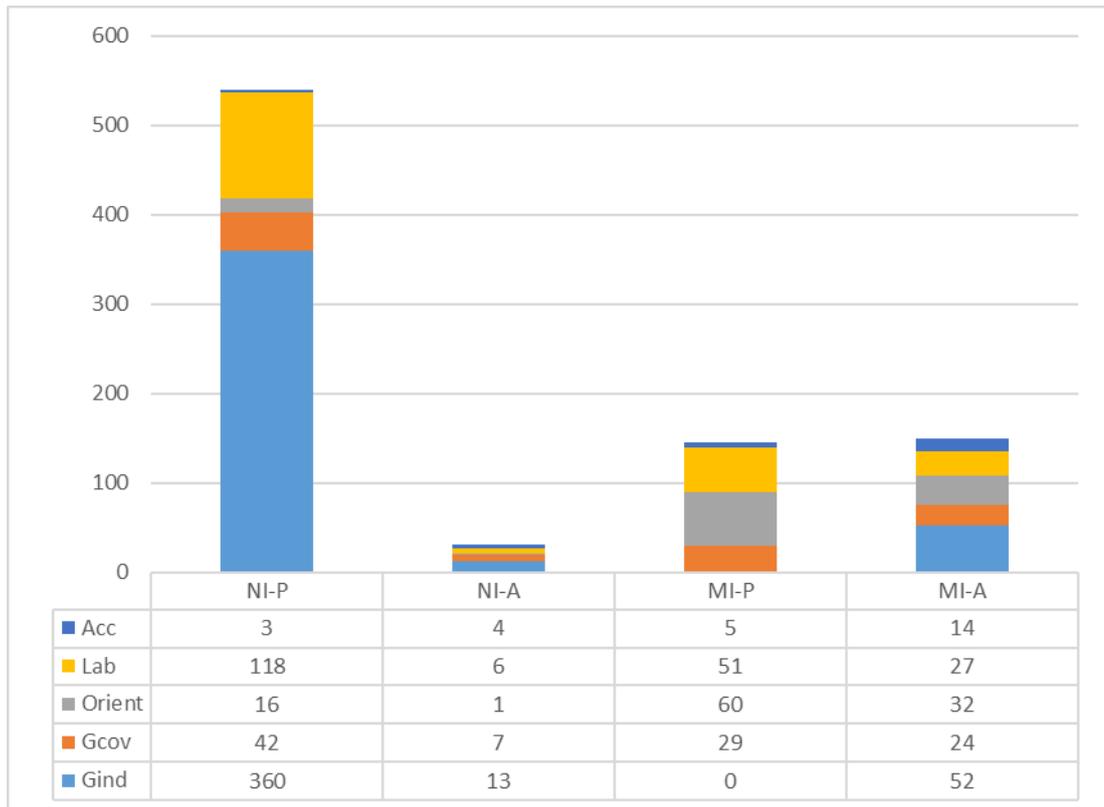


Gráfico 7. Causas visuales de los cambios estructurales, dependiendo de los efectos en la intervención propia o la ajena.

Los cambios que más se producen son NI-P, especialmente gracias a los gestos independientes y a la lectura de labios en segundo lugar. Los gestos independientes no solo son nuevas intervenciones por sí mismas, sino que 13 intervenciones más (NI-A) se descubren gracias al efecto de los gestos independientes en otro interlocutor. Por definición, los gestos independientes no pueden generar ninguna MI-P porque no existe una intervención verbal propia a la que esté vinculada. Sin embargo, los gestos independientes son causa de más MI-A que el resto de elementos visuales separadamente.

Mientras que la distancia entre NI-P y NI-A es mayor en los gestos coverbales, las MI-P y las MI-A están más igualadas. Tanto la orientación corporal como la lectura de labios suelen afectar más a las intervenciones propias. Aun así, en el caso de la orientación, el cambio más común es MI-P, y en el de los labios, NI-P.

El total de MI-P y MI-A es casi igual, pero gracias a la cantidad de MI-A por gestos independientes. El resto de las causas (excepto las acciones), afectan más a las MI-P que a las MI-A.

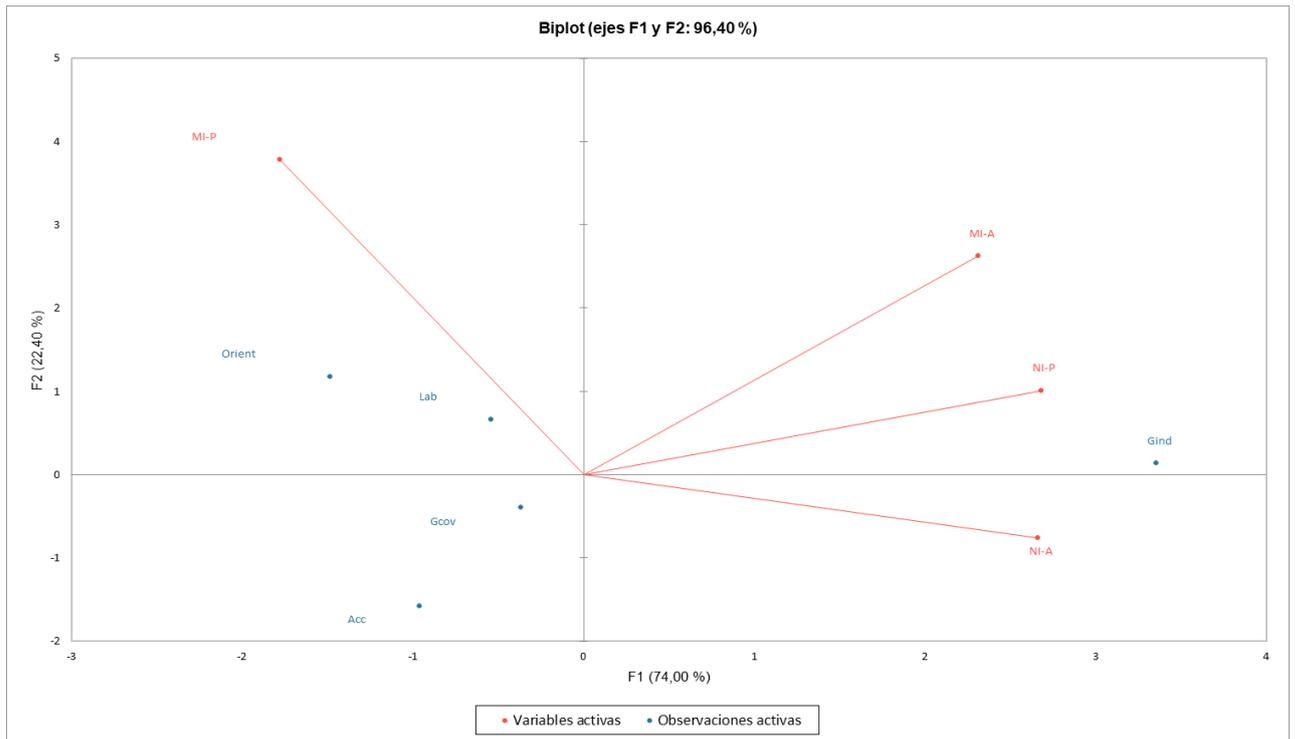


Gráfico 8. Análisis de componentes principales de causas visuales de los cambios estructurales, dependiendo de los efectos en la intervención propia o la ajena.

El análisis de componentes principales (Gráfico 8) muestra que los gestos independientes generan más cambios de todos los tipos (excepto MI-P) que el resto de los elementos visuales, especialmente NI-P. Como también se veía en el Gráfico 7, en general, la creación de una nueva intervención ajena (NI-A) es minoritaria respecto al resto de cambios (NI-P, MI-P, MI-A); pero los gestos independientes son los elementos que más causan NI-A. La distribución más estable de los elementos que la causan es la de MI-A.

Aunque la orientación afecta más a las MI-P, estas no se ven muy afectadas en general. Los gestos independientes y la orientación son las causas principales, mientras la lectura de labios, los gestos coverbales y las acciones sirven de apoyo, sobre todo a MI-P.

6.1.4. Causas visuales de cambios en la segmentación social

Respecto a los cambios producidos en el nivel social, ya hemos visto los cambios en el Gráfico 5; a continuación, en el Gráfico 9 se observan las causas de estos cambios.

Se cumple la hipótesis de que la mayoría de las NI causadas por gestos independientes son I, no T. Esta tendencia también se cumple en las NI descubiertas por

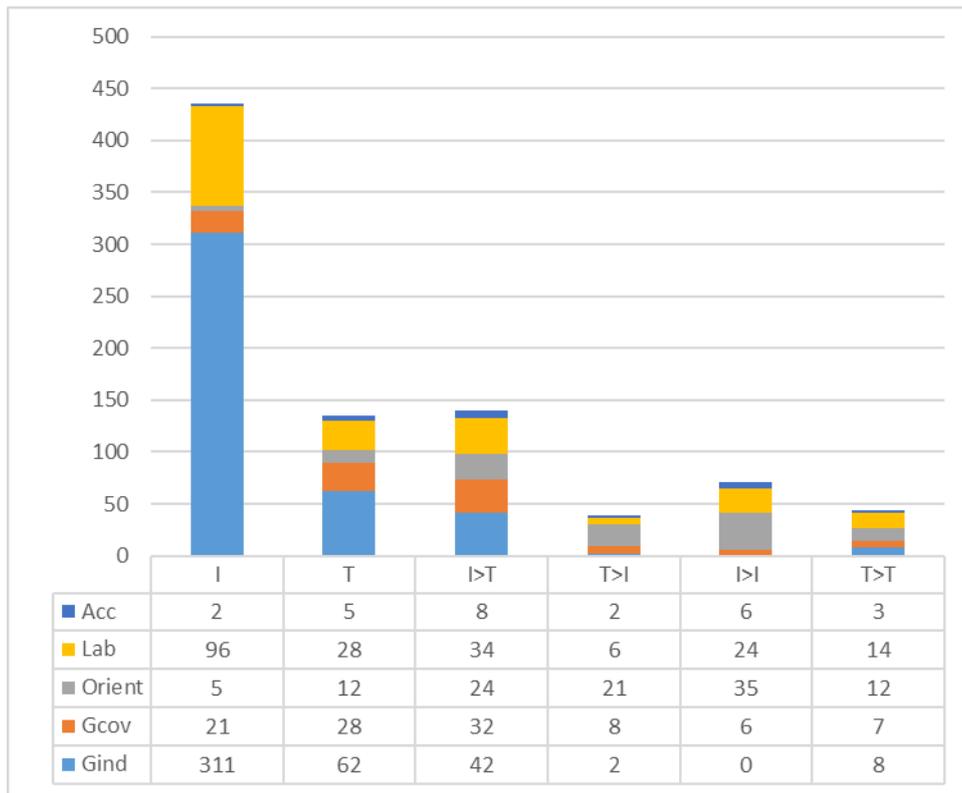


Gráfico 9. Causas visuales de los cambios sociales

la lectura de labios, pero sin una diferencia tan acusada. Quizás se deba a que las Ir tiendan a pronunciarse a menor volumen y no se captan en la grabación. El alto número de nuevas I por estas causas dispara la diferencia entre nuevas I y T. Sin embargo, en el resto de causas (gestos coverbales, orientación y acciones), aunque afecten en menor medida a las NI, se invierten los resultados: la cantidad de nuevos turnos es superior a la de intervenciones.

Los gestos independientes, aunque generan NI, no afectan mucho a la modificación social o estructural de las intervenciones. Se entiende de esta manera que los nuevos intercambios que generan estas NI con las intervenciones verbales no modifican la segmentación de estas últimas, pero la refuerzan. Es decir, que las intervenciones verbales no solo están validadas por la reacción verbal de un hablante, sino también por la reacción gestual de otro.

Los labios afectan algo más a las MI, pero su mayor efecto también está en las nuevas I; en cambio, el resto de las causas afectan más a las MI que a las NI. En conjunto, contribuyen en mayor medida a la las MV que a las NMV, pero la orientación, los labios y las acciones mantienen bastante simetría entre MV y NMV.

Generalmente, todas las causas provocan una proporción de cambios I>T mayor que T>I, aunque en la orientación la diferencia no es tan acusada. De los 39 cambios identificados de T>I, la mayoría los causan la orientación de los interlocutores, seguidos en menor medida de los gestos coverbales y la lectura de labios. De T>I solamente hay dos casos que combinan causas, Lab+Gcov y Lab+Orient. En los otros 35 cambios una sola causa es suficiente para reinterpretar un supuesto turno como intervención no validada en la secuencia interactiva.

Aunque no esté reflejado en las variables independientes, informamos de que algunas intervenciones que mantienen su estado social no han cambiado tampoco de tipo de intervención porque la lectura de labios ha generado solamente un cambio de hablante; por ejemplo, se observa que quien emite una Ir no es A, sino C, por lo que es un cambio estructural, pero sigue siendo una Ir.

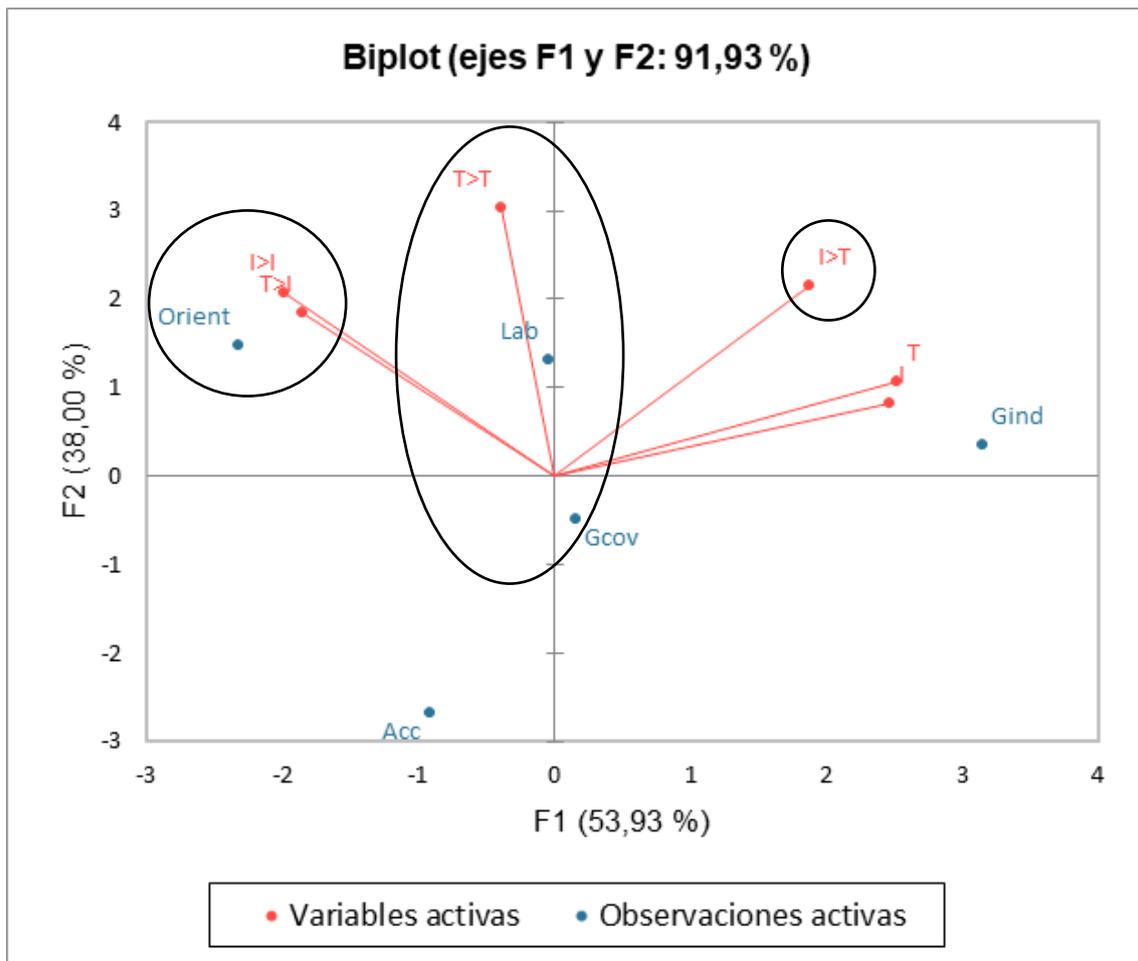


Gráfico 10. Análisis de componentes principales de causas de cambios sociales.

En el análisis de componentes principales, se corroboran los cambios sociales:

- Los gestos independientes crean sobre todo NI, pero también modifica intervenciones existentes de I>T.
- Siendo más numerosa la causa de la lectura de labios, a nivel estadístico, en el cómputo total existe una relación entre los labios y los gestos coverbales que modifican el tipo de intervención y mantienen su estado de turno.
- Contrastando con las otras causas, la orientación es la causa que más cambios provoca de tipo I>I y T>I; por tanto, generalmente confirma que la intervención no es turno.
- La acción sigue siendo una causa residual de cambios.

Por último, igual que hemos relacionado las variables Propio/Ajeno con las causas de los cambios estructurales (Gráfico 7), a continuación, las relacionamos con las causas de los cambios sociales (Gráfico 11). De esta manera observamos si los elementos visuales modifican o no modifican la validación de las intervenciones propias o ajenas. Las columnas MI-A y MI-P se desglosan en MV-P (modifica validación propia), MV-A (modifica validación ajena), NMV-P (no modifica validación propia) y NMV-A (no modifica validación ajena).

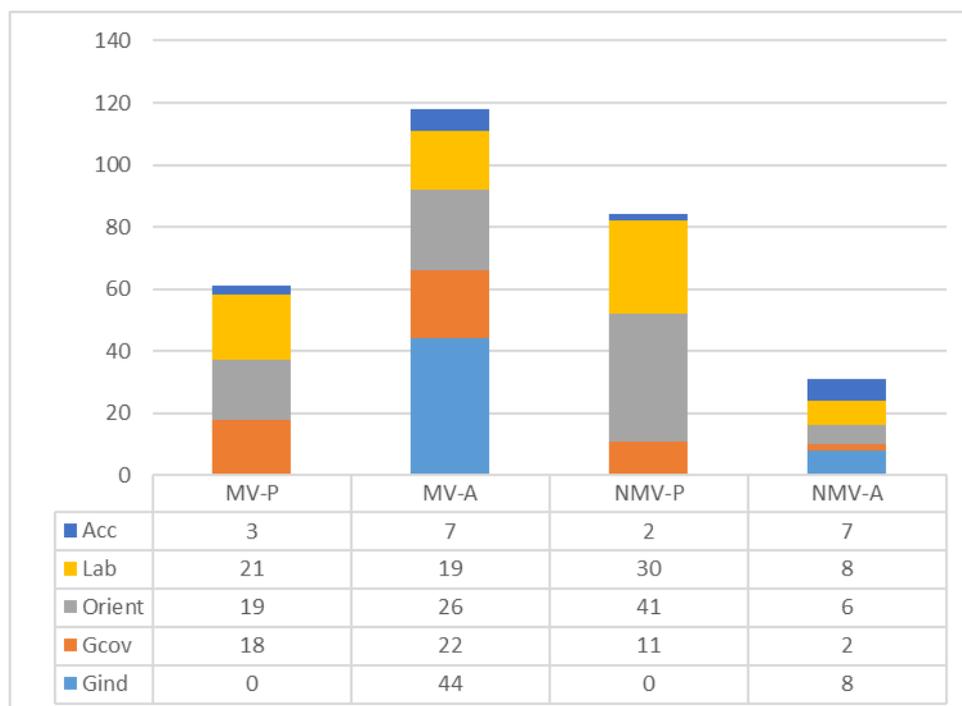


Gráfico 11. Causas visuales de los cambios sociales por sus efectos en la intervención propia o la ajena.

Como ya hemos comentado, es imposible que los gestos independientes afecten a intervenciones verbales propias, en consecuencia, solamente afectan a intervenciones verbales ajenas, ya sea estructuralmente (NMV-A) o estructural y socialmente (MV-A) (Gráfico 11). Contrastando estas dos columnas, todas las causas suelen afectar más a la MV-A que a la NMV-A, con una diferencia más acusada en los gestos independientes; excepto las acciones, que han afectado un mismo número de veces a unas y otras.

Comparando dentro de las MV, las MV-A son algo superiores a las MV-P a causa de los gestos coverbales, la orientación o, casi iguales, por los labios. Es una diferencia importante respecto a los gestos independientes, que, al afectar tanto a los MV-A, aumentan la distancia en contra de los MV-P totales.

En cambio, los gestos coverbales, la orientación y los labios muestran una diferencia más clara entre las NMV-P y las NMV-A, siendo las primeras superiores a las segundas.

6.1.5. Resumen

En este primer análisis, hemos observado que el número de cambios producidos entre la segmentación oral y visual es significativa. De estos cambios, un 67% está producido por intervenciones completamente nuevas que no existían en la segmentación oral. De estas intervenciones nuevas, un 65% son gestos en posición independiente de intervención, de las cuales un 83% son intervenciones no turno. Por tanto, 311 gestos independientes en intervenciones no turno son el 36% de todas las causas de cambio en la segmentación (311 de 864). Este efecto por esta causa agrupa el porcentaje más alto de cambios en la segmentación, pero el resto de las combinaciones y causas suman hasta llegar al 100%. La suma de otras causas supera por poco margen a todos los cambios que generan los gestos independientes (51%-49%). La causa con menos repercusión es la acción extralingüística (5% de los cambios), en consecuencia, podemos decir que no son determinantes en la segmentación de la conversación.

La lectura de labios es la segunda causa de aparición de nuevas intervenciones en la segmentación visual. Este es un factor que podría cambiar dependiendo de la calidad de los micrófonos, pero es bueno saber que en última instancia la lectura de labios podrá identificar cualquier intervención que escape a la grabación de audio. Del conjunto de nuevas intervenciones, la mayoría no está validada.

El resto de las causas (especialmente, la orientación corporal) afecta más a la modificación de intervenciones existentes en la segmentación oral que a la creación de nuevas intervenciones. Además, hay más modificaciones de intervenciones que suponen también un cambio social y no solo del tipo de intervención estructural (pero la orientación, los labios y las acciones mantienen bastante simetría). De los cambios producidos en la validación de intervenciones existentes, es mayor el cambio de intervención a turno que viceversa.

La distinción entre los efectos en la intervención propia o en la ajena nos da datos interesantes como que los gestos independientes, aunque generen muchas intervenciones nuevas, no suelen modificar la segmentación de intervenciones ajenas. En general, la creación de una nueva intervención ajena al sujeto que realiza el elemento visual es minoritaria respecto al resto de cambios. Respecto a las intervenciones modificadas, prácticamente los efectos se producen por igual en intervenciones propias y ajenas, pero con diferentes tendencias en cada causa.

Con el objetivo de extraer los datos más significativos de este análisis, hemos realizado un árbol de decisiones que relaciona los tres tipos principales de cambio (NI, MV, NMV), las causas y los efectos en la intervención propia o ajena para encontrar relaciones que no se aprecien a simple vista en los gráficos previos (Gráfico 12).

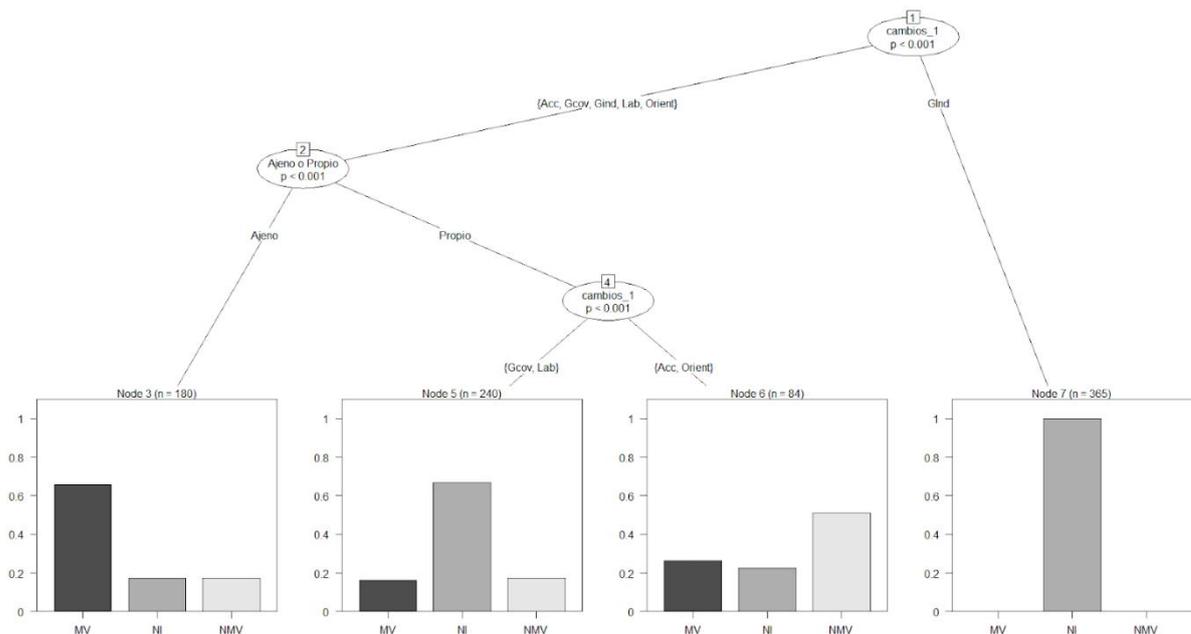


Gráfico 12. Árbol de decisiones que relaciona el tipo de cambio social, las causas y los efectos en la intervención propia o ajena.

En el árbol se observa que en los datos hay un gran nodo de nuevas intervenciones que son gestos independientes, después, todas las causas se distinguen entre su efecto en la intervención ajena y la propia. Por un lado, todos los cambios modifican generalmente la validación de una intervención verbal ajena; por otro lado, los efectos en las intervenciones propias dependen de las causas. Las nuevas intervenciones causadas por gestos independientes ya se han extraído en el primer paso; después de esta causa, los gestos coverbales y la lectura de labios suelen afectar con mayor frecuencia a la aparición de nuevas intervenciones. Por otro lado, el conjunto de la orientación corporal y las acciones ha afectado más a cambios en el tipo de intervención, pero no a su validación social.

Centrándonos en los gestos, que son los elementos visuales que cumplen funciones similares a los elementos lingüísticos, los 360 gestos independientes encontrados en el corpus causan el 49% de los cambios en la segmentación y los 93 gestos coverbales, el 12%. Para comprender mejor cómo han producido los cambios, en el próximo apartado, se realiza el análisis de la función, la posición y la forma de estos gestos.

6.2. Análisis de gestos independientes y coverbales

En el primer análisis hemos visto que los gestos coverbales suponen 102 cambios en la segmentación de la conversación (12%), mientras que los gestos independientes producen 425 cambios (49%). De estos últimos, la principal razón es que 360 gestos forman 351 intervenciones únicamente gestuales. Debido a este peso en los cambios de la segmentación dialógica y a las propiedades segmentales que comparte con el habla esta modalidad (en tanto que compatibles con las unidades conversacionales de Val.Es.Co.), vemos importante averiguar más detalles sobre la función y la forma de estos gestos.

En este segundo análisis observamos las funciones que cumplen en la conversación los gestos coverbales e independientes que modifican la segmentación dialógica, así como sus formas. Para clasificar los gestos del corpus en las funciones preestablecidas (referencial, pragmática performativa, pragmática interactiva, pragmática modal, pragmática discursiva) (§5.5.4.), hemos considerado su posición y su uso en la secuencia en relación con el habla y otros gestos y su forma a través de todos los casos, buscando patrones que formen criterios de clasificación, pues cada factor por sí mismo (forma, posición, contexto secuencial) no es suficiente para establecer la función.

En primer lugar (§6.2.1.), mostramos las funciones generales más frecuentes anotadas en el corpus, así como su reparto entre la posición independiente y la coverbal. Después dividimos estas funciones en funciones menores encontradas a lo largo del corpus y explicamos cada una para que se comprendan los motivos por los cuales los gestos se han clasificado en una u otra función.

En segundo lugar (§6.2.2), aunque al explicar las funciones del gesto se han mostrado las posiciones que ocupan en la intervención y en la secuencia que, junto a las formas, contribuyen a la clasificación funcional; se observa únicamente la relación función-posición para encontrar algún patrón.

Asimismo, en tercer lugar (§6.2.3), se observan todos los datos desde la perspectiva formal para averiguar qué funciones suelen realizarse con una parte del cuerpo o una configuración de movimiento concreta. Así podemos atender a la multifuncionalidad de los gestos y encontrar la forma como criterio para distinguir ciertas funciones o vincular las similitudes entre ellas a tipos de funciones similares.

Por último (§6.2.4), relacionamos las variables posición y forma para ver si hay partes del cuerpo que favorezcan la posición independiente o coverbal.

6.2.1. Función

En este primer recuento observamos que, como era esperable, la mayoría de los gestos independientes y coverbales que afectan estructural y socialmente en la segmentación dialógica son pragmáticos.

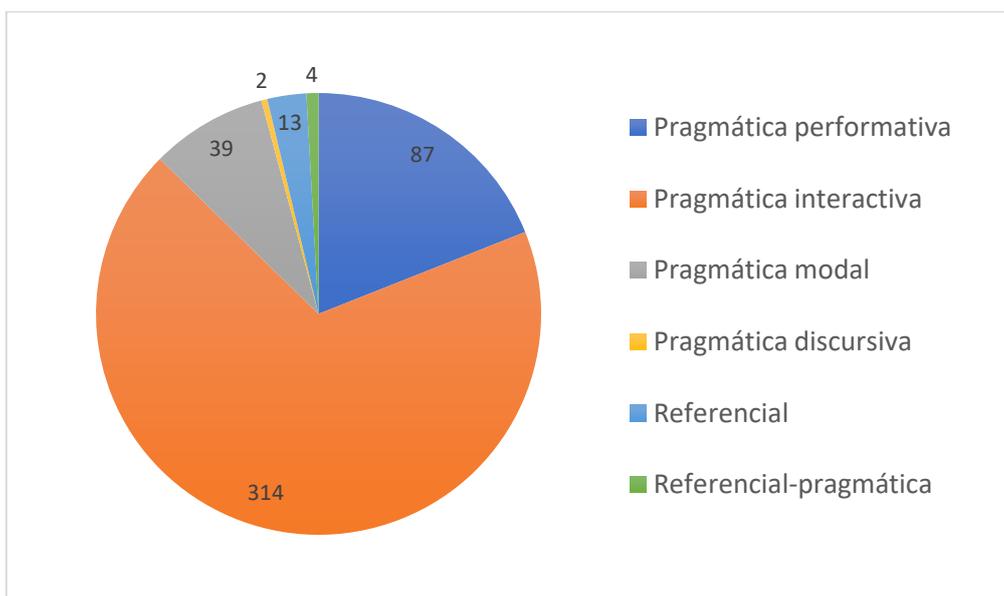


Gráfico 13. Funciones del corpus.

Se cumple nuestra hipótesis de que la función interactiva es la más frecuente, con gran margen de las otras funciones frecuentes: la performativa y la modal (Gráfico 13). Por su número, son las únicas funciones de las que podríamos proponer un comportamiento típico. El resto de las funciones son anecdóticas en un corpus de 131 minutos, así que podemos afirmar que no son las que suponen mayores cambios en la segmentación de las unidades dialógicas.

La función referencial atiende a los gestos que contribuyen al contenido semántico de la intervención, pero también se han incluido categorías híbridas cuando esta función se combina con un efecto interactivo o performativo, es decir, gestionando la alternancia de turnos o contribuyendo a distinguir el tipo de acto. Estos 4 casos híbridos son marginales y podrían unirse tanto a la categoría de referencial como a la de interactivo o performativo.

La clasificación de los gestos del corpus en estas funciones ha venido de la formulación de funciones menores que ayudaran a determinar la función general a la que pertenecían. Este proceso forma parte de las fases 3 y 4 de la MGA (§5.5): primero, se analizan los gestos en su contexto local para, después, comparar funciones similares en diferentes contextos de uso y distinguir así una función principal independiente de variantes contextuales. Por tanto, hemos identificado una serie de funciones que encajan en las funciones mayores preestablecidas (Gráfico 14).

El seguimiento es la función más frecuente (152 gestos), la que hace que la función interactiva despunte, además del tercero terminal (73 gestos) y el acuerdo (36). También como gestos performativos y modales, son frecuentes las evaluaciones (52) y las muestras de afiliación (31).

Algunas microfunciones se repiten en distintas macrofunciones, pues no solamente una misma forma puede tener distintas funciones, sino que, dependiendo del contexto, hay elementos pragmáticos multifuncionales que combinan dos funciones y en unos casos una tiene más importancia que otra (Pons Bordería 2022). Esto ocurre con el acuerdo, el desacuerdo, el desconocimiento y la comprensión. Los dos primeros porque también se realizan mediante formas referenciales y no solo gestos más abstractos y convencionalizados; los dos segundos porque pueden tener función modal o performativa (se incluyen dentro de las funciones epistémicas) según su posición. En los siguientes subapartados se explica en detalle cada caso, así como los motivos para incluir los gestos en cada categoría funcional.

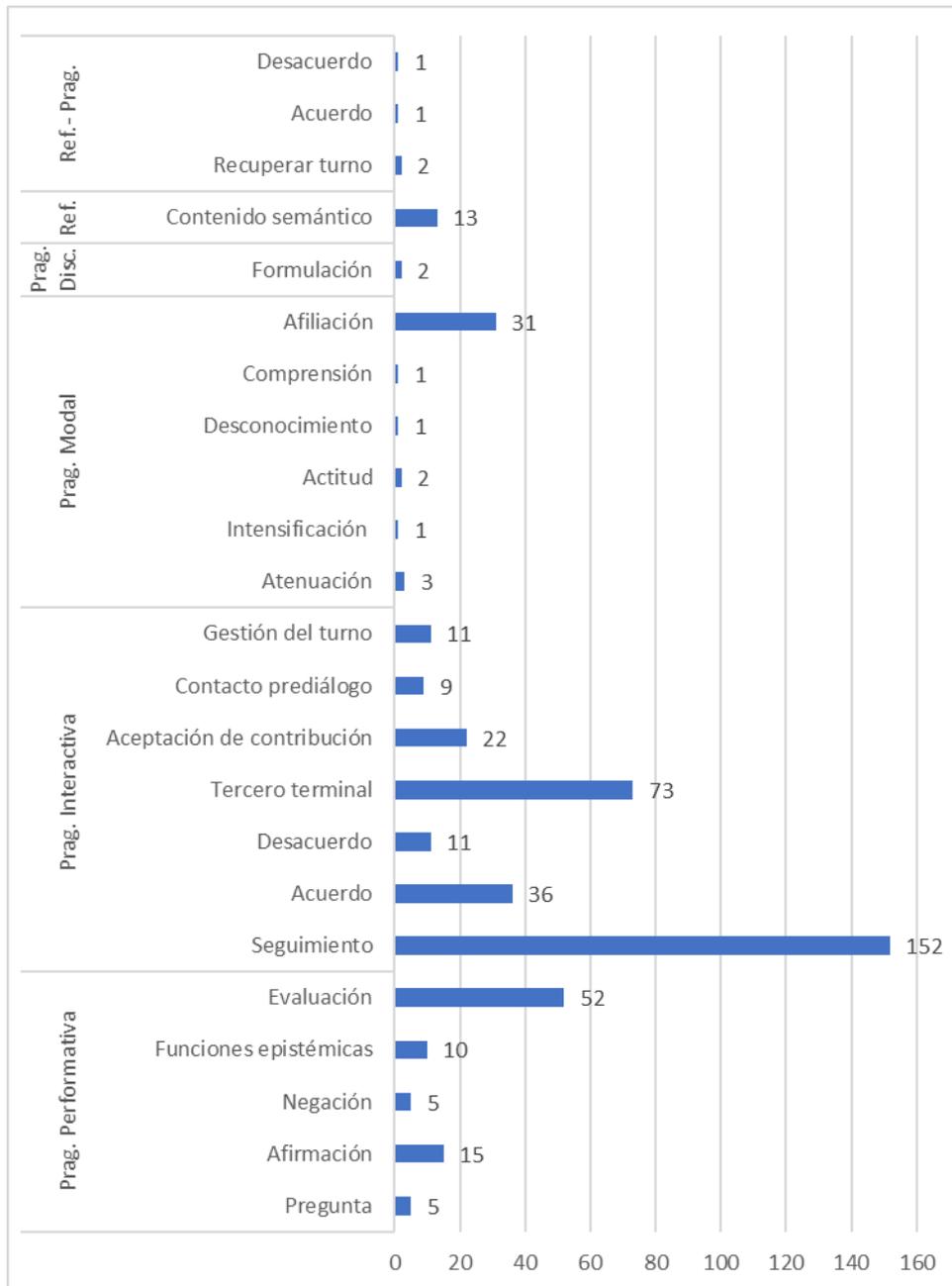


Gráfico 14. Microfunciones de los gestos del corpus.

6.2.1.1. Gestos pragmáticos performativos

Los clasificados como gestos principalmente performativos son los que acompañan al habla para realizar un acto o que por sí mismos forman el acto.

Pregunta

La pregunta gestual, ya sea como acto propio o acompañando a una pregunta verbal, no es frecuente en el corpus. Los tres ejemplos coverbales toman formas distintas que dependen del acto verbal al que acompañan.

En el ejemplo 28, una sacudida de cabeza acompaña a un “mm” y ayuda a interpretar la intervención como una pregunta de confirmación de la declaración negativa del interlocutor. Es equivalente a “¿no?”.

(28)

C: Aurora por ejemplo no quiere más

B: [m[m]]
[CabSacude]

C: [[mi ami]]ga Aurora dice que no quiere más porque le dolió un montón

(Conv5, 0:26:33)

Las cejas fruncidas (ejemplo 29) hacia el interlocutor mientras expresa una pregunta indirecta confirman que no es un pensamiento propio en voz alta, sino que busca ayuda.

(29)

1C1: y luego se fueron a trabajar/ él estaba- trabajaba en u[un/// no en Barcelona más
[CejFrunce----->

arriba]
----->]

2B1: [ComisAbajo] [°(((madre mía)) °]
[CejEleva----]

3C2: en uun

4B2: sí en un pueblo

(Conv4, 0:14:48-0:14:53)

Las cejas elevadas (ejemplo 30) sirven para repetir la pregunta que ya ha expresado verbalmente y asegurar que el interlocutor sepa que todavía quiere una respuesta

(30)

1A1: la mitad no se me han hecho// que ponía dos minutos

2B1: ((¿qué ha pasao?))

A: pero me da que no ⇒B

B: CejEleva

3A2: como tengo más bolsas no hay problema

(Conv2, 0:08:48)

Los dos casos en posición independiente realmente van precedidos por un gesto interactivo: un tercero terminal antes de pedir evaluación y un seguimiento antes de preguntar. El segundo es la intervención C1 del ejemplo 31.

(31)

1B1: y como Daniel no va a venir porque es un gilipollas

A1: (RISAS)[(RISAS) [alto y claro]]
 C1: [CejEleva-----]
 B: [[y además]] tra[baja pero sobre todo]
 C: [CejFrunce-----]
 B: [porque es gilipollas]
 2A1: [pero [sobre todo por lo pri]me]ro¹⁰⁴
 C2: [Sonrisa-----]
 ⇒A
 3C3: ¿por qué no va?
 ⇒B

(Conv5, 0:07:24)

Afirmación

Otros gestos performativos son semánticamente afirmaciones formadas por algún tipo de asentimiento. Corresponden a la segunda parte de un par adyacente que generalmente es respuesta a una pregunta confirmatoria, pero también a una duda o a una respuesta que la persona puede confirmar.

De estas afirmaciones, hay 11 casos en posición independiente (5 Ir-i y 6 Ir), todos turno, excepto 2 Ir que confirman la información de otro, así que no están validados previamente por la exigencia de una pregunta.

(32)

1A1: ¿qué os vais al campo aquí?

2D1: Asiente

(Conv1, 0:21:55)

Hay 4 casos en posición coverbal: 3 respondiendo a una pregunta y 1 a una duda. Por ejemplo, en 33, el asentimiento facilita que se distinga un “mm” casi imperceptible del mismo hablante:

(33)

1A1: hoy no me he tomado postre y me está apeteciendo ahora tomarme unos petit suisse de chocolate [()]

2C1: [¿tie]n[es?]

3A2: [Asiente] mm

4C2: pues chico cómetelo

¹⁰⁴ Entre risas

5A3: ahora ahora

(Conv3, 0:10:23)

Generalmente, la forma utilizada es el asentimiento (5 simples, 8 repetidos, 2 hacia arriba) orientado al hablante que pregunta. Pero no hay una relación clara entre la forma y la posición.

Negación

Como segunda parte de par adyacente también ha habido negaciones, pero solo 5 casos: 3 coverbales y 2 independientes. 3 responden a preguntas que piden información, 1 a la refutación de una opinión, que “Vodafone debería ir bien” (2D.Cab.3), y 1 rechaza un ofrecimiento (1A.ManCab.2). Formalmente, todos los gestos realizan una sacudida de cabeza orientada al interlocutor; además, 1A.ManCab.2 utiliza la mano también para rechazar.

Identificación	Posición	Contexto de la negación	Forma: cabeza	Forma: manos		Forma: orientación
				Orientación	Movimiento	
2D.Cab.2	Coverbal	Respuesta negativa	CabSacude			A Interlocutor
5C.Cab.54	Coverbal	Respuesta negativa	CabSacude			A Interlocutor
2D.Cab.3	Coverbal	Refutación de opinión	CabSacude			A Interlocutor
1A.ManCab.2	Indep.	Rechazo de ofrecimiento	CabSacude	PalmaVertical HaciaCentro-Frente	A Izquierda	A Interlocutor
2C.Cab.10	Indep.	Respuesta negativa	CabSacude			A Interlocutor

Tabla 36. Resumen las características de los gestos de negación.

Funciones epistémicas

Por un lado, 5 gestos independientes que transmiten el conocimiento del hablante sobre el tema tratado (conocimiento, desconocimiento, escepticismo) responden a preguntas, dudas o aclaraciones. Por otro lado, hay 5 gestos independientes que muestran comprensión, es decir, un cambio epistémico debido a una explicación de otro hablante.

Por su función epistémica, podrían vincularse a la modalización del hablante, pero, como no modaliza una intervención verbal, sino que forman por sí mismos una intervención que es segunda parte de un par adyacente, los hemos clasificado como

performativos. Los ejemplos 34 y 35 corresponden las funciones de conocimiento y comprensión respectivamente.

(34)

1B1: eso es que yo lo vi con el Samuel // con el Samuel Tapias ese que se calló del→

2C1: Asiente

B: y se quedó→ / atontao

C2: atontao

(Conv4. 0:09:55)

(35)

1B1: aunque él también dice *si las bajamos seguro quee en diez [minutos ya se las han llevado]*

2C1: [ya pero tenéis] que llamar aal [ayuntamiento]

A1: [pues se lla- se lla]ma ((se lla-)) [el ayuntamiento ()]

2B1: [no pero me refiero quee]
luego se lo lleva la [gente]

A: [solo] [tienes [que] avisarle]

C2: [Asiente----]
[Ce]Eleva--]

B: [[cuando las] cosas están] bien/ pero si no se lo
llevan→

(Conv3, 0:05:08)

Evaluación

En el corpus hay 43 gestos independientes (solamente 9 reciben reacción) y 9 coverbales con la función de evaluación. En posición coverbal acompañan a evaluaciones verbales (ejemplo 36) o a elementos paralingüísticos (*mm*, *pff*) que toman significado más claro gracias al gesto (ejemplo 37). En cambio, en posición independiente ellos solos constituyen un acto (C2 y C3 en ejemplo 38).

(36)

1C1: te cuesta un poco enseñarte a usarlas y tal pero es súper

2B1: °(ya que luego ya)°

3C2: mm

4B2: [°((qué guay)) °]

[AsienteRep-----]

[Sonrisa-----]

5C3: [Sonrisa] [no] a mí me gusta un montón

(Conv5, 0:17:15)

(37)

1B1: ee ((de- está ya eso)) cuatro años desde que acá[bamos la carrera viviendo ahí]

A1: [AsienteRep-----]

C1: [mm]
[CabInclina]
[CejEleva--]

(Conv5, 0:04:00)

(38)

1C1: ¿por qué no va?

2B1: porque es gilipollas y no quiere ir [porque]

C2: [CabSacude]
[Sonrisa]

B: odia Barcelona

A1: [oye que René [también-]]

C3: [OjosAbre-----]

B: [[y no-] y no le cae bien el- que es verdad a mí tampoco me cae muy bien

(Conv5, 0:07:27-0:07:36)

Semánticamente, estas reacciones al interlocutor expresan evaluaciones de todo tipo: sorpresa, indignación, incredulidad, hartazgo... En consecuencia, las formas de expresarlas son variadas, pero en la mayoría se utiliza la expresión facial sola o combinada con otra parte del cuerpo (42), en segundo lugar, la cabeza (32).

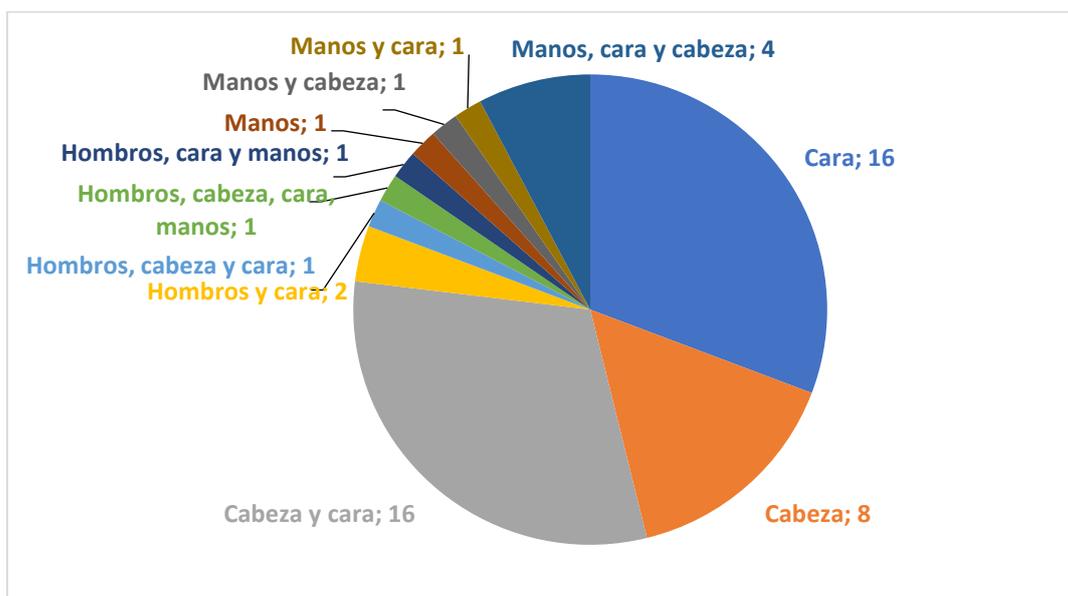


Gráfico 15. Partes del cuerpo en la función evaluativa

6.2.1.2. Gestos pragmáticos interactivos

Los gestos principalmente interactivos muestran funciones procedimentales relacionadas con la gestión de la interacción: la atención del oyente al hablante (*seguimiento*, Cestero 2000b; *marca de reacción*, Pons et al. en prensa), el acuerdo y el desacuerdo (Pons et al. en prensa), la resolución satisfactoria del par adyacente (*tercero terminal*, AC), la aceptación de una contribución, la apelación, la toma de contacto entre interlocutores antes de iniciar el diálogo, el rechazo de la toma de turno, la cesión del turno y el mantenimiento del turno... El seguimiento se diferencia de otras funciones que aparecen como Ir (entendimiento, conocimiento, evaluación) en que no aporta ningún significado adicional al mantenimiento del rol de oyente. Las reacciones del oyente que sí contienen otros significados se han categorizado como performativos o modales.

Seguimiento

Los gestos de seguimiento son los más numerosos del corpus, en su mayoría son independientes Ir (145), de los cuales solo 2 están validados después por otro interlocutor, y solo 7 son coverbales que indican seguimiento e introducen la intervención verbal propia.

Respecto a su forma, la mayoría (130) es realizada únicamente con la cabeza (los 7 coverbales son con la cabeza). Estos movimientos de cabeza son: 64 Asiente, 54 AsienteRep, 9 AsienteArriba, 2 CabInclinada. La siguiente parte del cuerpo utilizada es la cara: 8 se realizan solo con la cara, todos ellos con elevamiento de cejas y dos, además, con movimiento de boca: 11 combinan cabeza y cara (asentimiento y cejas o labios).

Destacan 3 casos en los que se utilizan los hombros o las manos porque el gesto de seguimiento repite los actos del hablante. En el primero, un hablante hace un gesto independiente juntando y separando los dedos repetidamente para referirse al sonido de las palomitas que se están haciendo en el microondas, el oyente reacciona repitiendo el mismo gesto. En el segundo, no repite un gesto del hablante, sino que representa lo que dice: el hablante está narrando una historia y pone en boca de un personaje “oye perdona”, el oyente adelanta la cabeza y encoge los hombros actuando con los gestos que imagina al personaje. En el tercero, el hablante hace una pregunta retórica “¿qué más dará?” y el oyente reproduce un gesto que podría interpretarse como una repetición de la pregunta: ceja elevada y encogimiento de hombros.

Acuerdo

Originalmente (Briz y Val.Es.Co. 2003), se incluían como SAI únicamente los recursos apelativos del hablante hacia el oyente y muestras de atención del oyente hacia su interlocutor; mientras se explicaba que las marcas lingüísticas de acuerdo o desacuerdo confirman la reacción de una intervención a otra y forman parte del proceso de aceptación o validación del turno. Después, en la categorización funcional de las partículas en el DPDE (Pons et al. en prensa) y en la segmentación del Corpus Val.Es.Co. 3.0. (Pons Bordería 2022) se incluyen las partículas que muestran acuerdo y desacuerdo como interpersonales también, pues confirman la postura del receptor como favorable o desfavorable para conectar su intervención con la intervención previa. Esta visión procedimental del acuerdo se vincula con el enfoque de la conversación como negociación (§2.2.2), que requiere del acuerdo entre interlocutores para que avance.

El asentimiento, a menudo repetido (14 de 35), es una forma común para el acuerdo. También la sacudida de cabeza (5) cuando se está de acuerdo con una proposición negativa.

El principal criterio para diferenciarlo del seguimiento es que reacciona a una opinión. Además, el asentimiento repetido aporta énfasis para que se entienda como un acuerdo y no solo seguimiento, así como que esté acompañado por elevamiento de cejas (Ejemplo 39, intervención C2).

(39)

1B1: *¿y eso te lo has cosido tú? es pa decirle fíjate lo que sorprende porque aquí son todos una pan[da de inútiles]*

D1: [Asiente-----]

B: no me refiero a- a las [drags sino me refiero]

A1: [sí que no es Maestros] de la cos[tura]

2D2: [al jurado]

3B2: [al ju- al] jurao

D3: ¡oy!

B: [digo la Locking que no la ha oído]=

C1: [AsienteArriba-----]

B: =en la vida ni Dios (RISAS) que no la conoce [ni su [puta] madre/ qué bueno [fue]]

C2: [AsienteRep-----]
[CejEleva-----]

[OjosCierra-----]

4D4: [qué-] [qué]
(Conv2, 0:12:02-0:12:16)

Respecto al desacuerdo (11 casos), esta función ha afectado a la segmentación dialógica únicamente en posición independiente. En 9 casos ha sido una Ir (intervención D2 en ejemplo 37), los otros dos forman parte del gesto complejo 5C.ManCarCab.2 (§6.2.2)

(40)

1D1: que para ellos ((sé))- yo sé que/ estuvo hablando de Marta [o sea de Marta de E]=

C1: [AsienteRep-----]

D: =va estuvo hablando de- de Eva y tal pero

2C2: y de Marta también [cuando le dice]

D2: [CabSacude----]
[CejFrunce-----]
[BocaOtro¹⁰⁵-----]

C: noo porque te comentó de-/ de que no hicieras el doctorado→ de quee- de que fueras directamente/ a dar clase ahí en la ESO tal o no sé a- algo d- así dijo

(Conv2, 0:35:53)

Generalmente, se ha utilizado la cabeza y la cara para expresar desacuerdo, aunque también se han usado las manos (Tabla 37).

	Cabeza	Cejas	Boca	Mano		
				Forma de dedos	Orientación	Dirección de movimiento
1D.Car.4		CejFrunce				
2D.CabCar.12	CabAdelante		ComisurAbajo			
2D.CabCar.11	CabSacude	CejFrunce	BocaOtro			
4C.Man.2				ÍndiceEstirado	PalmaVertical HaciaCentroFrente	Otro (Giro lateral de la muñeca)
1A.CabCar.3	CabInclinada	CejEleva				
1A.CabCar.4	CabInclinada	CejEleva	ComisurAbajo			
5C.CabCar.2	CabInclinada	CejEleva				
2D.Cab.52	CabInclinada					

¹⁰⁵ Boca entreabierta

5C.CabCar.15	CabOtro	CejEleva	Sonrisa			
5C.ManCarCab.2(1)	CabSacude			ManoPlana	PalmaArriba	
5C.ManCarCab.2(6)	CabOtro		Sonrisa	ManosPlanas	PalmasArriba	AFuera

Tabla 37. Forma de los gestos de desacuerdo

Las que combinan la cabeza inclinada y la elevación de cejas las hemos incluido dentro del desacuerdo, aunque no es un rechazo completo a la opinión del hablante, ni tampoco muestra acuerdo. Podríamos verlo como un *acuerdo parcial* o *acuerdo atenuado*, al estilo del marcador “bueno” (DPDE, Briz, Pons y Portolés 2008). No se encuentra un patrón en la orientación del gesto.

(41)

1B1: uno- una posibilidad de diez una de diez

2C1: ¡bo[o! pe-]

B: [tiene nueve] que no

C: pero es bastante

B2: ya ya

C: es mucho [[no es una] entre] cien

B3: [Asiente----]
[OjosCierra]
[Mano¹⁰⁶-----]

A1: [CabInclina-----]
[CejEleva-----]

B: no no no

(Conv1, 0:14:08)

A1 sopesa con su gesto la valoración de C de que una probabilidad de 10 es mucha probabilidad, no muestra ni acuerdo ni desacuerdo; por el contrario, B da la razón a C con B2 y B3, utilizando el asentimiento en el segundo.

Tercero terminal

La muestra interactiva de seguimiento que se produce mientras un hablante está realizando su turno y el oyente reacciona a lo largo de él es un uso que distinguimos del tercero terminal, que confirma que un par adyacente se ha resuelto satisfactoriamente.

¹⁰⁶ Mano: PalmaVertical HaciaCentro AFuera

El tipo de pares adyacentes que cierran en el corpus son de pregunta-respuesta, pregunta-respuesta-entendimiento, explicación-entendimiento, explicación-confirmación, aclaración-conocimiento, duda-confirmación, entendimiento-confirmación, información-seguimiento, información-evaluación, evaluación-acuerdo, evaluación-sorpresa, evaluación-humor, sugerencia-aceptación. Salvo pregunta-respuesta, el resto de las funciones en la segunda parte del par podrían considerarse Ir del oyente no turno normalmente, pero el tercero terminal del hablante los valida; en el caso de los coverbales, antes de continuar el turno. En posición coverbal se producen 20 y en independiente, 48.

Al igual que el seguimiento, suele realizarse solamente moviendo la cabeza (50). 6 se combinan con la cara (cejas elevadas o algún movimiento de labios), 1 con hombros; 1 Manos, Hombros y Cabeza y 1 Manos. El movimiento de cabeza más común, combinado con otro movimiento o no, es Asiente (39), seguido de AsienteRep (13), AsienteArriba (4), CabInclianda (1). También hay sonrisas solas (8), especialmente cuando el par incluye una respuesta de humor o la pregunta o la respuesta se ha formulado con sonrisa. Generalmente, se realizan orientándose hacia el interlocutor (42) o retirándose de él (18) (6 a nadie).

Sin embargo, hay ciertos movimientos de cabeza que no son verticales sino diagonales u horizontales que desvían la mirada del oyente y que posicionados después de la retroalimentación nos plantean la posibilidad de que no sean realmente una reacción a esta. No consideramos que respondan validando de alguna forma la retroalimentación, sino que es un movimiento monológico que le sirve al hablante para confirmarse a sí mismo el final de una idea ya desarrollada y retomar su discurso. Al estilo de algunos marcadores fáticos que funcionan “como reafirmación o justificación del yo, de su actuación o de lo dicho, es decir, fórmulas autorreafirmativas que refuerzan o justifican razonamientos de los hablantes ante su(s) interlocutor(es), sean argumentos o conclusiones” (Briz 1998: 225). Véase el ejemplo 39.

(42)

1B1: y el otro día dijo que estaba aquí en Villena y fue como *vale venga pues t[al]*

A1: [mm]

B: =*pues nos lo hacemos conti[go]*

A2: [AsienteRep] [AsienteRep]
⇒B

C: [Asiente] / y el diseño → o sea este
 ⇒ ∅

nos ha molao el primero que me pasó no me gustó luego nos pasó otro

(Conv. 5, 0:20:54)

Aceptación de contribución

El seguimiento se ha restringido en este estudio a la validación de una información en curso o terminada, pero nos ha resultado interesante marcar con otra etiqueta cierto uso de esta validación. La *aceptación de contribución* sirve para marcar dos supuestos:

- 1) El hablante está dando una información y otro participante toma el turno momentáneamente para añadir información pertinente que contribuye a la información del hablante, entonces el hablante hace algún gesto con el que acepta esa contribución y continúa con su turno o continúa otro interlocutor.
- 2) El hablante está dando una información y otro participante toma el turno momentáneamente para añadir información pertinente que contribuye a la información del hablante, entonces un tercer participante, un oyente, hace algún gesto con el que acepta la contribución del segundo hablante. Es una estructura que solo puede aparecer en conversaciones de varios participantes.

Esta función interactiva no dista mucho del seguimiento o del tercero terminal, pero encontramos que algunos casos diferían en este comportamiento del resto y vimos pertinente esta categoría. Mientras el seguimiento valida al hablante principal, esta aceptación de contribución valida al secundario. Respecto a la diferencia con el tercero terminal, la contribución del segundo hablante no encaja como acto de acuerdo, evaluación, confirmación, etc. Ambos actos, el del primer hablante y el del segundo, son de información, formando, en todo caso, un par información-información.

En el corpus, este gesto se realiza con un asentimiento simple o repetido orientado al interlocutor que ha dado la contribución. No es muy frecuente, 10 de ellos son coverbales y 12, independientes. Los coverbales siempre los realiza el hablante principal, que continúa su turno (ejemplo 43). De los independientes, 8 son realizados por la tercera persona oyente que no toma el turno (5 de ellos del hablante 2C) (ejemplo 44).

(43)

1C1: porque claro la gente sí que quiere hacerlo pero ne- falta mucho pa que [se pueda ((hacer))]

2A1: [muchos medio]s

3C2: AsienteRep

A2: Asiente

C: porque es que los profesores/ estamos viendo un montón de casos que cogen dicen no le das una ficha que pinten y ya está

(Conv. 5, 0:14:58-)

(44)

1D1: (BALBUCEO) *he llegao un poquito más tarde porque he tenido: un percan- un percance familiar [((tal))]*

2B1: [un perca]n- per- ¿usó esa pa[labra?]

3D2: [dij-] un percance no- primero dijo un percance
[antes de venir]

4A1: [()]

C1: [Asiente]
⇒A

5D3: [pero] es que luego dijo↑ un problema familiar cuando llegó allí y dicen

(Conv2, 0:23:52)

Su baja frecuencia se debe a que la función está restringida a unos supuestos muy concretos. A menudo, la información que da el segundo hablante sí supone un acto de evaluación, acuerdo u otro y su validación se etiqueta como tercero terminal; o la emisión es suficientemente larga para considerar que se valida con un seguimiento. Por consiguiente, la aceptación de contribución no afecta a tantos casos como otros gestos interactivos, pero ha resultado necesaria porque no correspondía a la descripción de las otras funciones interactivas.

Contacto prediálogo

Lo que hemos llamado *contacto prediálogo* corresponde a la presecuencia de invocación del AC (Schegloff 2007: 48, Lerner 2003: 188): el nombre del receptor tiene función al inicio de turno como presecuencia de invocación para asegurar que se establece el contacto entre los participantes que posibilita la secuencia principal. En el estudio de Lerner, tras recibir el cambio de mirada y/o postura del oyente hacia el hablante, este último inicia la secuencia. Poyatos (1996: 230-246) también menciona el ofrecimiento y la petición del turno entre oyentes, mediante gestos deícticos, enarcamiento de cejas y/o el contacto ocular para ofrecer y miradas al cooyente, apertura labial y/o ligero enarcamiento de cejas para pedirlo.

En el corpus, hemos encontrado algunos gestos que parecen cumplir una función

similar, tanto iniciando el contacto como aceptándolo. Hay dos tomas de contacto que las hemos categorizado como coverbales, ya que la aceptación del oyente la consideramos una Ir y el gesto del hablante es coverbal al resto de su intervención verbal (ejemplo 45).

(45)

(Inicio de turno)

1C1: CejEleva CejEleva [CejEleva-----] [pues vaya / entonces]
 OjosAbre [OjosAbre-----]
 CabInclinada [CabInclinada]

B1: [AsienteRep-----]

C: no van a aparecer

(Conv5, 0:06:03)

En posición independiente hay 7 gestos. En dos casos (ejemplo 46), la toma de contacto (1D1, 1C1) es un gesto independiente porque luego el hablante no llega a desarrollar el turno verbal, pero tampoco el receptor toma el turno verbal (2C1, 1D1), por lo que se queda como un diálogo paralelo entre oyentes.

(46)

1B1: ¿estás- estás regulinchi?

2D1: tengo frío

3B2: [ah ay eso también me ha pasao]

A1: [AsienteRep-----]

B: antes a las cuatro que era como [ten[go frí]o]

1D1: [Encoge-----]

2C1: [CabInclinada]

B: tengo calor tengo frío

1C1: CejFrunce
 ComisurasArriba

2D1: CejEleva

D2: [((la vida))]

B: [¿no te has to]mao algo?

4D3: no/ me iba a tomar un para[cetamol ahora si mee-]

A1: [tóma]telo

D: si me ponía muy mal

(Conv5, 0:04:40-0:04:57)

Las formas de los pocos gestos que hay en el corpus son variadas, pero se puede intuir que la toma de contacto utiliza movimientos que llaman la atención del receptor

como cejas elevadas, la cabeza inclinada o las manos, y la aceptación utiliza formas similares a otros gestos interactivos reactivos (asentimiento, sonrisa) o que muestran indeterminación para no tomar el turno (encoger hombros, inclinar la cabeza).

		Cabeza	Cejas/ojos	Boca	Hombros	Manos		
1A.Man.1	Coverbal					PalmaAbajo- HaciaFrente	Aleteo	Toma de contacto para iniciar diálogo
5C.HomCabCar.3	Coverbal	CabInclinada	OjosAbre		Encoge			Toma de contacto para iniciar diálogo
1D.Car.7	Indep.		CejEleva					Toma de contacto prediálogo
5C.Car.12	Indep.		CejFrunce	Labios- Retraidos				Toma de contacto prediálogo
5D	Indep.				Encoge			Toma de contacto prediálogo
5B.Cab.20	Indep.	AsienteRep						Acepta contacto prediálogo
1D.Car.1	Indep.			Sonrisa				Acepta contacto prediálogo
5C.Cab.51	Indep.	CabInclinada						Acepta contacto prediálogo
5D	Indep.	CabInclinada	CejEleva					Acepta contacto prediálogo

Tabla 38. Forma de los gestos de contacto prediálogo.

Ya que en nuestro estudio la mirada o la postura corporal no son gestos por sí mismos, no los consideramos gestos de inicio o aceptación del diálogo. Generalmente, se produce el encuentro de miradas entre los interlocutores y uno de ellos tantea al otro con un gesto para iniciar el diálogo o animarle a iniciarlo.

El gesto con la mano (1A.Man.1) ilustra que la manera de llamar la atención del receptor debe ser más visible que la de un movimiento de cabeza si el hablante no está recibiendo la mirada del receptor. En el ejemplo 47, C está escuchando atentamente el turno de B, así que A necesita llamar la atención de C para que escuche su pregunta: usa las manos, no la cabeza o la cara. El primer movimiento de mano, junto a “oye” repetido sirve para apelar a C. Sin estos gestos coverbales de A, no se hubiera encontrado este dialogo paralelo.

(47)

1B1: es- es una prueba como que compruebaan varios factores y hay [dos que han [sali]=

2A1:

[Mano¹⁰⁷-----
 ⇒C
 [((oye))]

B: =do posible [eso sig]=

A: Mano-----
 [((oye))]

B: =nifica que hay una pro]ba[[bi[li- una de]

A: Mano-----] [Mano¹⁰⁸-----
 [OjoOtro¹⁰⁹-----

A: [()]

B: =diez] de que] [sea efec]=

A: Mano-----]
 Ojo--]

3C1: [((¿te lo hago?))]
 ⇒A

B: =ti[vamente] Síndrome de Down

A: [AsienteRep]

(Conv1, 0:12:50-13:01)

Gestión del turno

En una menor cantidad, encontramos otras funciones interactivas relacionada con la gestión del turno: cesión (5), mantenimiento (3), rechazo (2) y toma (1).

La cesión y el mantenimiento son siempre gestos coverbales porque suponen que el hablante ya tiene el turno. El rechazo siempre es independiente porque rechaza tomar el turno, por tanto, no continúa la intervención verbalmente. La toma puede ser coverbal o independiente, dependiendo de si cumple su intención; en el único caso del corpus, no lo consigue, así que es independiente.

La cesión se realiza al final del turno y puede combinar cabeza, cara y hombros con configuraciones fijas de cada parte. Las cejas elevadas las vinculamos con la apelación al interlocutor; la cabeza inclinada, que ya hemos visto en los gestos de acuerdo

¹⁰⁷ PalmaAbajo DedosAFrente Aleteo

¹⁰⁸ Mano: Señala

¹⁰⁹ Ojo: Señala

parcial, y el encogimiento de hombros, visto también en gestos de desconocimiento, junto con los labios retraídos dan vía libre al interlocutor para que dé su opinión.

Identificación	Parte del cuerpo	Cabeza	Cejas	Boca	Hombros	Orientación
5C.HomCar.2	Hombros y Cara		CejEleva	ComisurAbajo	Encoge	aInterlocutor
4C.HomCar.1	Hombros y Cara	CabInclina	CejEleva	LabRetrae	Encoge	aInterlocutor
5C.CabCar.3	Cabeza y Cara	CabInclina	CejEleva	LabRetrae		aInterlocutor
5C.Car.8	Cara		CejEleva	LabRetrae		RetiradaDesde Interlocutor
5C.HomCabCar.4	Hombros, Cabeza y Cara	CabInclina		LabRetrae	Encoge	aInterlocutor

Tabla 39. Forma de los gestos de cesión del turno.

En el ejemplo 48, la cesión del turno de C hace que el asentimiento de A4 no forme parte del siguiente turno, sino que sea una Ir y que 2A5 sea la reacción al gesto de cesión.

(48)

1C1: porque es que los [profesores]

A3: [CejEleva-]

C: estamos viendo [un montón de casos que co]=

A: [CabSacude-----]

C: gen dicen *no le das u[na ficha que pinten y ya está]*

A4: [AsienteRep-----]
[CejEleva-----]

**C: CabInclina
CejasEleva
LabRetrae**

2A5: ya no y no le da lo que necesita

(5C.CabCar.3, Conv5, 0:15:09)

Los que mantienen el turno verbal son necesariamente gestos coverbales que mantienen dos emisiones verbales como parte del mismo turno. Los tres ejemplos encontrados utilizan la mano, con la palma vertical o hacia abajo (ej. 49), pero dirigida al frente para frenar a los interlocutores, o la boca abierta, la [a]-face (Streeck y Hartge 1992).

(49)

1C1: son/ cuatro meses y medio o así será [() se[mana]]

2D1: [((vamos a [ver]))]

[Mano¹¹⁰----->

3B1: [[por ahí] por ahí no estoy segura]

D: [((vamos a ver))]

----->]

C2: [[() ()]

⇒D

B: [[pe]ro el caso es que le han dicho que] está como en el filo ahí que a lo mejor no puede

(1D.Man.1, Conv1, 0:15:21)

La toma de turno utiliza una articulación de mano similar a las del mantenimiento.

En el único caso, C no consigue tomarlo (ejemplo 50).

(50)

1B1: yy [le- y le dieron seis]=

C1: [pero hace muchos años]

B: =meses de vida

A1: mm

B: dice *eso se va volver* [a reprodu[cir] en- en seis meses]

C2: [Mano¹¹¹-----]

A: [((dice))] *difícil quee*] no nos hayamos dejao [algo]

B: [y se fue]ron

a Lourdes

(4C.Man.3, Conv4, 0:11:33)

Los dos rechazos del turno se dan repitiendo el gesto elicitor del interlocutor, B devuelve la elicitación a C (ejemplo 51), y cuando el interlocutor no está interesado en seguir hablando del tópico (ejemplo 52).

(51)

1C1: ¿en tu coche?

2A1: sí sí he venido en- en el mío

¹¹⁰ PalmaAbajoDiHaciaFrente

¹¹¹ ManoPlana PalmaVerticalHaciaFrente

3C2: [Asiente-----] [CabInclina]
 [ComisurasAbajo] [CejEleva--]
 ⇒B

4B1: CabInclina

5C3: lo valientes que son ahora

(4B.Cab.2, Conv4, 0:01:58)

(52)

1A1: porque [la forma que tiene]

2C1: [ahí a las últimas de la]s- las últimas de los casos lo pones encima y les
 guardas las patas y cuando te cambies de casa [OjoCierraB]
 [BocaOtro¹¹²-]
 ⇒B

A2: [mm]

3B1: Asiente
 CejEleva

C2: Sonrisa

A3: [pero vamos]

C: [te pones] las patas

A4: CabSacude

(3A.Cab.1, Conv3, 0:18:06)

6.2.1.3. Gestos pragmáticos modales

Tan solo hay 8 gestos en posición coverbal que impliquen algún cambio en la segmentación dialógica. Generalmente, todas las funciones modales acompañan a una intervención verbal, añadiendo a su interpretación algún valor modal como, en los casos encontrados en el corpus, la atenuación, la intensificación, el desconocimiento, la comprensión o una actitud.

En el ejemplo 53, hay dos gestos modales de una misma intervención. Mientras B se levanta para dirigirse a la cocina, hace un movimiento de mano hacia A, que junto a la prosodia y llamarla “señora” le añade una actitud de desdén a la intervención; cuando pasa cerca de A de camino a la cocina, le toca el brazo y sonrío, atenuando lo que acaba de decir para que se interprete como una broma.

¹¹² Una comisura abajo.

(53)

1B1: yo quiero tomarme algo

D1: pues [tómeselo]

2A1: [((tómatalo))]

C1: [hombre]

D: ni que estu[viera en [su casa]]

3B2: [no me ha [sacaa nada se]]ñora
 [Mano¹¹³-----] Mano¹¹⁴

4A2: pero- per- ¿de bebida? pues [es tu ca]=

5B3: [yo qué sé]

A: =sa cógete

(Conv2, 0:05:15-0:05:22)

Obsérvese que ya habían aparecido el desconocimiento y la comprensión como gestos performativos en posición independiente, en cambio, en posición coverbal, los consideramos modales epistémicos. En el ejemplo 54, el gesto de B añade este sentido de desconocimiento a la pregunta que formula.

(54)

1B1: ((me-)) era como me decía *¿qué tamaño [lo queréis?]*

2A1: [Mano¹¹⁵-----]
 [Encoge-----]
 [CejEleva---]

B: =[y yo]
 [Mano¹¹⁶----->
 [OjosAbre]
 [BocaOtro]

A: [hombre tam[bién lo que tú veas que quieres tener]]

B: [[¿((cómo es más bien de))] grande?
 ----->]

(Conv5. 0:25:52)

¹¹³ DedosEstirados PalmaArribaDiagonal Arriba Aleteo

¹¹⁴ ManoOtro: Toca brazo de A

¹¹⁵

¹¹⁶

Por otro lado, hay 31 gestos en posición independiente que reaccionan al hablante con sonrisas o risas únicamente. Esta forma, que no va acompañada de asentimientos u otro movimiento, nos ha planteado qué está transmitiendo el emisor. Podría ser otra forma de indicar seguimiento o acuerdo o de evaluar algo dicho como gracioso, entonces los categorizaríamos como interactivos o performativos. Sin embargo, se pueden interpretar también como elementos modales que contribuyen a reforzar la relación entre los interlocutores.

Sin haber hecho un análisis minucioso de las intervenciones que provocan estas sonrisas, en general, se puede observar que suelen ser imitaciones de personajes producidos durante una narración e intervenciones dichas entre risas o con un tono jocoso, quizás necesario para que no se malinterprete como un comentario que pueda causar conflicto ("no hace falta que te den el premio al mejor disfraz del año", "ponía dos minutos pero me da que no", "¡pues a la cárcel!", "pero tampoco los vas a denunciar", "esperemos que sea tatuaje", "yo me lo hubiera sacado antes que él", "tengo ahí supurando el pus", "me puede dar un empacho", "que lo apuntes", "estás de buen ver guapa")

Entre otras funciones en la conversación, como cubrir necesidades psicológicas o ejercer poder, el humor construye solidaridad entre los interlocutores (Hay 2000). El hablante la usa para "mantener su imagen positiva y para defender su imagen negativa ante el oyente" (Alvarado Ortega 2012). Cuando el hablante pretende ser gracioso, las risas del oyente afiliativas refuerzan su imagen (Alvarado Ortega y Cortés Rodríguez 2017). Pero no siempre hay intención humorística de partida, sino que también "el humor surge como consecuencia de la oposición de guiones semánticos y de la violación del Principio de Informatividad y de Manera" (Alvarado Ortega 2012), creando complicidad entre hablante y oyente. Realmente, la risa aparece más asociada declaraciones y comentarios que a bromas (Provine 2004, Scott et al. 2014).

También hay risas de castigo o ataque (Alvarado Ortega y Cortés Rodríguez 2017, Alvarado Ortega 2021), con las que el oyente se burla del hablante, rebajando lo que dice, atacando a su imagen pública. Tanto las de afiliación como las de castigo aparecen en el Corpus Val.Es.Co. (Briz y el Grupo Val.Es.Co 2002b) y el Corpus Val.Es.Co. 2.0 (Cabedo y Pons) en intervenciones verbales del que reacciona; las de castigo especialmente entre hombres (Alvarado Ortega 2021). En nuestro corpus, generalmente, la reacción al humor es positiva hacia el interlocutor, al menos en los gestos independientes, pero no es un factor que hayamos medido.

Antes que la causa humorística de la sonrisa, que podría llevarnos a la exigencia del par chiste-risa; lo que hace que nos decantemos por su función modal es su uso en la creación de complicidad y el refuerzo de la imagen del hablante. En consecuencia, aunque en otras funciones, como las epistémicas, la posición independiente ha sido un criterio para catalogarlas como performativas, creemos que esta función afiliativa de la sonrisa en posición independiente tiene una razón de ser modal. Al igual que el seguimiento o el acuerdo independiente tienen un valor predominantemente interactivo.

6.2.1.4. Gestos pragmáticos discursivos

El único gesto discursivo que ha modificado la segmentación dialógica en dos ocasiones es formulativo. Uno es un soplo con la mirada desviada con el que se interpreta que no sabe cómo formular una respuesta a lo que se le ha preguntado (ejemplo 55, 6C3). Otro sigue un sonido ininteligible, intentando encontrar las palabras moviendo las manos de un lado a otro (ejemplo 56). El primero es independiente y el segundo coverbal, pero ninguno aparece en solitario porque realmente el independiente acompaña a otro gesto (performativo: desconocimiento) con el que forman una intervención gestual.

(55)

1C1: pues mira () otra opción que a veces voy allí y hay d'eso// y luego no hay un arroz que me guste *venga el arroz y el arroz* anda ee// () (RISAS) ((este)) a[rroz]

D1: [haber] abierto
la ventana a[yer]

(C y A cruzan miradas)

C: [¿eh?]
⇒A

2A1: [sigues] enfadá con la- con la [que-]

3C2: [()]

D: [esta noche]

4A2: [¿por qué?]

5B1: [()]

6C3: Encoge BocaOtro¹¹⁷

A3: [()]

B: [no-no me-] no me- ¿de qué?/ ¿con alguien de la carnicería?

¹¹⁷ Sopla

(Conv1, 0:07:01)

(56)

1B1: la abuela de René ochenta- noventa años dice *yo no quiero ir a una residencia que están s- está lleno de viejos*

C2: hi [()]

B: [y ella no] está- no forma parte de e[se grupo]

C3: [(RISAS)] (RISAS)// °(está lleno de vie-)°

D1: ((mm)) Manos¹¹⁸ [Manos¹¹⁹-----]
OjoCierra

B: [el otro día cumplió] el- el herma- ell abuelo de René noventa y cinco

(Conv1, 0:30:12)

6.2.1.5. Gestos referenciales

Los gestos que se han categorizado como principalmente referenciales son los que aportan un contenido semántico al acto verbal (4 casos) o gestual principal (4 casos) al que acompañan del propio hablante o de otro hablante (3 casos).

Nos referimos a un acto gestual principal porque hay un caso de un turno gestual complejo en el que aparece un gesto performativo, cuatro gestos deícticos que señalan a distintas partes del cuerpo y un nuevo gesto performativo para cerrar. Entonces, 4 casos de los gestos independientes referenciales forman parte de este turno, el cual describimos con más detalle en §6.2.2. Pero hay otros 3 gestos independientes referenciales. Dos colaboran con el contenido verbal de otro hablante: uno es replicado por el receptor y, por tanto, validado; pero el otro no, es una Ir; el último es la respuesta a una pregunta aclaratoria.

Todos los gestos referenciales se realizan con acciones manuales. Cuatro son representativos e indican un tamaño, una relación temporal o una acción física; o son gestos deícticos que señalan a un objeto de la habitación. Aparte del turno gestual complejo mencionado y la respuesta aclaratoria, los deícticos también se han utilizado para acompañar una petición verbal (coverbal) para que el receptor dé un objeto de la habitación.

¹¹⁸ PalmaLateral HaciaCentro AIzquierda

¹¹⁹ PalmaArribaDi HaciaCentro Arriba alternativamente

6.2.1.6. Gestos referenciales-pragmáticos

Para algunos gestos independientes, consideramos que la función referencial se combina con otras funciones pragmáticas interactivas o performativas. En dos casos continuados, el hablante 1C intenta recuperar su turno y para ello adelanta la intervención con un gesto representativo de objeto sobre el que va a hablar. Por este uso para tomar el turno, lo catalogamos como gesto referencial-pragmático. Como ninguna de las veces consigue tomar el turno, los gestos quedan en posición independiente.

Respecto a la función referencial-performativa, los dos casos son deícticos y sustituyen a otra modalidad. En uno, 2D señala con el dedo para mostrar acuerdo hacia 2A porque está bebiendo y no puede hablar ni realizar ningún gesto con la cabeza o la cara. En otro, 1D señala con la cabeza y los ojos porque está sosteniendo su comida con las manos, lo realiza para llamar la atención a 1B sobre 1C para expresar que no está de acuerdo con su actitud o le exaspera.

6.2.2. Función y posición

En total hay más gestos independientes que coverbales y, respecto a la distribución de estos en las funciones, el 93% de los gestos ocupan funciones que tienen más gestos independientes que coverbales.

Si tomamos las macrofunciones, sin distinguir microfunciones concretas, la distinción entre independientes y coverbales queda como en el Gráfico 16. Las barras muestran el porcentaje de gestos coverbales e independientes que tiene cada función, pero la tabla de datos indica que hay una gran diferencia en el número de muestras de cada función.

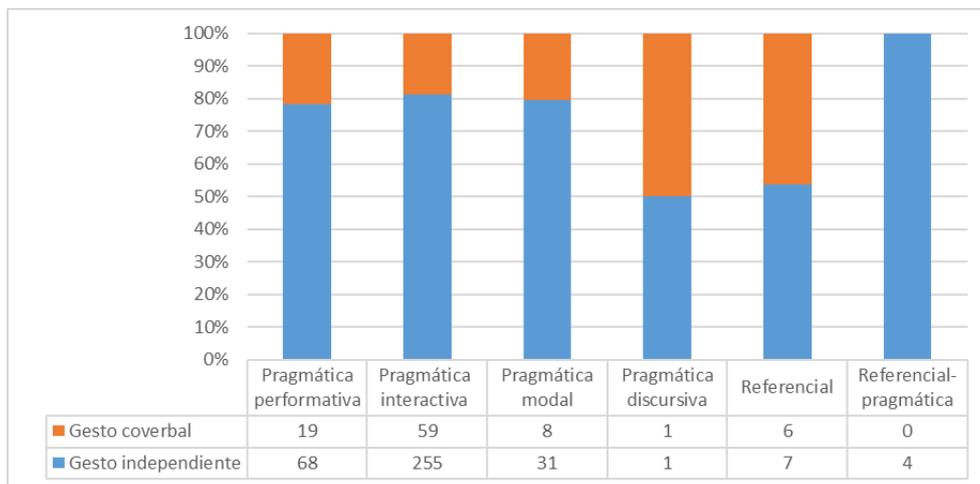


Gráfico 16. Macrofunciones de gestos independientes y coverbales

Ordenando las microfunciones mostradas en el apartado anterior por su porcentaje de gestos independientes, extraemos los siguientes resultados sobre la relación entre posición y función.

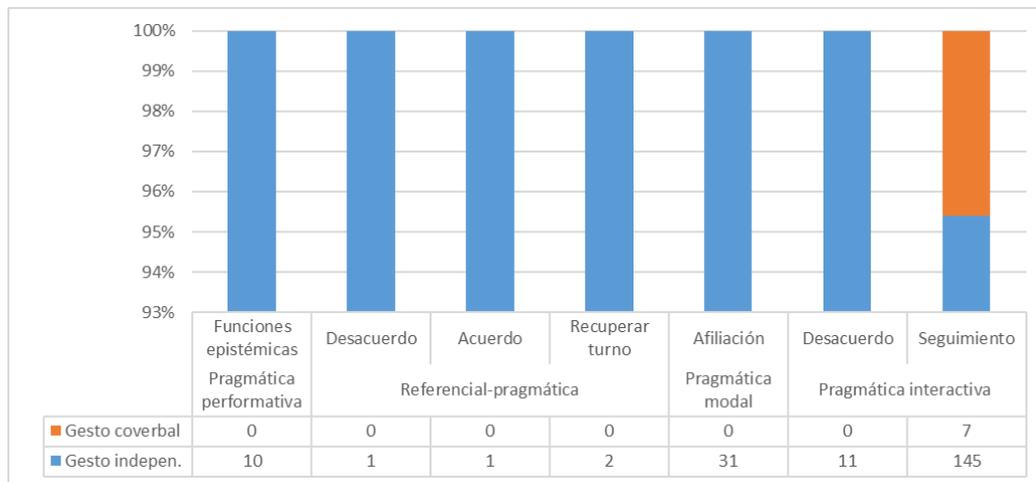


Gráfico 17. Funciones con posición independientes predominante.

Las funciones que siempre ocupan la posición independiente son la función epistémica performativa, la referencial-pragmática, la afiliativa y el desacuerdo (Gráfico 17). Precisamente, estas funciones epistémicas, por ocupar esa posición independiente, se convierten en el único acto de la intervención y son capaces de responder a preguntas o dudas, mientras que en posición coverbal se han clasificado como gestos modales. En consecuencia, en estas funciones, la relación posición-macrofunción es directa.

En cambio, en los gestos referenciales, no todos los independientes se han clasificado con la doble etiqueta referencial-pragmático, sino solamente aquellos que en esta posición tienen una función diferente a la de transmitir un contenido semántico, como recuperar el turno (interactivo) o expresar acuerdo o desacuerdo (performativo). Asimismo, el desacuerdo sin una forma referencial también aparece en posición independiente.

Las afiliativa es la única función modal en posición independiente y se corresponde con las sonrisas o risas de los oyentes para reforzar la imagen del hablante (§6.2.1.3).

Mientras las funciones mencionadas no suponen una gran cantidad de gestos del corpus, el seguimiento destaca por ser la función más frecuente en posición independiente, con 145 casos; solo 7 casos coverbales no permiten que el 100% sean independientes. Por supuesto, habrá muchos más seguimientos coverbales en la

conversación, pero no que supongan un cambio en la segmentación, que es el objetivo de este estudio.

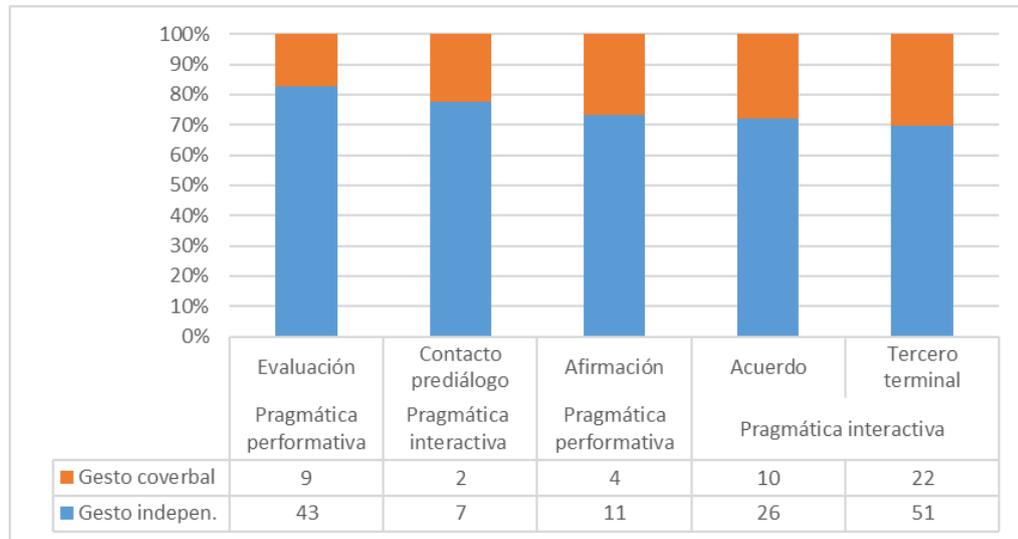


Gráfico 18. Funciones con posición independiente frecuente.

Otras funciones cuyos gestos independientes están en la horquilla del 83%-70% de frecuencia son la evaluación y la afirmación, como performativos, y el contacto prediálogo, el acuerdo y el tercero terminal, como interactivos (Gráfico 18). Tras el seguimiento, estas funciones suman una cantidad considerable de gestos del corpus (185).

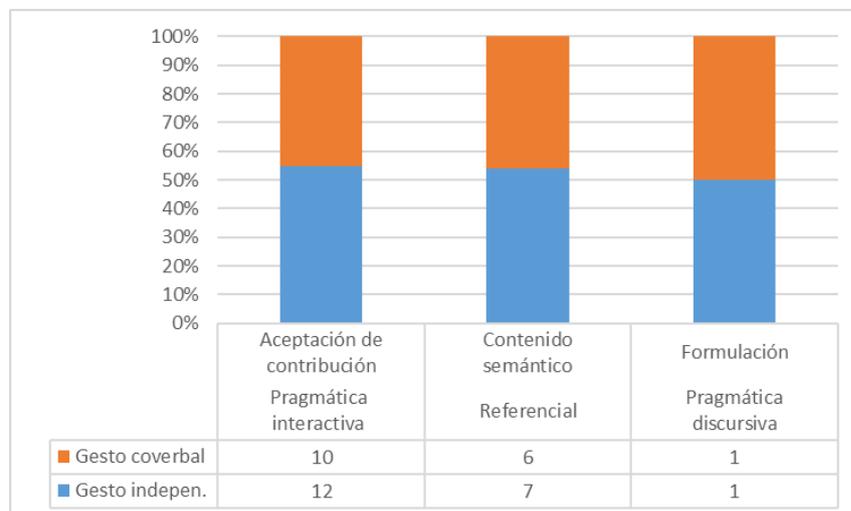


Gráfico 19. Funciones con posición independiente y coverbal igualada.

En unos porcentajes muy igualados entre los gestos independientes y los coverbales (Gráfico 19), la aceptación de contribución es la función interactiva con menor porcentaje de independientes porque también suele aparecer al inicio del turno del hablante para validar la contribución del interlocutor y seguir con el turno planeado. Ya en menor cantidad de casos, la función referencial aparece tanto en posición

independiente como coverbal afectando a la segmentación dialógica. Y con una aparición anecdótica en el total del corpus, la formulación ocupa una posición en cada una de las veces.

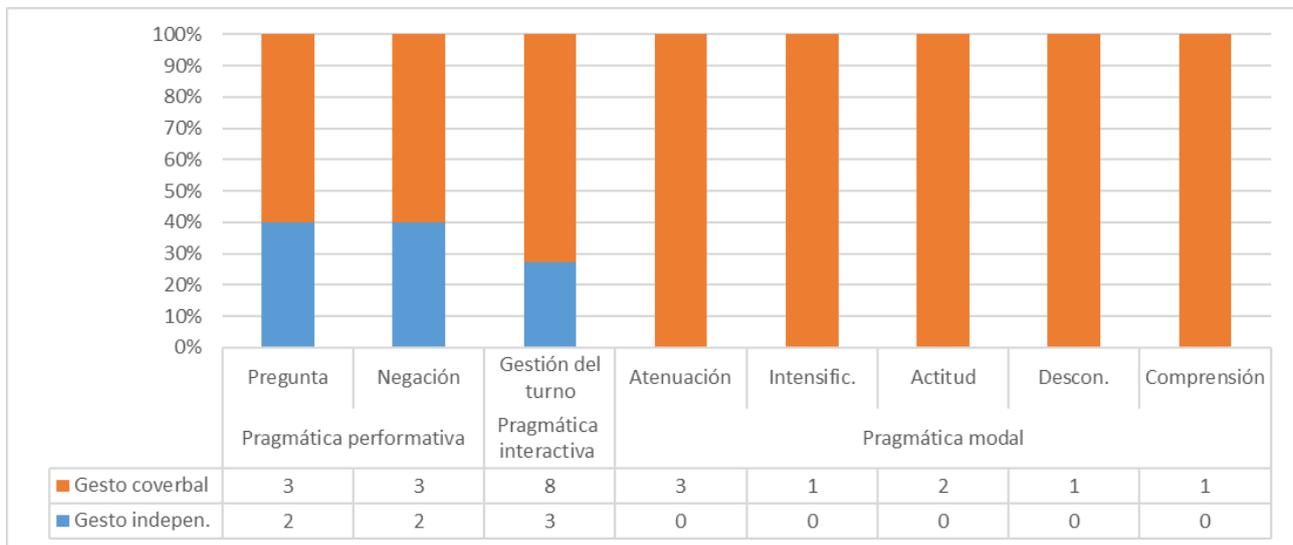


Gráfico 20. Funciones con posición coverbal predominante.

Las funciones que ya ocupan más la posición coverbal que la independiente (Gráfico 20) son la pregunta y la negación (performativos) y la gestión del turno (interactivo). Otras funciones interactivas han tenido más efectos en posición independiente, pero, precisamente, estos gestionan turnos verbales, así que suelen ser coverbales a ellos.

Por último, salvo la afiliación, el resto de las funciones modales, que tan solo son 8 del corpus, siempre son coverbales, pues afectan a la interpretación modal de la intervención verbal.

En cuanto a la segmentación de los gestos independientes como intervenciones, la gran mayoría, 318, son Ir no validadas; los demás tipos de intervención tan solo son ocupados por 37 gestos. Sin embargo, en la evaluación la proporción no es tan grande (21% validadas - 79% no validadas). Las únicas funciones, de uso minoritario en el análisis, que no cumplen esta tendencia son la afirmación, la negación, el contacto prediálogo y el contenido semántico, que en posición independiente suelen ocupar intervenciones validadas (Gráfico 21). La causa puede estar en que estas funciones performativas forman parte de pares adyacentes, así como la función referencial cuando es el único acto de la intervención; el contacto prediálogo, en nuestros casos, suele lograr su objetivo de establecer el inicio del diálogo, así que está validado, no se ignora.

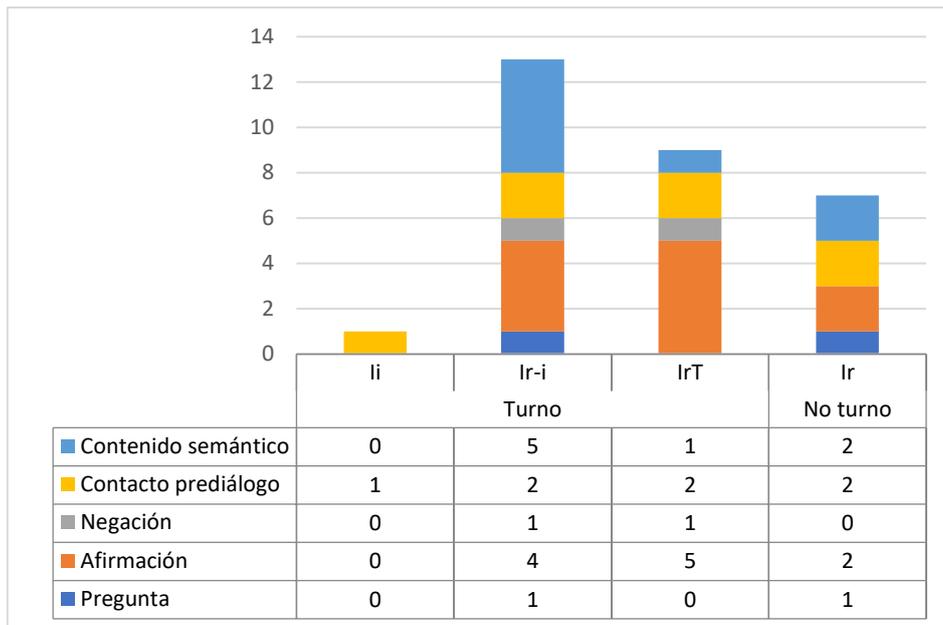


Gráfico 21. Funciones predominantes en intervenciones gestuales validadas.

Una última cuestión sobre la función y la posición es que una intervención gestual no solo puede estar formada por un único acto gestual en posición independiente de intervención, sino que puede tener dos actos gestuales (§5.4.3). Esto ocurre en el corpus en 14 intervenciones gestuales. Todas ellas están formadas por dos gestos, excepto de un caso complejo en el que se combinan 6 gestos.

En 6 de estas intervenciones se combinan funciones interactivas en posición reactiva: seguimiento y acuerdo (2), seguimiento y desacuerdo (1), aceptación de contribución y seguimiento (1), seguimiento y seguimiento (1). Siguiendo la tipología de Cestero (2000b), podemos denominarlos *apoyos compuestos*.

Otras 4 combinaciones en posición Ir no validadas combinan afiliación y evaluación (modal+performativo), seguimiento y evaluación (interactivo+performativo), seguimiento y pregunta que no recibe respuesta (interactivo+ performativo), y acuerdo y contenido semántico (performativo+referencial) que el hablante no llega a formular verbalmente porque se interrumpe para escuchar a otro hablante.

Ya como turnos (3 casos), en posición de Ir-i encontramos tercero terminal y pregunta (interactivo+performativo) y evaluación y comprensión (performativo+performativo), y en Ir desconocimiento y formulación (performativo+discursivo).

El turno que combina 6 gestos se inicia con un gesto interactivo de desacuerdo, continúa con 4 gestos referenciales deícticos que señalan a distintas partes del cuerpo y termina con un último desacuerdo (Ejemplo 57, Imagen 16). B explica que el precio de

su tatuaje está justificado porque se está curando bien, pero C utiliza como ejemplos los tatuajes que tiene en cada parte del cuerpo para argumentar su desacuerdo con B. Realmente, el último gesto interactivo viene acompañado de una risa, por lo que estrictamente deberíamos clasificar los gestos de este turno como coverbales. Pero precisamente en este caso, hay más contenido proposicional en los propios gestos que en la risa. Podríamos decir que la risa es la que acompaña al turno gestual y no al revés.

(57)

1B1: mm (1.6) pero sí es verdad que// no me duele o sea me lo estoy curando ahora y no me [pi]=

A1: [mm]

B: =ca no me es[cuece][e]=

2C1: [Manos¹²⁰]
[CabSacude]

B: =[/ no me está]=

C: [Manos¹²¹--]
[CabOtro¹²²--]
[Sonrisa-----]

B: =[molestando]=

C: [Manos¹²³----]
CabOtro-----
Sonrisa-----

B: =[na]=

C: [Manos¹²⁴]
CabOtro-
Sonrisa--

B: =[da]

C: [Manos¹²⁵]
CabOtro-
Sonrisa--

3B2: ya [[pero no]]=

C: [[(RISAS)]]
[CabSacude]
[Manos¹²⁶--]

¹²⁰ Manos: ManoPlana PalmaArriba HaciaFrente

¹²¹ ManoPlana PalmaArriba HaciaBrazoDch

¹²² Cabeza inclinada ligeramente hacia delante.

¹²³ ManosPlanas PalmasArriba HaciaVientre

¹²⁴ ManosPlanas PalmasArriba HaciaCostillas

¹²⁵ ManosPlanas PalmasArriba HaciaPeriferiaArriba

¹²⁶ PalmasArribaAFuera

Sonrisa---]

B: =sé pero Emy que sí se ha hecho otros sí que ha sido como *los otros tal*

(5C.ManCarCab.2, Conv5. 0:23:50)



Imagen 16. Secuencia de gestos 5C.ManCarCab.2.

6.2.3. Forma y función

En §6.2.1 se ha descrito la forma de algunos gestos o se han mencionado algunas formas frecuentes de cada microfunción. Como partimos de una tipología funcional muy amplia, hay también mucha variedad formal. De algunas funciones hay muy pocos casos (prag. discursivo, prag. modal, referencial, referencial-pragmático), por lo que no podemos afirmar una tendencia en la configuración de los gestos utilizada. Además, los pocos referenciales hacen referencia a objetos físicos o abstractos muy específicos. En todo caso, el gesto modal de afiliación, con 31 unidades, es el único cuyo criterio formal (sonrisa o risa) es clave para determinar la función.

Hay 87 gestos performativos, pero sus formas dependen del tipo de acto que realizan. Hemos visto algunas formas asociadas con la pregunta, la afirmación, la negación y las funciones epistémicas, pero no tenemos suficientes muestras para afirmar una tendencia. La mayoría de los performativos son de evaluación (52), pero también el significado de cada evaluación (sorpresa, incredulidad, hartazgo, etc.) afecta a la forma del gesto.

En cambio, hay 314 gestos interactivos. La distinción entre el seguimiento, el acuerdo, el tercero terminal y la aceptación de contribución, generalmente, ha sido una combinación de la forma y el lugar que el gesto ocupaba en la secuencia, pues la forma a menudo coincidía. Como de estas funciones hay bastantes muestras, podemos observar qué configuración formal específica es más frecuente para cada una (Tabla 40).

Gestos interactivos	Seguimiento	152
	Acuerdo	36
	Desacuerdo	11
	Tercero terminal	73
	Aceptación de contribución	22
	Contacto prediálogo	9
	Gestión del turno	11

Tabla 40. Total de gestos interactivos.

En el resto de los gestos interactivos (desacuerdo, contacto prediálogo, gestión del turno) la forma era más específica para cada función. En consecuencia, tanto para estas como para las macrofunciones mencionadas anteriormente, podemos observar las partes del cuerpo más utilizadas (cabeza, manos, cara, etc.), pero por la variedad de funciones y

la poca cantidad de casos, no creemos que sea útil un análisis de formas específicas, más allá de las comentadas en §6.2.1.

Hay 336 gestos que solamente utilizan una parte del cuerpo (cabeza, cara, hombros o manos), mientras que 109 combinan dos o más partes. Para simplificar el análisis y no pasar de 4 partes del cuerpo a 16 combinaciones (algunas de usos marginales), contabilizamos cada parte del cuerpo como una unidad de movimiento; es decir, un gesto facial se cuenta como un movimiento de cara, pero si en la realización de un gesto se usa la cabeza y la cara, se cuenta como un movimiento de cabeza y un movimiento de cara. Además, como tenemos cantidades de gestos muy dispares para cada función, utilizamos en la Gráfico 22 los porcentajes de frecuencia de cada movimiento en la función¹²⁷.

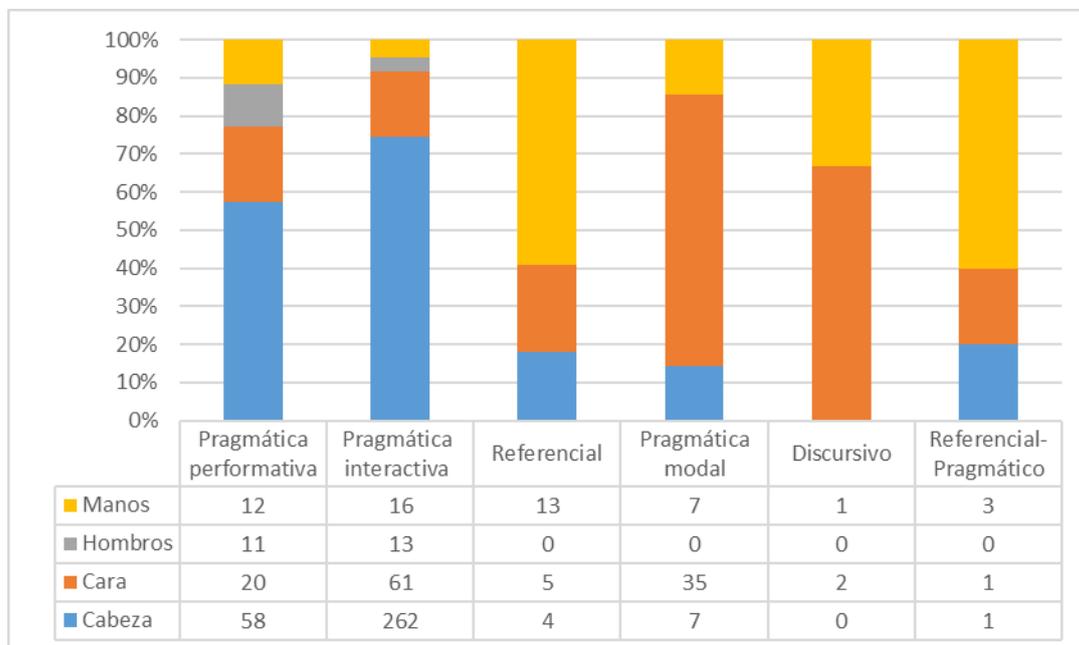


Gráfico 22. Partes del cuerpo utilizadas para cada función.

El resultado más claro es que los gestos performativos e interactivos suelen realizarse con la cabeza, especialmente los interactivos; con menos frecuencia se utiliza la cara, los hombros o las manos, en ese orden. Como son los gestos de los que más muestras contamos, estos son los resultados más fiables.

Teniendo en cuenta que la mayor parte de los gestos modales encontrados eran risas y sonrisas de afiliación, es obvio el 71% de frecuencia de los movimientos faciales.

¹²⁷ Como en Tabacaru (2019), a menudo varios movimientos corporales se combinan. Ella elige contabilizar cada movimiento sin diferenciar si aparecen solos o en combinación con otros. Entonces tiene más formas que gestos.

Los gestos referenciales, ya sean utilizados para expresar contenido semántico o pragmático, suelen realizarse sobre un 60% de las veces con las manos. Este resultado es compatible con la descripción general de los gestos referenciales como gestos manuales deícticos o que representan objetos o acciones concretas o abstractas.

En el Gráfico 23, mostramos las partes del cuerpo de las funciones más frecuentes (seguimiento, acuerdo, tercero terminal y aceptación de contribución).

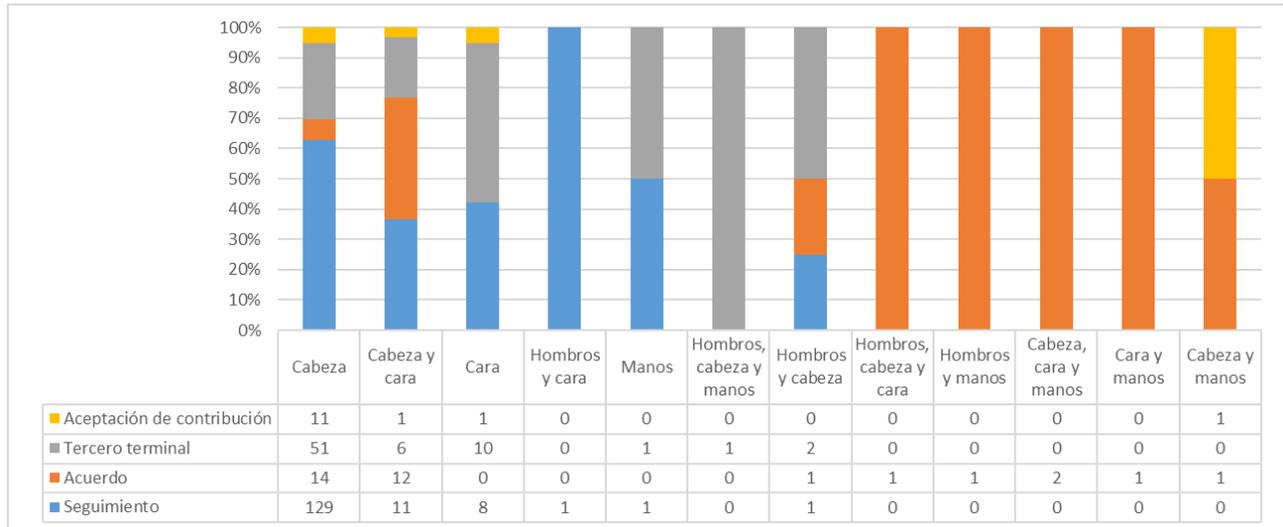


Gráfico 23. Funciones interactivas más frecuentes y partes del cuerpo utilizadas

El movimiento únicamente de cabeza es el más frecuente en todas las funciones, pero especialmente en el seguimiento. Los gestos que combinan cabeza y cara ya son 30, pero hay un aumento de la función de acuerdo. Solamente hay 19 gestos faciales, casi 50% seguimiento y 50% tercero terminal; el acuerdo no se expresa con la cara únicamente. En el resto de las combinaciones, tan solo hay de 1 a 4 casos, así que no podemos generalizar; sin embargo, las combinaciones que incluyen el movimiento de la mano suelen ser tener función de acuerdo; y cuando hay hombros o manos implicadas, no suele ser por seguimiento.

8 de los 10 gestos faciales de tercero terminal son una sonrisa. Se le ha dado esa función por su posición en la secuencia, pero en los casos que no cierran un par adyacente, se han clasificado como función modal de afiliación. Podría discutirse si en estos casos predomina la función afiliativa también o la de cierre de secuencia, o si son multifuncionales.

Puesto que los gestos de cabeza son los más numerosos, en el Gráfico 24 vemos qué movimientos se usan en cada función.

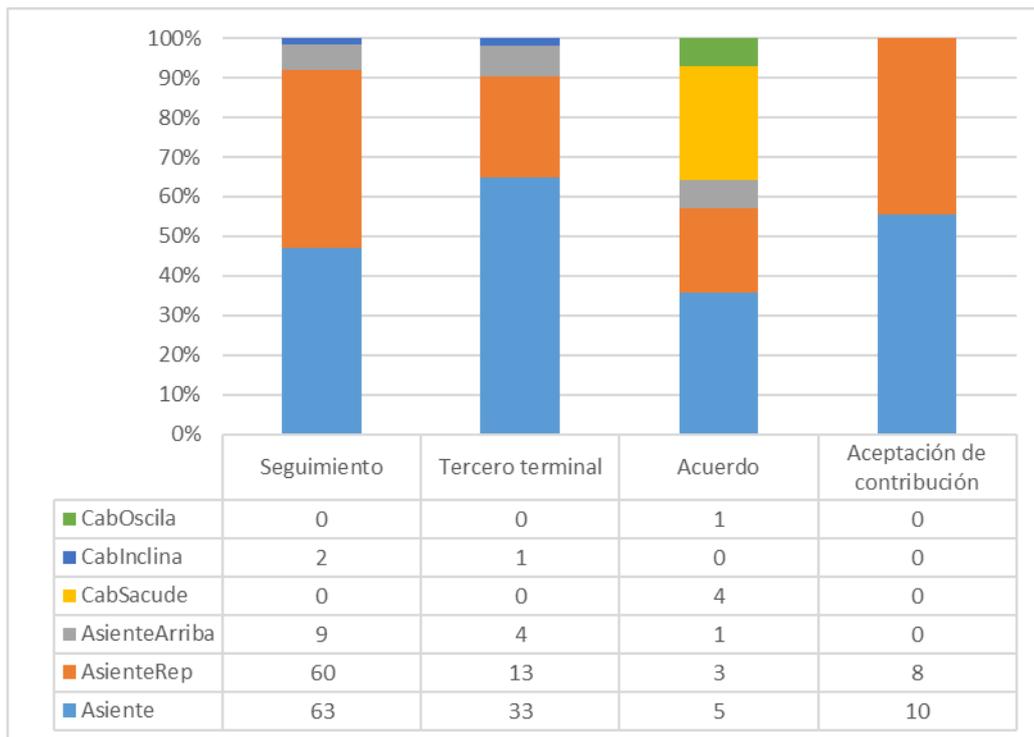


Gráfico 24. Funciones interactivas más frecuentes y movimientos de cabeza.

El asentimiento simple, de una sola cabezada, y el repetido son los movimientos más utilizados generalmente, pero es más notable en el tercero terminal, seguido de la aceptación de contribución y el seguimiento. El acuerdo, que expresa no solo la atención hacia el hablante o que se ha cerrado un par adyacente, sino que el oyente está de acuerdo, parece utilizar más recursos, tanto en el Gráfico 24 con la cabeza, como en el Gráfico 23 con las manos. Es destacable también la sacudida de cabeza para expresar acuerdo, pues no expresa negación, sino que el oyente está de acuerdo con la opinión del hablante formulada como proposición negativa.

En general, la forma de estos gestos interactivos no es determinante para diferenciar qué función interactiva específica cumple el gesto, sino que depende del contexto secuencial. Quizás el asentimiento repetido es menos frecuente en el tercero terminal que en el seguimiento o la aceptación de contribución. Y los casos en los que el acuerdo combina varias partes del cuerpo, como las manos o los hombros, también son más reconocibles.

6.2.4. Posición y forma

Según Kendon (1967), la orientación de la cabeza hacia el interlocutor es más frecuente durante el rol de oyente. Sin embargo, por último, comprobamos si esto sucede así en las intervenciones gestuales (Gráfico 25).

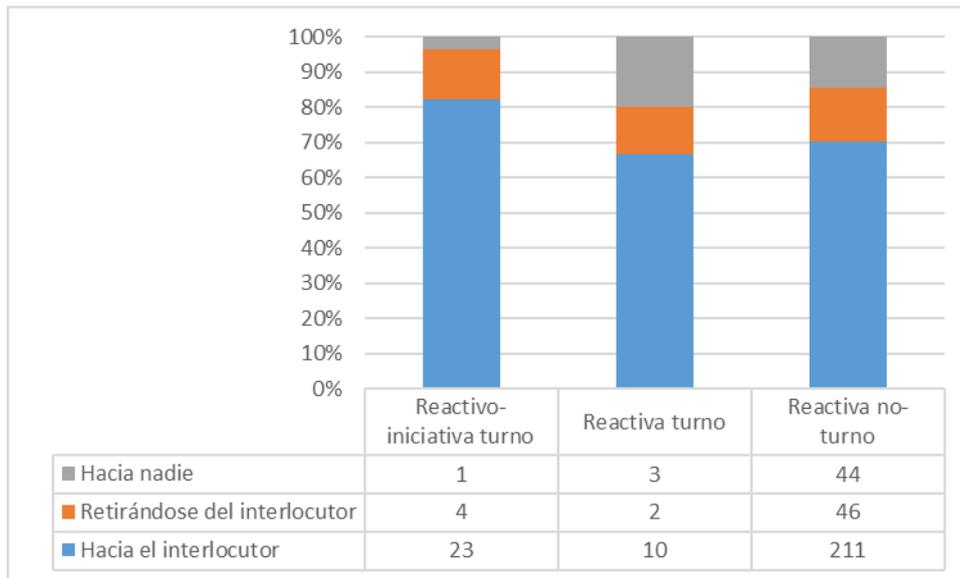


Gráfico 25. Orientación de las intervenciones gestuales.

Como tan solo hay dos casos de intervenciones iniciativas e independientes, no las incluimos en el análisis. Puesto que hay muchas más intervenciones reactivas que de otro tipo, mostramos los resultados con porcentajes.

En general, en todos los tipos de intervención la orientación suele dirigirse hacia el interlocutor. No podemos afirmar que haya una diferencia notable entre las intervenciones que son turno y las que no están validadas como turno. Pero sí hay una distancia de un 10% aproximadamente entre las reactivas (que reaccionan al hablante, pero no reciben reacción posterior) y las reactivo-iniciativas (que sí reciben una reacción posterior). Quizás se debe a la función del gesto (un 60% de las Ir-i orientadas al interlocutor son performativas), que en combinación con la mirada favorece la reacción; pero no podemos asegurarlo.

6.2.5. Resumen

En este segundo análisis se ha podido comprobar que en los cambios producidos en la segmentación dialógica los gestos predominantes son los gestos independientes con función interactiva que ocupan una intervención reactiva. Se cumple la hipótesis de que la función interactiva es la más frecuente tanto en gestos independientes como en

coverbales. A partir de los casos encontrados en el corpus, hemos desarrollado una serie de funciones interactivas que afectan a la segmentación: el seguimiento, el tercero terminal, el acuerdo y el desacuerdo, la aceptación de la contribución del oyente, la gestión del turno y el contacto prediálogo. Seguimiento, tercero terminal, acuerdo y aceptación de contribución son los más frecuentes, especialmente el primero, y su distinción depende especialmente del contexto secuencial, pues coinciden con frecuencia en su forma: el asentimiento de cabeza es su forma predominante.

Fuera de la función interactiva, la evaluación (performativa) y la afiliación (modal) son funciones bastantes comunes, ya que también son propias de la intervención reactiva del oyente. En cambio, hay menos de 15 casos cada una de las demás funciones (performativas, modales, referenciales o discursivas), así que no podemos validar una posición, un tipo de intervención o una forma que sean prototípicas de este tipo de gestos, pero podemos plantear hipótesis para futuros estudios multimodales de la conversación.

Las funciones performativas (evaluación, pregunta, afirmación, negación, función epistémica) aparecen generalmente en posición independiente, aunque la pregunta y la negación tienden más a la coverbal. Además, la intervención gestual formada por una pregunta, una afirmación o una negación, suele ser turno (en contraste con el resto de los gestos del corpus). Aunque más de la mitad de los casos se realizan con la cabeza, es la macrofunción en las que las cuatro partes del cuerpo (cabeza, cara, hombros, manos) están más repartidas.

En funciones modales hay una gran diferencia entre la afiliación, que siempre la encontramos en posición independiente, y el resto (atenuación, intensificación, actitud, comprensión, desconocimiento), que son todas coverbales. Respecto a la forma, la mayoría se realizan con la cara y son de afiliación; entre el resto de las funciones modales también se utiliza en el mismo grado la cabeza y las manos.

Las funciones referenciales se reparten de igual manera entre gestos independientes y coverbales cuando simplemente transmiten un contenido semántico, pero la balanza se inclina hacia los independientes al incluir los gestos referenciales que en esa posición adoptan una función pragmática. En esta función es donde encontramos más gesticulación de manos, como era de esperar.

Apenas hay dos casos de funciones discursivas de formulación, una independiente y otra coverbal, con la boca y otro con las manos. En consecuencia, no nos atrevemos a lanzar ninguna hipótesis.

CAPÍTULO 7. CONCLUSIONES

En este estudio hemos abordado la interacción multimodal de la conversación desde un enfoque pragmático estructural que no se había utilizado hasta ahora para este campo. El modelo Val.Es.Co. (Briz y Grupo Val.Es.Co. 2003, Pons Bordería 2022) presenta un sistema jerárquico en el que unidades de nivel superior están formadas por unidades de nivel inferior, de manera que esta estructura de la conversación permite sistematizar los fenómenos lingüísticos, pragmáticos y sociales que aparecen en ella. Esto ha permitido incorporar los gestos —es decir, los movimientos kinésicos de los interlocutores con una intencionalidad y una función en la interacción— al nivel dialógico de la estructura, que comprende la unidad *intervención* y su relación pragmática con otras intervenciones. Aplicando la teoría de posiciones (Briz y Pons Bordería 2010, Pons et al. en prensa) a los gestos, hemos diferenciado entre gestos independientes, que no acompañan a una intervención verbal, sino que por sí mismos son el único acto (o conjunto de actos) que forma una intervención gestual, y los gestos coverbales, que forman parte de una intervención verbal a modo de actos o subactos. En cuanto a la dimensión social del nivel dialógico —que atiende al intercambio de papeles sociales entre hablante y oyente que permite el dinamismo y el desarrollo de la conversación—, hemos relacionado los gestos con la validación de la intervención como unidad *turno*.

Además de los propios resultados del análisis, este estudio también supone una aportación a nivel metodológico, pues nuestra metodología para la recopilación de un corpus de conversaciones reales en vídeo y para su anotación puede servir de pauta para futuros estudios multimodales que se interesen en este género, en este nivel estructural o en estos gestos.

Hemos creado un corpus de conversaciones en español en vídeo de 131 minutos (alrededor de 28.300 palabras) preservando las características de la conversación coloquial. La grabación de manera informada pero secreta ha sido complicada y ha requerido finalmente de la presencia de la investigadora por la dificultad de entrenar a otra persona para seguir el protocolo; en consecuencia, comprendemos la decisión de otros corpus de grabar a informantes conscientes de la grabación. Pero, al menos para este estudio, hemos podido analizar los datos con la seguridad de que eran los más auténticos que podíamos conseguir, descartando los gestos del gancho. En futuros estudios se podrían contrastar nuestros resultados con los de grabaciones sin una metodología de grabación secreta. Asimismo, un corpus más amplio podría incluir un análisis sociolingüístico de los gestos que cada tipo de hablante realiza.

También continuamos el camino iniciado por Val.Es.Co. para llevar al soporte informático la jerarquía de unidades de la conversación (Pons Bordería 2022), ampliando el contenido verbal de las unidades al contenido gestual mediante una metodología de anotación que vincula la descripción del gesto con las unidades conversacionales. Asimismo, nuestra anotación completa la transcripción de Val.Es.Co. en papel (Briz y Grupo Val.Es.Co. 2002a) integrando los gestos: la codificación de los gestos es lo suficiente inteligible y económica para que no requiera un esfuerzo del lector y su alineación con las intervenciones verbales permite identificar los gestos coverbales y los independientes. Es un sistema apto para los gestos que afectan al nivel dialógico, pues se identifica si un gesto forma por sí mismo una intervención gestual y la codificación abarca las formas más comunes de gestos coverbales que afectan a la relación entre intervenciones. No obstante, habría que desarrollar de otra manera el sistema de anotación si se tuviera interés en gestos que generalmente no afectan al nivel dialógico, como los referenciales y los discursivos, que son muy abundantes y suelen utilizar configuraciones de la mano muy variadas.

El objetivo del estudio era comprobar que la multimodalidad podía causar un cambio sustancial en la segmentación de la conversación mediante las unidades dialógicas de Val.Es.Co., así como describir el tipo de cambios y sus causas. La comparación entre el audio y el vídeo de las conversaciones segmentadas mediante las unidades conversacionales dialógicas, efectivamente, ha mostrado que se produce alrededor de un 22% de cambios, que suponen un valor-p de la X^2 bajo ($<0,0001$). Estos cambios han sido nuevas intervenciones (67%) y modificaciones de las ya existentes (33%).

La causa que ha aunado más cambios entre una segmentación y otra son los gestos independientes. Por tanto, el elemento más destacable del análisis es esta intervención gestual que, además, suele ser una intervención reactiva con una función interactiva con la que el oyente muestra que está atendiendo al hablante. Sin embargo, estos últimos son 145 gestos (seguimiento en Ir) de un total de 459 gestos que modifican la segmentación y de un total de 803 cambios producidos por diversas causas visuales —que incluyen gestos, pero también lectura de labios, orientación corporal durante la intervención verbal y acciones extralingüísticas—. Por consiguiente, no se debe centrar la segmentación dialógica multimodal de la conversación en este tipo de gesto, sino también en otros elementos, posiciones y funciones.

Los gestos independientes suponen un total de 360 intervenciones nuevas (de un corpus de 3561 intervenciones) y aunque su aparición afecta a la creación o modificación de intervenciones colindantes, no en una proporción alta (18%), posiblemente porque suelen ser intervenciones reactivas que reaccionan a intervenciones que ya estaban validadas por medios orales. La lectura de labios es la segunda causa de aparición de nuevas intervenciones, y los gestos coverbales, la tercera. En cambio, la modificación de intervenciones existentes en la segmentación oral se debe especialmente a la orientación corporal de los interlocutores. La acción extralingüística es una causa minoritaria que apenas incide en la segmentación.

Socialmente, las intervenciones validadas o turnos se han incrementado un 10,5% en la segmentación visual, por nuevos turno o modificaciones de las intervenciones en turnos; pero hay más nuevas intervenciones no validas que validas (419-120) —311 son gestos independientes reactivos— y más intervenciones que modifican su tipología sin cambiar la validación o cambiando de turno a intervención (148) que intervenciones que cambian a turno (116).

Respecto a las características de los 453 gestos que cambian la segmentación dialógica, el 79% son independientes y el 21%, coverbales. Esto no significa que en una conversación haya más gestos que no acompañen a una intervención verbal que los que sí acompañen, sino que los gestos coverbales suelen afectar a cuestiones internas de la interpretación del turno y no a su relación con otros turnos en el intercambio. En este sentido, son más los gestos independientes que afectan al nivel dialógico, sobre todo, porque ellos mismos son nuevas intervenciones. Cumpliendo nuestra hipótesis, la mayoría de los gestos independientes cumple funciones típicas de las Ir del oyente (seguimiento, evaluación) y la minoría, funciones que aparecen en pares adyacentes.

Hemos clasificado funcionalmente los gestos en referenciales, pragmáticos performativos, pragmáticos interactivos, pragmáticos modales y pragmáticos discursivos, inspirándonos en la clasificación de Kendon (2004) y Müller (1998, 2013, en prep.) y añadiendo la función interactiva (Ekman y Friesen 1969, Bavelas et al. 1992, Streeck 2005, Wehling 2017). Como resultado, hemos encontrado que la mayoría son pragmáticos (96%), específicamente interactivos (68%) y performativos (19%). Es un resultado esperable, pues los gestos referenciales, modales y discursivos suelen actuar como coverbales en el nivel interno del turno; aun así, hay un 8% de modales y un 4% de referenciales que sí afecta al nivel dialógico.

Las funciones interactivas se han relacionado, en su mayoría, con el seguimiento o el acuerdo del oyente al turno de hablante, a la aceptación de una contribución al turno propio o ajeno y al cierre de un par adyacente. Además, estas funciones han coincidido bastante en sus variantes formales, siendo la predominante el asentimiento de cabeza; por lo que el criterio principal para diferenciarlas es su lugar en la secuencia. Otras funciones interactivas menos numerosas han sido la gestión del turno, el contacto prediálogo y el desacuerdo, con formas más específicas para cada una.

Salvo los gestos evaluativos, que también suelen ser intervenciones reactivas como el resto de las funciones, los gestos performativos (pregunta, afirmación, negación) y referenciales, a pesar de su baja incidencia, han destacado por ocupar una intervención validada como turno cuando estaban en posición independiente.

En general, la posición independiente o coverbal del gesto no es determinante para identificar las funciones porque encontramos que todas pueden ocupar ambas posiciones y que en todas suele haber más independientes que modifiquen la segmentación que coverbales. La única función en la que parece haber diferencias más notables entre la posición independiente y la coverbal es la modal. Los usos modales (atenuación, intensificación, actitud, comprensión, desconocimiento) se han vinculado exclusivamente a la posición coverbal, pues, cuando han ocupado la independiente, se han clasificado como performativos. No han sido muchos casos, pero se ha hecho así porque el desconocimiento, por ejemplo, puede dar una interpretación epistémica de incertidumbre del hablante a lo que está diciendo verbalmente, pero en posición independiente resultaban ser respuestas a preguntas que podrían equivaler a “no lo sé”, por lo que consideramos que por sí mismos podían ser actos performativos. La única función clasificada como modal en posición independiente ha sido la afiliación, pues consideramos ese valor más importante en la secuencia que el hecho de que fuera un acto. De la misma manera, el resto de las funciones en posición independiente no se han clasificado automáticamente como performativas.

Respecto a la forma de los gestos, esta está muy vinculada al significado particular de cada función y de muchas de ellas no tenemos muestras suficientes como para formular unas formas prototípicas. Simplemente, notificamos la mayor incidencia de los movimientos de cabeza y faciales para los gestos interactivos y performativos, que son los que más modifican la segmentación dialógica, y confirmamos la mayor frecuencia de los gestos manuales en los referenciales. De los gestos interactivos sí tenemos más

muestras, especialmente de seguimiento, pero el análisis ha mostrado que la forma no es determinante para distinguir entre las funciones interactivas. Por tanto, en los gestos de nuestro corpus, la forma es un criterio que se combina necesariamente con el contexto secuencial para identificar la función, pero por sí misma no es criterio suficiente porque no tenemos suficientes muestras o, precisamente, los gestos interactivos dependen más del criterio secuencial.

Aunque no haya suficientes muestras de ciertas funciones o formas, los resultados de este estudio exploratorio nos ayudan a formular hipótesis para futuros estudios. En este estudio nos hemos centrado en las diferencias entre la segmentación oral y la visual, pero sería interesante observar cómo estos gestos se combinan con la información verbal en la segmentación. Por ejemplo, hemos encontrado un gran número de gestos independientes que funcionan como intervenciones reactivas, pero no todas validan otras intervenciones verbales, sino que refuerzan la validación de intervenciones ya validas verbalmente; entonces, se podría observar la validación de turnos mediante intervenciones reactivas tanto verbales como gestuales, si unos validadores son más frecuentes que otros o si suelen combinarse. La función interactiva de los terceros terminales también ha sido bastante numerosa en el corpus, por lo que también sería interesante contrastar el uso de su forma verbal y gestual.

Dentro de las funciones performativas han aparecido bastantes preguntas a las que se ha respondido con una afirmación gestual (asentimiento). Dos hipótesis pueden explicar esto: 1) una preferencia por el gesto para responder a las preguntas tipo confirmatoria, 2) una preferencia por el gesto para responder únicamente “sí”, pero las preguntas absolutas son normalmente confirmatorias. Se podría contrastar estas respuestas gestuales con las respuestas verbales a preguntas absolutas, si aparecen y qué tipo de preguntas son.

Quienes estén interesados en las funciones performativas, referenciales, modales o discursivas, deberían también incluir gestos que no afectan exclusivamente a la segmentación dialógica, sino a cualquier nivel de la estructura. Las cuestiones formales sobre la identificación de estas funciones se solucionarían con una muestra más grande de estas.

En definitiva, este trabajo ha aportado al estudio de los gestos un nuevo marco teórico y metodológico basado en la pragmática-estructural. Hasta ahora no se había

cuantificado la aparición de intervenciones gestuales, así como los efectos de la información multimodal en la interpretación de las relaciones estructurales y sociales entre intervenciones. Hemos conseguido sistematizar este análisis gestual gracias a la creación de una anotación desde la que se organiza esta información multimodal en la conversación. Además, se ha aportado un corpus de estudio con conversaciones reales que avala los datos obtenidos en este trabajo. En este sentido, esta tesis tiene un enfoque doble, dado que perfila teóricamente un modelo de análisis y; al mismo tiempo, añade una metodología novedosa con la que establecer nuevas investigaciones en el futuro.

Generalmente, los estudios sobre gestos se han centrado en una función, en un esquema semántico o en una forma muy concretas. Esta forma de proceder permite un profundo conocimiento de esos gestos; no obstante, nuestro estudio resulta de interés, ya que permite ampliar el foco en todos los elementos visuales que afectan al nivel dialógico, a todas las funciones del gesto que pueden encontrarse en una posición y a varias partes del cuerpo. Ambos enfoques, el amplio y el estrecho, se nutren mutuamente. Por un lado, el amplio permite una visión de conjunto que da cuenta de los patrones entre gestos de distintas características, pero no puede profundizar en todas las posibilidades de realización de cada gesto. Por otro lado, el estrecho sí permite el análisis de un solo gesto mediante muchas muestras que demuestren un patrón, pero pierde las relaciones que puede tener con otros gestos más distantes. Como se puede comprobar, ambas perspectivas de análisis son necesarias, dado que los resultados de una ayudan en las hipótesis y los argumentos de la otra. Una muestra de ello es que los resultados de este estudio nos han permitido proponer líneas de investigación de funciones específicas para el futuro.

CHAPTER 7. CONCLUSIONS

In this study, the multimodal interaction of conversation is addressed from a pragmatic structural approach that had not been used before in this field. The Val.Es.Co. model (Briz and Val.Es.Co. Group 2003; Pons Bordería 2022) presents a hierarchical system in which higher-level units are formed of lower-level units. So this conversational structure allows for the systematization of linguistic, pragmatic, and social phenomena that appear in it. This has allowed the incorporation of gestures—that is, the kinesic movements with intentionality and a function in the interaction—at the dialogical level of the structure, which encompasses the unit called intervention and its pragmatic relation to other interventions. By applying position theory (Briz and Pons Bordería 2010, Pons et al. in press) to gestures, independent gestures have been distinguished from coverbal gestures. Independent gestures do not accompany verbal interventions but they are the sole act (or set of acts) that forms a gestural intervention, meanwhile coverbal gestures are part of a verbal intervention as acts or subacts. Regarding the social dimension of the dialogical level—which attends to the exchange of social roles between speaker and listener and allows for the dynamism and development of the conversation— gestures have been related to the validation of the intervention as a turn unit.

In addition to the results of the analysis, this study also represents a methodological contribution, as our methodology for collecting a corpus of real video conversations and annotating them can serve as a guideline for future multimodal studies focused on in this genre, this structural level, or these gestures.

We have created a corpus of Spanish video conversations totaling 131 minutes (around 28,300 words), preserving the characteristics of informal conversation. The informed but secret recording was challenging. Ultimately, it required the presence of the researcher due to the difficulty of training another person to follow the protocol. Therefore, the decision of other corpora to record informants who are aware of the recording is understandable. However, at least for this study, the analysis is done with the assurance that the data is the most authentic that could be obtained (excluding gestures from the confederate). In future studies, our results could be compared with recordings without a secret recording methodology. Furthermore, a larger corpus could include a sociolinguistic analysis of the gestures performed by each kind of speaker.

This work continues the path initiated by Val.Es.Co. to make this hierarchy of conversation units available digitally (Pons Bordería 2022). The verbal content of the units is expanded to include gestural content through an annotation methodology that

links gesture descriptions to conversational units. Likewise, the annotation complements the paper transcription of Val.Es.Co. (Briz and Val.Es.Co. Group 2002) by integrating gestures. The encoding of gestures is sufficiently intelligible and economical, so it does not require an effort from the reader. The gesture alignment with verbal interventions allows for the identification of coverbal and independent gestures. It is a suitable system for gestures that affect the dialogical level: on the one hand, it identifies whether a gesture in itself forms a gestural intervention; on the other hand, the coding encompasses the most common forms of coverbal gestures that affect the relationship between interventions. However, the annotation system would need to be developed differently if there was interest in gestures that generally do not affect the dialogical level, such as referential and discursive gestures, because they are very abundant and often use a wide range of hand configurations.

The objective of the study was to verify that multimodality could cause a substantial change in the segmentation of conversation through the dialogical units of Val.Es.Co., as well as to describe the type of changes and their causes. The comparison between audio and video recordings of conversations, segmented using dialogical conversational units, shows indeed that approximately 22% of changes yield a low p-value of the X2 test (<0.0001). These changes consist of new interventions (67%) and modifications of existing ones (33%).

The factor that contributes to the most changes between one segmentation and another is independent gestures. Therefore, the most noteworthy element in the analysis is this gestural intervention. It also tends to be a reactive intervention with an interactive function through which the listener shows attentiveness to the speaker. However, these latter gestures are 145 (with “following” function) out of a total of 459 gestures that modify the segmentation. And they are out of a total of 803 changes caused by various visual factors (including gestures, but also lip reading, body orientation during verbal intervention, and extralinguistic actions). Therefore, the multimodal dialogical segmentation of conversation should not only focus on this type of gesture but also consider other elements, positions, and functions.

Independent gestures account for a total of 360 new interventions (out of a corpus of 3561 interventions). Although their occurrence affects the creation or modification of neighboring interventions, it does not do so to a high proportion (18%). Possibly, they are reactive interventions that react to interventions that have already been validated through

oral means. Lip reading is the second cause of new intervention appearance, followed by co-verbal gestures as the third cause. On the other hand, the modification of existing interventions in oral segmentation is primarily due to the body orientation of the interlocutors. Extralinguistic actions are a minor cause that barely impacts segmentation.

From a social perspective, validated interventions or turns increased by 10.5% in visual segmentation, due to new turns or modifications of interventions within turns. However, there are more new invalid interventions than valid ones (419-120), with 311 of them being reactive independent gestures. Additionally, there are more interventions that modify their typology without changing validation or transitioning to turns (148) than interventions that change into turns (116).

Regarding the characteristics of the 453 gestures that change dialogic segmentation, 79% are independent gestures and 21% are co-verbal gestures. This does not mean that in a conversation there are more gestures that do not accompany verbal interventions than those that do. Rather, co-verbal gestures tend to affect internal aspects of turn interpretation rather than their relationship with other turns in the exchange. In this sense, independent gestures have a greater impact on the dialogic level, primarily because they themselves are new interventions. Consistent with the hypothesis proposed, the majority of independent gestures serve typical functions of listener response (such as “following” and “evaluation”), while the minority fulfill functions that appear in adjacent pairs.

Gestures are functionally classified into referential, performative pragmatic, interactive pragmatic, modal pragmatic, and discursive pragmatic. This classification is inspired by Kendon’s (2004) and Müller’s (1998, 2013, in prep.), adding the interactive function (Ekman and Friesen 1969, Bavelas et al. 1992, Streeck 2005, Wehling 2017). As a result, we find that the majority of gestures are pragmatic (96%), more specifically interactive (68%) and performative (19%). This is an expected result, as referential, modal, and discursive gestures often act as co-verbals at the internal level of the turn. However, 8% of modal gestures and 4% of referential gestures do impact the dialogic level.

The interactive functions are mostly related to the listener's actions as following, agreement, acceptance of a contribution to a turn, and the closure of an adjacent pair. Furthermore, these functions show considerable overlap in their formal variants, with

head nodding being the predominant form. Therefore, the main criterion for differentiating them is their position in the sequence. Other less frequent interactive functions include turn management, pre-dialogue contact, and disagreement, each with more specific forms.

Except for evaluative gestures—which also tend to be reactive interventions like the rest of the functions—, performative gestures (question, affirmation, denial) and referential gestures, despite their low incidence, stand out for occupying a validated intervention as a turn when they are in an independent position.

In general, the independent or co-verbal position of the gesture is not determinative for identifying the functions because it is found that all functions can occupy both positions and that there are usually more independent gestures that modify segmentation than co-verbal ones. The only function where there appear to be more notable differences between the independent and co-verbal positions is the modal function. The remaining modal uses (mitigation, intensification, attitude, understanding, unfamiliarity) are exclusively linked to the co-verbal position. When they occupy the independent position, they are classified as performative gestures (although there are not many cases). For example, a “lack of knowledge” can give an epistemic interpretation of uncertainty to what the speaker is verbally expressing; but in an independent position, it turned out to be a response to a question that could be equivalent to “I don't know”. So this kind of gesture can be considered a performative act in its own right. The only function classified as modal in an independent position is “affiliation” because that value was more important in the sequence than the fact that it is an act. Similarly, for the rest of the functions in independent position, they are not directly classified as performative.

Regarding the form of the gestures, it is closely linked to the particular meaning of each function. But, for many of them, there are not sufficient samples to formulate prototypical forms. The higher incidence of head and facial movements is noted for interactive and performative gestures, which are the ones that most modify dialogic segmentation. And the greater frequency of manual gestures in referential gestures is confirmed. There are more samples for interactive gestures, especially for “following”, but the analysis shows that form is not determinative in distinguishing between interactive functions. Therefore, in the gestures of this corpus, form is a criterion that necessarily combines with the sequential context to identify the function. However, on its own, form

is not a sufficient criterion because there are not enough samples or, maybe, because interactive gestures depend more on the sequential criterion.

Although there are not enough samples of certain functions or forms, the results of this exploratory study help to formulate hypotheses for future studies. In this study, we focus on the differences between oral and visual segmentation, but it would be interesting to observe how these gestures combine with verbal information in segmentation. For example, a large number of independent gestures that function as reactive interventions are found, but not all of them validate other verbal interventions. Instead, they reinforce the validation of already verbally validated interventions. Therefore, it would be possible to observe the validation of turns through both verbal and gestural reactive interventions. Some validators are more frequent than others or they tend to combine. The interactive function of third parties is also quite numerous in the corpus, so it would also be interesting to contrast the use of their verbal and gestural forms.

Within the performative functions, several questions appear to be answered with a gestural affirmation (nodding). Two hypotheses can explain this: 1) a preference for the gesture to respond to confirmatory-type questions, 2) a preference for the gesture to respond "yes", but absolute questions are usually confirmatory. These gestural responses could be contrasted with verbal responses to absolute questions. If they occur and what type of questions they are responding to.

Those interested in performative, referential, modal, or discursive functions should also include gestures that do not exclusively affect dialogic segmentation but any level of structure. Formal issues regarding the identification of these functions would be resolved with a larger sample of these gestures.

In conclusion, this work contributes to the study of gestures by providing a new theoretical and methodological framework based on pragmatic-structural analysis. Until now, the occurrence of gestural interventions and the effects of multimodal information on the interpretation of structural and social relationships between interventions had not been quantified. We systematize this gestural analysis successfully through the creation of annotation that organizes multimodal information in conversation. Additionally, a study corpus consisting of real conversations has been developed, which supports the data obtained in this work. Thus, this dissertation has a dual focus: it provides a theoretical analysis model, and simultaneously introduces a novel methodology for future research.

Generally, studies on gestures have focused on a specific function, a semantic schema, or very specific forms. This approach allows for a deep understanding of those particular gestures. However, our study is of interest as it expands the focus to include all visual elements that affect the dialogic level, encompassing all functions of gestures that can be found in various positions and which involve different body parts. Both the broad and the narrow approaches complement each other. On one hand, the broad perspective provides an overview that reveals patterns between gestures with different characteristics, but may not delve into all the possibilities of each gesture's realization. On the other hand, the narrow perspective allows for the analysis of a single gesture through multiple samples that demonstrate a pattern, but it may overlook the relationships it may have with other more distant gestures. As can be seen, both analytical perspectives are necessary, as the results from one inform the hypotheses and arguments of the other. An example of this is that the findings of this study have enabled us to propose future research directions for specific functions.

Bibliografía

- Abercrombie, D. (1968). Paralanguage. *British Journal of Disorders of Communication*, 3(1), 55-59. <https://doi.org/10.3109/13682826809011441>
- Adolphs, R., Damasio, H., Tranel, D. y Damasio A. R. (1996). Cortical systems for the recognition of emotion in facial expression. *The Journal in Neuroscience*, 16(23): 7678-7687.
- Aiko, H. (2011). Some Functions of Speaker Head Nods, en J. Streeck, C. Goodwin, C. LeBaron (Eds.), *Embodied Interaction: Language and Body in the Material World* (93-105), Cambridge University Press.
- Alam, F., Rosemberg, C. R. y Scheuer, N. (2020). Gestos y habla en la construcción infantil de narrativas entre pares. *Cuadernos del Centro de Estudios en Diseño y Comunicación. Ensayos* 89, 187-212.
- Albelda, M. y Estellés, M. (coords.): *Corpus Ameresco*, Universitat de València, ISSN: 2659-8337, www.corpusameresco.com.
- Alcaraz Carrión, D. y Valenzuela, J. (2019). Distant time, distant gesture: speech and gesture correlate to express temporal distance. *Semiotica*, 241, 159-183.
- Alibali, M. W., Heath, D. C. y 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, 169–188. doi: 10.1006/jmla.2000.2752
- Allwood, J. (2001). Dialog Coding – Function and Grammar: Göteborg Coding Schemas. *Gothenburg Papers in Theoretical Linguistics*, 85. Dept. of Linguistics, University of Göteborg, 1-67.
- Allwood, J. (2008). Multimodal Corpora, en A. Lüdeling y M. Kytö. (Eds.), *Corpus Linguistics. An International Handbook* (207-225). Berlin: Mouton de Gruyter.
- Allwood, J., Björnberg, M., Grönqvist, L., Ahlsén, E., Ottesjö, C. (2000). The Spoken Language Corpus at the Dept of Linguistics, Göteborg University. *Forum Qualitative Social Research*, 1(3), 22.
- Allwood, J. y Cerrato L. (2003). A Study of Gestural Feedback Expressions. *First Nordic Symposium on Multimodal Communication*, Copenhagen.
- Allwood, J., Cerrato, L., Dybkjaer, Jokinen, K., Navaretta, C., y Paggio, P. (2005). The MUMIN Multimodal Coding Scheme, en *NorFa Yearbook 2005*.
- Allwood, J., Cerrato, L., Jokinen, K., Navarretta, C. y Paggio, P. (2007). The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena, en J.C. Martin, P. Paggio, P. Kuehnlein, R. Stiefelwagen, F. Pianesi (Eds.), *Multimodal corpora for modelling human multimodal behaviour*, número especial de International Journal of Language Resources and Evaluation (41) (273-287). Berlin: Springer.
- Alturo, N., Clemente, I., Payrató, Ll. (2016): Notes for a multilingual and multimodal functional discourse grammar, en M. Fernández-Villanueva, K. Jungbluth (Ed.): *Beyond Language Boundaries. Multilingual Use in Multilingual Contexts* (3-33). Berlin: De Gruyter Mouton.

- Alvarado Ortega, M. B. (2012). Una propuesta de estudio para el humor en la conversación coloquial, *ELUA*, 26, 7-28.
- Alvarado Ortega, M. B. (2021). El humor y las risas en la conversación coloquial: estudio pragmático de la identidad de género. *ORALIA*, 24/1, 7-26.
- Alvarado Ortega, M.B., Cortés Rodríguez, L. (2017) La risa en los debates en torno al estado de la nación, *LEA*, 39(1), 5-30.
- Alvarado Cantero, L. (2016). El gesto señalar con los labios: hacia una integración de lo no verbal en el estudio del español de los Estados Unidos, *Glosas* 8/9, 5-25.
- André-Larochebouvy, D. (1984). *La conversation quotidienne*, Paris, Didier Crédif.
- Andrén, M. (2010). *Children's gestures from 18 to 30 months*. Tesis doctoral, Centre for Languages and Literature, Lund University.
- Anscombe, J. C. y Ducrot, O. (1983). *L'argumentation dans le langage*, Bruselas, Mardaga.
- Argyle, M. (1988). *Bodily Communication* (2ª ed.). Routledge.
- J. M. Atkinson and J. Heritage (Eds.) (1984), *Structures of social action: studies in conversation analysis*. Cambridge: Cambridge University Press and Editions de la Maison des Sciences de l'Homme.
- Auer, P. (2021). Turn-allocation and gaze: A multimodal revision of the “current-speaker-selects-next” rule of the turn-taking system of conversation analysis. *Discourse Studies*, 23(2), 117–140. <https://doi.org/10.1177/1461445620966922>
- Austin, G. (1966[1806]). *Chironomia, or a treatise on rhetorical delivery*. M.M. Robb y L. Thonssen (eds.), Carbondale & Edwardsville: Southern Illinois University Press.
- Austin, J. L. (1962). *How to do things with words*. Oxford: Oxford University Press.
- Badia Climent, S. (2018). La conversación coloquial y los videojuegos MOBA: estudio sobre la fragmentariedad, *Estudios Interlingüísticos*, 6, 27-41.
- Badia Climent, S. (2023). *Lenguaje coloquial y videojuegos en línea. El caso de los MOBA: hacia un nuevo tipo de comunicación*. Tesis doctoral. Universidad de Valencia.
- Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, 15, 415–419. doi: 10.1111/j.0956-7976.2004.00694.x
- Barakat R. A. (1969). Gesture systems. *Keystone Folklore Quarterly*, 14, 105-121.
- Barros, P. (ed.) (2012). *COGILA. Español oral conversacional: corpus y guía didáctica*. Granada: Editorial Universidad de Granada.
- Bateman, J., Wildfeuer, J. & T. Hiippala, (2017). *Multimodality: Foundations, Research and Analysis – A Problem-Oriented Introduction*. Berlin: De Gruyter Mouton.
- Battersby, S. A., Lavelle, M., Healey, P. G. T., & McCabe, R. (2008). Analysing Interaction: A comparison of 2D and 3D techniques, en *Conference on multimodal corpora*. Marrakech.
- Battersby, S. A. y Healey, P.G. (2010). Head and hand movements in the orchestration of dialogue, en S. Ohlsson & R. Catrambone (eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*.

- Battersby, S. (2011). *Moving together: The organization of non-verbal cues during multiparty conversation*. Tesis doctoral, Queen Mary, University of London.
- Bavelas, J. B. (1994). Gestures as part of speech: Methodological implications. *Research on Language and Social Interaction*, 27, 201-221.
- Bavelas, J. B., Chovil, N., Lawrie, D. A. y Wade, A. (1992). *Interactive gestures*. *Discourse Processes*, 15, 469–489.
- Bavelas, J., Chovil, N., Coates, L. y Roe, L. (1995). Gestures specialized for dialogue. *Personality and Social Psychology Bulletin*, 21 (4), 394–405. <https://doi.org/10.1177/0146167295214010>
- Bavelas, J. B., Coates L., Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology Bulletin*, 21, 395-405.
- Bavelas, J. B., Holt, T. y Allison, M. (2000). *Links between gestures and words*. No publicado.
- Bavelas, J. B., Coates L., Johnson, T. (2002) Listeners Responses as a Collaborative Process: The Role of Gaze. *Journal of Communication*, 52(3), 566-580.
- Bavelas, J. B., T. J. Kenwood y B. Phillips (2002). An experimental study of when and how speakers use gesture to communicate. *Gesture* 2(1), 1-17.
- Bavelas, J., Gerwing, J., Sutton, C. y & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, 58(2), 495–520. <https://doi.org/10.1016/j.jml.2007.02.004>
- Bavelas J, Gerwing J, Healing S. (2014). Including facial gestures in gesture-speech ensembles, en M. Seyfeddinipur y M. Gullberg (Eds.), *Conversation to Visible Action as Utterance: Essays in honor of Adam Kendon*. John Benjamins Publishing Company.
- Bavelas, J. B., Gerwing, J., Healing, S., & Tomori, C. (2016). Microanalysis of Face-to-face Dialogue. An Inductive Approach, en C. A. VanLear y D. J. Canary (Eds.), *Researching communication interaction behavior: A sourcebook of methods and measures* (129-157). Thousand Oaks, CA: Sage.
- Bavelas, J. y Chovil, N. (1997). Faces in dialogue, en J. A. Russell y J. M. Fernandez-Dols (Eds.), *The psychology of facial expression* (334–346). Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9780511659911.017>
- Bavelas J. B. y Chovil, N. (2000). Visible acts of meaning: An integrated message model of language in face-to-face dialogue. *Journal of Language and Social Psychology*, 19(2), 163–94.
- Bavelas, J. B., y Chovil, N. (2006). Hand gestures and facial displays as part of language use in face-to-face dialogue, en V. Manusov y M. Patterson (Eds.), *Handbook of nonverbal communication* (97-115). Thousand Oaks, CA: Sage.
- Bavelas, J. B. y Chovil, N. (2018) Some pragmatic functions of conversational facial gestures. *Gesture*, 17(1), 98-127.
- Belio-Apaolaza, H. S. (2019). La naturaleza lingüístico-comunicativa de los gestos emblemáticos: una propuesta de representación, *Círculo de Lingüística Aplicada a la Comunicación (CLAC)*, 80, 115-135.

- Belío-Apaolaza, H. S. (2022). Propuesta taxonómica para el análisis contrastivo de los gestos emblemáticos, *ELUA: Estudios de Lingüística Universidad de Alicante*, 37, 235-252.
- Bernstein, B. (1962). Linguistic Codes, Hesitation Phenomena and Intelligence. *Language and Speech*, 5, 31-47.
- Berrendonner, A. (1990), Pour une macro-syntaxe, *Travaux de linguistique*, 21, 25-36.
- Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., et al. (2008). Le CID—Corpus of interactional data—Annotation et exploitation multimodale de parole conversationnelle. *Traitement automatique des langues*, 49, 105-134.
- Betti, S. (2007). Comunicación no verbal y gestualidad: El cómo se dice algo es más importante que lo que se dice. Estudio contrastivo español e italiano, *Cuadernos de Italia y Grecia*, 6.
- Birdwhistell, R.L. (1952). *Introduction to Kinesics: An Annotated System for Analysis of Body Motion and Gesture*. Washington, D.C.: Department of State, Foreign Service Institute; Louisville: University of Kentucky.
- Birdwhistell, R. L. (1970). *Kinesics and context*. Philadelphia, PA: University of Pennsylvania Press.
- Blache, P., Bertrand, R., Ferré, G., Pallaud, B., Prévot, L., Rauzy, S. (2017). The Corpus of Interactional Data: A Large Multimodal Annotated Resource, en N. Ide, J. Pustejovsky (Eds.) *Handbook of Linguistic Annotation*. Springer, Dordrecht. https://doi.org/10.1007/978-94-024-0881-2_51
- Blanche-Benveniste, C. y Jeanjean, C. (1987). *Le français parlé*. Paris: Didier Erudition.
- Blakemore, D. (1987). *Semantic Constraints on Relevance*. London: Basil-Blackwell.
- Bohle, U. (2007). *Das Wort ergreifen - das Wort übergeben. Explorative Studie zur Rolle redegleitender Gesten in der Organisation des Sprecherwechsels*. Berlin: Weidler.
- Bohle, U. (2013). Gesture and conversational units. En C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (eds.), en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 2: An International Handbook on Multimodality in Human Interaction* (1360-1367). Berlin: De Gruyter Mouton.
- Bohle, U. (2014). Contemporary classification systems, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 2: An International Handbook on Multimodality in Human Interaction* (1453-1461). Berlin: De Gruyter Mouton.
- Bolden, G. (2003). Multiple modalities in collaborative turn sequences. *Gesture*, 3(2), 187–212.
- Bouvet, D. y M. A. Morel. (2002). *Le ballet et la musique de la parole. Le geste et l'intonation dans le dialogue oral en français*. Paris-Gap, Ophrys, Bibliothèque de Faits de Langues.
- Borod, J. C., Obler, L. K., Erhan, H. M., Grunwald, I. S., Cicero, B. A., Welkowitz, J., Santschi, C., Agosti, R. M. y Whalen J. R. (1998). Right hemisphere emotional perception: evidence across multiple channels. *Neuropsychology*, 12(3), 446-458.

- Bravo, D. (1999). ¿Imagen 'positiva' vs. imagen 'negativa'?, *Oralia* 2, 155-184.
- Bressemer, J. (2012). *Repetitions in gesture: Structures, functions, and cognitive aspects*. Tesis doctoral, Faculty of Social and Cultural Sciences, European University Viadrina, Frankfurt (Oder).
- Bressemer, J. (2013a). Transcription systems for gestures, speech, prosody, postures, and gaze. En C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tesselndorf (eds.), en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tesselndorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (1037-1059). Berlin: De Gruyter Mouton.
- Bressemer, J. (2013b). A linguistic perspective on the notation of form features in gestures, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tesselndorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (1079-1097). Berlin: De Gruyter Mouton.
- Bressemer, J. (2021). *Repetitions in Gesture: A Cognitive-Linguistic and Usage-Based Perspective*. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110697902>
- Bressemer, J., Ladewig, S., Müller, C., (2013). Linguistic Annotation System for Gestures. en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tesselndorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (1098- 1124). Berlin: De Gruyter Mouton.
- Bressemer, J., y Müller, C. (2014a). A repertoire of German recurrent gestures with pragmatic functions, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tesselndorf (Eds.), *Body - Language - Communication. Volume 2: An International Handbook on Multimodality in Human Interaction* (1575–1592). Berlin: De Gruyter Mouton.
- Bressemer, J., y Müller, C. (2014b). The family of Away gestures: Negation, refusal, and negative assessment, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tesselndorf (Eds.), *Body - Language - Communication. Volume 2: An International Handbook on Multimodality in Human Interaction* (1592-1604). Berlin: De Gruyter Mouton.
- Bressemer, J. y Müller, C. (2017). The “Negative-Assessment-Construction” – A multimodal pattern based on a recurrent gesture?. *Linguistics Vanguard*, 3(s1).
- Briz, A. (coord. 1995): La conversación coloquial (Materiales para su estudio). *Cuadernos de Filología*, Anejo XVI. València: Universitat de València.
- Briz, A. (1996): *El español coloquial: situación y uso*. Madrid: Arco Libros.
- Briz, A. (1998): *El español coloquial en la conversación. Esbozo de pragmagramática*. Barcelona: Ariel.
- Briz, A. (2000). Turnos y alternancia de turnos en la conversación. *Revista Argentina de Lingüística*, 16, 9-32.
- Briz, A. y Grupo Val.Es.Co (2002a). La transcripción de la lengua hablada: el sistema del grupo Val.Es.Co, *Español actual*, 77.
- Briz, A. y Grupo Val.Es.Co. (2002b). *Corpus de conversaciones coloquiales. Anejo de la Revista Oralia*. Madrid: Arco/Libros.

- Briz, A. (2002-2004). La estructura de la conversación. Orden externo y orden interno. *Archivo de filología aragonesa (AFA)*, 59-60, 265-280.
- Briz, A. y grupo Val.Es.Co (2003). *La estructura de la conversación coloquial. Unidades y estructuras. Actas del XXIII Congreso Internacional de Lingüística y Filología Románica*. Salamanca: Max Miemeyer Verlag.
- Briz, A. (2006). La segmentación de una conversación en diálogos. *Oralia* 9: 45-71.
- Briz, A. (2007a). Límites para el análisis de la conversación. Órdenes y unidades. *Revista Internacional de Lingüística Iberoamericana (RILI)*, 5, 23-37.
- Briz, A. (2007b). La unidad superior del discurso (conversacional): el diálogo, en L. Cortés, A. Bañón, M. M. Espejo y J. L. Muñío (Coords.) *Discurso y oralidad. Homenaje al profesor José Jesús de Bustos Tovar. Anejo 3/1 de la Revista Oralia*, 15-40.
- Briz, A. (2010a): Lo coloquial y lo formal, el eje de la variedad lingüística, en R. M. Castañer Martín y V. Lagüéns García (Coords.), *De moneda nunca usada: Estudios dedicados a José M^a. Enguita Utrilla* (125-133). Zaragoza: Instituto Fernando El Católico.
- Briz, A. (2010b): El registro como centro de la variedad situacional. Esbozo de la propuesta del grupo Val.Es.Co. sobre las variedades diafásicas, en I. Fonte y L. Rodríguez (Comps.), *Perspectivas dialógicas en estudios del lenguaje* (21-56). México: Universidad Autónoma de Nuevo León.
- Briz, A. (2012a). La “constelación” comunicativa coloquial: hacia un modo más dinámico de entender “lo coloquial”, *Español actual: Revista de español vivo*, 98, 217-226.
- Briz, A. (2012b). La definición de las partículas discursivas *hombre* y *mujer*. *Anuario de Lingüística Hispánica*, 28, 27-55.
- Briz, A. y Pons, S. (2010). Unidades, marcadores discursivos y posición. En Ó. Loureda y E. Acín (Coords.), *Los estudios sobre marcadores del discurso en español, hoy* (327-258). Madrid: Arco Libros.
- Briz, A., Pons, S. y J. Portolés (Coords.) (desde 2008). Diccionario de partículas del español. Recuperado el 20 de junio de 2023 de <<http://www.dpde.es>>.
- Briz y Villalba. *hombre*¹, en Briz, A., S. Pons y J. Portolés (Coords.) Diccionario de partículas discursivas del español. Recuperado el 20 de junio de 2023 de <<http://www.dpde.es/#/entry/hombre1>>.
- Brookes, H. J. (2001). O clever ‘He’s streetwise’. When gestures become quotable: The case of the clever gesture. *Gesture*, 1, 167-184. doi: 10.1075/gest.1.2.05bro
- Brône, G., Oben, B. (2015) InSight Interaction: a multimodal and multifocal dialogue corpus, *Language Resources & Evaluation* 49, 195-214. DOI 10.1007/s10579-014-9283-2
- Brown, R. y Levinson, A. G. (1978; 1987): *Politeness. Some Universals in Language Use*, Cambridge University Press.
- Brunner, L.J. (1979). Smiles can be back channels. *Journal of Personality and Social Psychology*, 37, 728-734.

- Bugental, D.B., Kaswan, J.W. y Love, L.R. (1970). Perception of Contradictory Meanings Conveyed by Verbal and Nonverbal Channels. *Journal of Personality and Social Psychology*, 16(64), 7-55.
- Burgoon, J., Guerrero, L. y White, C. (2013). The codes and functions of nonverbal communication. en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (609-626). Berlin: De Gruyter Mouton.
- Buschmeier, H., Malisz, Z., Skubisz, J., Włodarczak, M., Wachsmuth, I., Kopp, S., et al. (2014). ALICO: A multimodal corpus for the study of active listening, en *Proceedings of the 9th conference on language resources and evaluation* (3638-3643). Iceland, Reykjavík.
- Cabanes Pérez, S. (2020), La quinésica en la identificación de intervenciones y turnos, *Estudios Interlingüísticos*, 8, 39-40.
- Cabedo, A. y Pons, S. (Eds.). *Corpus Val.Es.Co 2.0*. Recuperado el 20 de junio de 2023 de <http://www.valesco.es>.
- Calbris, G. (1990). *The Semiotics of French Gestures*. Bloomington: Indiana University Press.
- Calbris, G. (2011). *Elements of meaning in gesture*. Amsterdam & Philadelphia: John Benjamins Publishing. <https://doi.org/10.1075/g5>
- Canavan, A. y Zipperlen G. (1996a). *CALLHOME Spanish Speech*. Philadelphia: Linguistic Data Consortium.
- Canavan, A. y Zipperlen G. (1996b). *CALLHOME Spanish Non-Caribbean Dialect*. Philadelphia: Linguistic Data Consortium.
- Cestero, A. M. (1996): Funciones de la risa en la conversación en lengua española, *Lingüística Española Actual*, 18(2), 279-289.
- Cestero, A. M. (Coord.) (1998). *Estudios de comunicación no verbal*. Madrid: Edinumen.
- Cestero, A. M. (1999). *Repertorio Básico de Signos no Verbales del Español*. Madrid: Arco Libros.
- Cestero Mancera, A. M. (2000a). *El intercambio de turnos de habla en la conversación (Análisis sociolingüístico)*. Universidad de Alcalá.
- Cestero Mancera, A. M. (2000b). *Los turnos de apoyo conversacionales*. Servicio de Publicaciones de la Universidad de Cádiz.
- Cestero, A. M. (2005). *Conversación y enseñanza de lenguas extranjeras*. Madrid: Arco Libros.
- Cestero, A.M. (2006). La comunicación no verbal y el estudio de su incidencia en fenómenos discursivos como la ironía. *ELUA*, 20, 57-77.
- Cestero, A.M. (2009). Marcas paralingüísticas y quinésicas de la ironía, en Ruiz Gurillo, L. y Padilla García, X. A. (eds.), *Dime cómo ironizas y te diré quién eres. Una aproximación pragmática a la ironía* (167-190). Frankfurt am Main, Peter Lang (Internationaler Verlag der Wissenschaften).
- Cestero, A. M. (2016), La Comunicación no verbal: propuestas metodológicas para su estudio. *LinRed, monográfico*, Universidad de Alcalá.

- Cestero, A.M. (2017). La comunicación no verbal en discurso persuasivo empresarial. *Pragmalingüística*, 25, 124-145.
- Cestero, A.M., Forment, M., Gelabert, M.J., Martinell, E. (2020). *Hablar en español sin palabras Diccionario audiovisual de gestos españoles*. Alcalá de Henares, Servicio de Publicaciones de la Universidad de Alcalá. http://www3.uah.es/meleuah/diccionario_gestos/#
- Chovil, N. (1991). Discourse-oriented facial displays in conversation. *Research on Language and Social Interaction*, 25, 163–194. <https://doi.org/10.1080/08351819109389361>
- Cienki A. (1998). Metaphoric gestures and some of their relations to verbal metaphoric expressions, en J.-P. Koenig (Ed.). *Discourse and cognition: Bridging the gap* (189-204). Stanford, CA: Center for the Study of Language and Information.
- Cienki, A. (2013). Image schemas and mimetic schemas in cognitive linguistics and gesture studies. *Review of Cognitive Linguistics*, 11(2): 417–432.
- Cienki, A. (2017). Gesture and pragmatics: From paralinguistic to variably linguistic, en A. Barron, G. Steen y Y. Gu (Eds.), *The Routledge Handbook of Pragmatics*. London: Routledge, 61–68.
- Cienki, A., y Muller, C. (2008a). Metaphor, Gesture, and Thought, en R. W. G. Jr. (Ed.), *The Cambridge Handbook of Metaphor and Thought* (483-501). Cambridge: Cambridge University Press.
- Cienki, A. y Müller, C. (eds.) (2008b). *Metaphor and Gesture*. Amsterdam: John Benjamins.
- Clark, H. H. (1996). *Using Language*. Volumen 4. Cambridge: Cambridge University Press.
- Clark H. H. y Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50(1), 62–81.
- Clift, R. (2022). Embodiment in dissent: The eye roll as an interactional practice, *Research on Language and Social Interaction*, 54(3), 261-276.
- Coates, L. J. (1991). *A Collaborative Theory of Inversion: Irony in Dialogue*. M.A. thesis, University of Victoria, BC.
- Coll, J., Gelabert, M. J. y Martinell, E. (1990). *Diccionario de gestos con sus giros más usuales*. Madrid: Edelsa.
- Cooperrider, K. (2017). Foreground gesture, background gesture. *Gesture* 16(2), 176-202. doi 10.1075/gest.16.2.02coo
- Cortés Rodríguez, L. (2002). Los estudios del español hablado entre 1950 y 1999. Periodos, disciplinas y corrientes. *Anejos de Oralia*, 2. Recuperado el 20 de junio de 2023 de: <https://ojs.uah.es/ojs/index.php/anejosoralia/article/view/8975>
- Cosnier, J. y Vaysse, J. (1997). Sémiotique des gestes communicatifs. *Nouveaux actes sémiotiques*, 52, 7-28.
- Coulthard, M. y D. Brazil (1992). Exchange structure, en M. Coulthard (ed.) *Advances in Spoken Discourse Analysis*, Routledge.
- Coulthard, M. y M.M. Montgomery (Eds.) (1981). *Studies in Discourse Analysis*, London: Routledge & Kegan Paul.

- Couper-Kuhlen, E. (2001). Interactional prosody: High onsets in reason-for-the-call turns. *Language in Society*, 30(1), 29-53.
- Couper-Kulen, E. y Selting, M. (2001). *Studies in Interactional Linguistics*. Amsterdam: John Benjamins.
- Couper-Kulen, E. y Selting, M. (2018). *Interactional Linguistics: Studying Language in Social Interaction*. Cambridge University Press.
- Crespo Sendra, V., C. Kaland, M. Swerts y Prieto, P. (2013): Perceiving incredulity: The role of intonation and facial gestures. *Journal of Pragmatics*, 47(1), 1-13.
- Cresti, E. y Moneglia, M. (2005). *C-ORAL-ROM Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/scl.15>
- Criado de Val, Manuel (1980). *Estructura general del coloquio*, Madrid: SGEL.
- Darwin, C. (1872). *Expression of the Emotions in Man and Animals*. London: John Murray.
- Debras, C. (2017). The shrug. Forms and meanings of a compound enactment. *Gesture* 16(1), 1-34. doi 10.1075/gest.16.1.01deb
- De Fornel, Michel (1992). The return gesture: Some remarks on context, inference and iconic gesture, en P. Auer y A. di Luzio (Eds.), *The contextualization of language*, (159-176). Amsterdam: John Benjamins. doi: 10.1075/pbns.22.11for
- Degand, L. y Simon, A. C. (2009). Minimal Discourse Units in Spoken French: On the Role of Syntactic and Prosodic Units in Discourse Segmentation. *Discours*, 4.
- Deppermann, A. (2013) Multimodal interaction from a conversation analytic perspective, *Journal of Pragmatics*, 46, 1-7.
- De Ruiter, J. P. (2000). The production of gesture and speech, en David McNeill (ed.), *Language and Gesture* (284-311). Cambridge: Cambridge University Press.
- De Ruiter, J. P., Mitterer, H. y Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation, *Language*, 82(3): 504–524.
- De Sanctis, S. (1902). *La Mimica del Pensiero*. Palermo: Sandron.
- Dittman, A.T. y L.G. Llewellyn (1967). The phonemic clause as a unit of speech decoding. *Journal of Personality and Social Psychology*, 6, 341-349.
- Dittman, A.T. y L.G. Llewellyn (1968). Relationship between vocalizations and head nods as listener responses. *Journal of Personality and Social Psychology*, 9, 79-84.
- Dressel, D. (2020). Multimodal Word Searches in Collaborative Storytelling. On the local mobilization and negotiation of coparticipation. *Journal of Pragmatics*, 170, 37-54.
- Ducrot, O. et al. (1980). *Les mots du discours*. Paris, Minuit.
- Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23, 283–292.
- Duncan, S. (1973). Toward a grammar for dyadic conversation. *Semiotica*, 9, 29-46.
- Duncan S. y D. W. Fiske (1977). *Face to Face Interaction: Research, Methods and Theory*. Hillsdale, N. J.: Laurence Erlbaum.

- Edelsky, C. (1981). Who's got the floor?. *Language in Society*, 10, 383-421. Cambridge University Press.
- Edmonson, W. (1981). *Spoken Discourse: A Model for Analysis*. London: Longman.
- Efron, D. (1970 [1941]). *Gesto, raza y cultura*. Buenos Aires, Nueva Visión.
- Ehlich, K. (2013). Nonverbal communication in a functional pragmatic perspective. en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessedorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (648-658). Berlin: De Gruyter Mouton.
- Ehlich, K. y Rehbein, J. (1979). Erweiterte halbinterpretative Arbeitstranskriptionen (HIAT2). *Linguistische Berichte*, 59, 51-75.
- Ekman, P. (1976). Movements with precise meaning. *Journal of Communication* 26(3), 14-26.
- Ekman, P. y Friesen, W. V. (1969). The repertoire of non-verbal behaviour: categories, origins, usage and coding, en A. Kendon (ed.), *Non-verbal communication, interaction, and gesture. Selections from Semiotica*. The Hague, Mouton Publishers, 1981.
- Ekman, P. y Friesen, W. V. (1974). Nonverbal behaviour and psychopathology, en R. J. Friedman y M. M. Katz (Eds.), *The psychology of depression: Contemporary theory and research* (203-232). Washinton, DC: J. Winston.
- Ekman, P. y Friesen, W. V. (1978). *The facial action coding system (FACS)*. Palo Alto, California: Consulting Psychologists Press.
- Elfenbein, H. A. y Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128, 203-235. doi: 10.1037/0033-2909.128.2.203
- Enfield, N.J. (2009). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. New York: Cambridge University Press.
- Enfield, N.J. (2011). Elements of formulation, en J. Streeck, C. Goodwin y C. Le-Baron (Eds.), *Embodied Interaction: Language and the Body in the Material World* (59-66). Cambridge: Cambridge University Press.
- Enfield, N. (2013). Doing fieldwork on the body, language, and communication, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessedorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (974-981). Berlin: De Gruyter Mouton.
- Engle, R. A. (2000). *Toward a theory of multimodal communication combining speech, gestures, diagrams, and demonstrations in instructional explanations*. Tesis doctoral, Stanford University.
- Eriksson, M. (2009). Referring as interaction: On the interplay between linguistic and bodily practices. *Journal of Pragmatics*, 41 (2), 240-262.
- Espinosa Guerri, G. (2016). Los dientes de sierra: una herramienta para el estudio de la estructura interactiva del discurso dialógico. *Normas*, 6, 13-26.
- Espinosa-Guerri, G. y García-Ramón, A. (2019): A preliminary typology of interactional figures based on a tool for visualizing conversational structure, en Ó. Loureda, I.

- Recio, L. Nadal y A. Cruz (eds.), *Empirical Studies of the Construction of Discourse* (93-130). Ámsterdam: John Benjamins.
- Estellés, M. y S. Pons (2014). Absolute initial position, en S. Pons Bordería (Ed.), *Discourse Segmentation in Romance Languages* (121–155). John Benjamins.
- Esteve-Gibert, N. (2016). La integración de gestos y habla en el discurso, en M.C. Horno Chéliz, I. Ibarretxe Antuñano, J.L. Mendivil Giró (Eds.), *Panorama actual de la ciencia del lenguaje, Primer sexenio de Zaragoza Lingüística*. Prensas de la Universidad de Zaragoza.
- Esteve-Gibert, N., Igualada, A., Prieto, P. (2019). El gest com a facilitador i precursor del desenvolupament del llenguatge. *Llengua, Societat i Comunicació*, 17, 26-41.
- Esteve-Gibert, N. y Prieto, P. (2013). Prosodic Structure Shapes the Temporal Realization of Intonation and Manual Gesture Movements, *Journal of Speech, Language, and Hearing Research*, 56(3), 850-864.
- Feyaerts, K., Papousek, I., Lackner, H. K., Gierlinger, G., Oben, B. (2012). *AliLeGra Corpus. Alignment Leuven Graz*. Dept. of Linguistics, KU Leuven.
- Forceville, C. (2020): *Visual and multimodal communication: Applying the relevance principle*. Oxford: Oxford University Press.
- Ford, C. E. y Fox, B. A. (2010). Multiple practices for constructing laughables, en E. Reber, D. Barth-Weingarten y M. Selting (Eds.), *Prosody in interaction* (339-368). Amsterdam: John Benjamins.
- Forment Fernández, M. (1997). La verbalización de la gestualidad en el aprendizaje de E/LE. *Frecuencia-L*, 4, 27-31.
- Fox, B. A. (2001). On the embodied nature of grammar: Embodied being-in-the-world, en J. Bybee y M. Noonan (Eds.), *Complex sentences in grammar and discourse* (79-100).
- Fox, B. A. y Thompson, S. A. (2010). Responses to Wh-Questions in English conversation. *Research on Language and Social Interaction*, 43 (2), 133-156.
- Fox, B., Thompson, S.A., Ford, C.E., Couper-Kuhlen, E. (2013). Conversation analysis and linguistics, en J. Sidnell y T. Stivers (Eds.), *The Handbook of Conversation Analysis* (726-740). Wiley-Blackwell.
- Frake, C.O. (1972). How to Ask for a Drink in Subanun, en P.P. Giglioli (Ed.): *Language and Social Context* (87-93). Baltimore: Penguin.
- Francis, G. y Hunston, S. (1992). Analysing everyday conversation, en M. Coulthard (Ed.) *Advances in Spoken Discourse Analysis* (123-161). Routledge.
- Frank, D. (1980): *Grammatik und Konversation*. Königstein, Scriptor.
- Freedman, N. (1977). Hands, words and mind: On the structuralization of body movements during discourse and the capacity for verbal representation, en N. Freedman y S. Grand (Eds.), *Communicative Structures and Psychic Structures* (109-132). New York: Plenum.
- Fries, C.C. (1952). *The structure of English*. New York: Harcourt, Brace.
- Fricke, E. (2007). *Origo, Geste und Raum: Lokaldeixis im Deutschen*. Berlin: Walter de Gruyter.

- Fricke, E. (2012). *Grammatik multimodal: Wie Wörter und Gesten zusammenwirken*. Berlin/Boston: De Gruyter.
- Fricke, E. (2013). Towards a unified grammar of gesture and speech: a multimodal approach, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessedorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction (733-754)*. Berlin: De Gruyter Mouton.
- Furuyama, N. (2000). Gestural interaction between the instructor and the learner in origami instruction, en D. McNeill (Ed.), *Language and Gesture (99-117)*. Cambridge: Cambridge University Press.
- Gallardo, B. (1992): El dinamismo conversacional: subsunción y feedback. Carácter acumulativo de los elementos conversacionales, *Communication and Society*, 5(1-2), 1-10.
- Gallardo, B. (1996). *Análisis conversacional y pragmática del receptor*. Ediciones Episteme
- Gallardo, B. (1998). *Comentario de textos conversacionales I. De la teoría al comentario*. Madrid, Arco/Libros.
- Gallardo Paúls, B. y Sanmartín Sáez, J. (2005). Afasia fluente. Materiales para su estudio (Corpus PerLA, V. 1).Valencia: Universitat/ AVaLCC.
- García García, M. (2004). El lugar de la comunicación no verbal en la clase de E.L.E., Kinésica contrastiva. *Biblioteca virtual redELE*, 2(2º semestre), 79.
- García-Ramón, A. (2020). Figuras interaccionales y propósito comunicativo en conversaciones coloquiales, entrevistas periodísticas y entrevistas del proyecto PRESEEA. *Círculo de Lingüística Aplicada a la Comunicación*, 82, 119-136, <http://dx.doi.org/10.5209/clac.68968>
- Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
- Gerwing, J. y Bavelas J. B. (2004). Linguistic influences on gesture's form. *Gesture*, 4 (2), 157-195.
- Gerwing, J. y Bavelas, J. B. (2013). The social interactive nature of gestures: theory, assumptions, methods, and findings. en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessedorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction (821-836)*. Berlin: De Gruyter Mouton.
- Givens, D. (1977). Shoulder shrugging: a densely communicative behavior. *Semiotica*, 19 (1/2), 13-29.
- Glenn, P. J. (1989). Initiating Shared Laughter in Multi-Party Conversations. *Western Journal of Speech Communication*, 53, 127-149.
- Goffman, E. (1961). *Encounters: Two Studies in the Sociology of Interaction*. Indianapolis: Bobbs-Merrill.
- Goffman, E. (1963). *Behavior in Public Places: Notes on the Social Organization of Gathering*. New York: Free Press.
- Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior*. Garden City, N.Y.: Doubleday/Anchor.

- Goffman, E. (1971). *Relations in Public: Microstudies of the Public Order*. New York: Harper and Row.
- Goffman, E. (1976). Replies and responses. *Language in Society*, 5, 257-313.
- Goffman (1983). The interaction order. *American Sociological Review*, 48,1-17.
- Goodwin, C. (1979). The Interactive Construction of a Sentence in Natural Conversation, en G. Psathas (Ed.), *Everyday Language: Studies in Ethnomethodology* (97-121). New York: Irvington Publishers.
- Goodwin, C. (1980). Restarts, pauses, and the achievement of mutual gaze at turn-beginning. *Sociological Inquiry*, 50, 272-302.
- Goffman, E. (1981). *Forms of Talk*. Philadelphia: Univ. Penn. Press
- Goodwin, C. (1981). *Conversational Organization. Interaction between Speakers and Hearers*. New York: Academic Press.
- Goodwin, C. (1986). Gesture as a resource for the organization of mutual orientation, *Semiotica*, 62(1/2), 29-49.
- Goodwin, C. (1993). Recording human interaction in natural settings. *Pragmatics*, 3(2), 181-209.
- Goodwin, C. (2000). Action and embodiment within human interaction. *Journal of Pragmatics*, 32, 1489-1522.
- Goodwin, C. (2003). Pointing as situated practice, en S. Kita (Ed.), *Pointing: Where Language, Culture and Cognition Meet* (217-241). Hillsdale, NJ: Lawrence Erlbaum.
- Goodwin, C. (2006). Human Sociality as Mutual Orientation in a Rich Interactive Environment: Multimodal Utterances and Pointing in Aphasia, en N. J. Enfield y S. C. Levinson (Eds.), *Roots of Human Sociality* (96-125). London: Berg.
- Goodwin, C. (2007). Environmentally coupled gestures, en S. Duncan, J. Cassel y E. Levy (Eds.), *Gesture and the Dynamic Dimensions of Language* (195-212). Amsterdam: John Benjamins.
- Goodwin, C. (2009). Embodied Hearers and Speakers Constructing Talk and Action in Interaction. *Cognitive Studies*, 16(1), 51-64.
- Goodwin, C. (2010). Constructing meaning through prosody in aphasia, en: D. Barth-Weingarten, E. Reber y M. Selting (Eds.), *Prosody in Interaction* (373-394). Amsterdam: John Benjamins.
- Goodwin, C. H. (2018). *Co-Operative Action*. Cambridge, MA: Cambridge University Press.
- Goodwin, C. y Heritage, J., (1990). Conversation Analysis. *Annual Review of Anthropology*, 283-307.
- Goodwin, M H. (1980). Processes of mutual monitoring implicated in the production of descriptive sequences. *Sociological Inquiry*, 50, 303-317.
- Goodwin, M. H. y Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62, 51-75.
- Green, J. R. (1968). *Gesture Inventory for the Teaching of Spanish*. Philadelphia: Chilton Books.

- Grice, H. P. (1975). Logic and conversation, en P. Cole y J. L. Morgan (Eds.). *Syntax and Semantics. Speech Acts* (41-58). Nueva York: Academic Press.
- Groupe ICOR (H. Baldauf-Quilliatre, I. Colon de Carvajal, C. Etienne, E. Jouin-Chardon, S. Teston-Bonnard, V. Traverso) (2016), CLAPI, une base de données multimodale pour la parole en interaction : apports et dilemmes , en M. Avanzi, M.J. Béguelin y F. Diémoz (Eds), *Corpus de français parlés et français parlés des corpus, Cahiers Corpus*.
- Grupo Val.Es.Co. y Briz Gómez, A. (Coord.) (1995): *La conversación coloquial (materiales para su estudio)*. València: Universitat de València, Departamento de Filología Española.
- Grupo Val.Es.Co. (2014). Las unidades del discurso oral. La propuesta Val.Es.Co. de segmentación de la conversación (coloquial). *Estudios de Lingüística del Español*, 35.1, 11-71.
- Gullber, M. y Holmqvist, K. (2006). What speakers do and what addressees look at. Visual attention to gestures in human interaction live and on video. *Pragmatics and Cognition*, 14(1), 53-82. doi: 10.1075/pc.14.1.05gul
- Gullberg, M. y Kita, S. (2009). Attention to speech-accompanying gestures: Eye movements and information uptake. *Journal of Nonverbal Behavioral*, 33 (4), 251-277. <https://doi.org/10.1007/s10919-009-0073-2>
- Gut, U., Looks, K., Thies A., Gibbon, D. (2002). *Cogest: Conversational gesture transcription system version 1.0*. Fakultät für Linguistik und Literaturwissenschaft, Universität Bielefeld, ModeLex Tech. Rep 1.
- Hadar, U., Steiner, T. J., Grant, E. C. y Clifford Rose, F. (1984) The timing of shifts of head postures during conversation. *Human Movement Science*, 3, 237-245.
- Haddington, P., Mondada, L., Nevile, M. (Eds.) (2013). *Interaction and Mobility: Language and the Body in Motion*. Berlin: De Gruyter.
- Haensel, J. X., Smith, T. J. y Senju, A. (2022). Cultural differences in mutual gaze during face-to-face interactions: A dual head-mounted eye-tracking study, *Visual Cognition*, 30(1-2), 100-115.
- Hagemann, J. (2014). Proxemics and axial orientation, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessedorf (Eds.), *Body - Language - Communication. Volume 2: An International Handbook on Multimodality in Human Interaction* (1310-1327). Berlin: De Gruyter Mouton.
- Hanna, B.E. (1996). Defining the emblem. *Semiotica*, 112(3/4), 289-358.
- Hall, E. T. (1963). Proxemics: The Study of Man's Spatial Relations, en I. Galdston (Ed.), *Man's Image in Medicine and Anthropology* (422-445). Nueva York: International University Press.
- Hall, E. T. (1966). *The Hidden Dimension*. Nueva York: Doubleday.
- Harris, Z. (1951). *Methods in Structural Linguistics*. Chicago: University of Chicago Press.
- Harrison, S. (2009). *Grammar, gesture, and cognition. The case of negation in English*. Tesis doctoral, Université Michel de Montaigne Bordeaux 3.

- Harrison, S. (2010). Evidence for node and scope of negation in coverbal gesture, *Gesture* 10(1), 29-51.
- Haverkate, H. (1994). *La cortesía verbal. Estudio pragmatolingüístico*. Madrid, Gredos.
- Haviland, J. B. (1993). Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology*, 3(1), 3-45.
- Hay, J. (2000). Functions of humor in the conversation of men and women. *Journal of Pragmatics*, 32, 709-742.
- Hayashi, M. (2003). Language and the body as resources for collaborative action: a study of word searches in Japanese conversation. *Research on Language and Social Interaction*, 36 (2), 109-141.
- Hayashi, M. (2005). Joint turn construction through language and the body: Notes on embodiment on coordinated participation in situated activities. *Semiotica*, 156(1/4), 21-53.
- Healey, P. G. T. y Battersby, S. A. (2009). The Interactional Geometry of a Three-way Conversation, en *Proceedings of the 31st annual conference of the cognitive science society* (785-790). Amsterdam.
- Heath, C. (1986). *Body movement and speech in medical interaction*. Cambridge: Cambridge University Press.
- Heath, C. (1992). The delivery and reception of diagnosis and assessment in the general practice consultation, en P. Drew y J. Heritage (Eds.), *Talk at work: Interaction in institutional settings* (235-267). Cambridge: Cambridge University Press.
- Heath, C. y Luff, P. (1992). Collaboration and control: Crisis management and multimedia technology in London Underground Line Control Rooms. *CSCW*, 1 (1-2), 69-94.
- Hengeveld, K. y Mackenzie, J. L. (2008). *Functional Discourse Grammar. A typologically based theory of language structure*. Oxford: Oxford University Press.
- Heritage, J. y J.M. Atkinson (1984). Introduction, en J.M. Atkinson y J. Heritage (Eds.), *Structures of Social Action* (1-16). Cambridge University Press.
- Hernández, A., Alemany, A., Badia, S., Cabanes, S., Castelló, C., Córcoles, T., Murillo, D., Pardo, A., Pascual, E., Salameh, S. (Grupo Val.Es.Co.) (2022). La coloquialidad en Paquita Salas desde el modelo de unidades conversacionales de Val.Es.Co., en V. Pérez Béjar y M. Méndez Orense (Eds.), *Perspectivas integradas para el análisis de la oralidad*. Sevilla: Universidad de Sevilla.
- Hidalgo Navarro, A. y Sanmartín Sáez, J. (2005). Los sistemas de transcripción de la lengua hablada. *Oralia: Análisis del discurso oral*, 8, 13-36. <https://ojs.ual.es/ojs/index.php/ORALIA/article/view/8325>
- Hoetjes, M., E. Krahmer, y M. Swerts (2014). On what happens in speech and gesture when communication is unsuccessful, en *Proceedings of the 36th annual meeting of the Cognitive Science Society (CogSci), July 2014, Québec City, Canada* (2369-2374). Austin, TX: Cognitive. Science Society.
- Holler, J. (2013). Experimental methods in co-speech gesture research, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 2: An International Handbook on Multimodality in Human Interaction* (837-856). Berlin: De Gruyter Mouton.

- Holler, J. y Wilkin, K. (2011). An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. *Journal of Pragmatics*, 43 (14), 3522-3536.
- Hömke, P., Holler, J. y Levinson, S. C. (2017). Eye Blinking as Addressee Feedback in Face-To-Face Conversation. *Research on Language and Social Interaction*, 50(1), 54-70.
- Hömke, P., Levinson, S. C. y Holler, J. (2022). Eyebrow movements as signals of communicative problems in human face-to-face interaction. *PsyArXiv*, 10.31234/osf.io/3jnmt.
- Hutchins, E. y Nomura, S. (2011). Collaborative Construction of Multimodal Utterances, en J. Streeck, C. Goodwin y C. Le-Baron (Eds.), *Embodied Interaction: Language and the Body in the Material World* (29-43). Cambridge: Cambridge University Press.
- Inhoffen, N. (1996). El papel de los gestos en la ordenación y estructuración de la lengua hablada española, en Th. Kotschi, W. Oesterreicher y K. Zimmermann (Coords.), *El español hablado y la cultura oral en España e Hispanoamérica* (45-70) Frankfurt/Madrid: Vervuert/Iberoamericana.
- Instituto Cervantes (2006). *Plan curricular del Instituto Cervantes: Niveles de referencia para el español*, 3 volúmenes. Madrid: Instituto Cervantes-Biblioteca Nueva.
- Iriskhanova, O. K. y Cienki, A. (2018). The Semiotics of Gestures in Cognitive Linguistics: Contribution and Challenges. *Voprosy Kognitivnoy Lingvistiki*, 4, 25-36. DOI: 10.20916/1812-3228-2018-4-25-36
- Iverson, J. M. & S. Goldin-Meadow (1997). What's communication got to do with it? Gesture in children blind from birth. *Developmental Psychology*, 33 (3), 453-467.
- Iverson, J. M. y Goldin-Meadow, S. (2001). The resilience of gesture in talk: Gesture in blind speakers and listeners. *Developmental Science*, 4 (4), 416-422.
- Iwasaki, S. (2009). Initiating Interactive Turn Spaces in Japanese Conversation: Local Projection and Collaborative Action. *Discourse Processes*, 46(2-3), 226-246.
- Jaffe, J. & S. Feldstein (1970). *Rhythms of dialogue*. New York Academic Press. Series: Personality and psychopathology, 8.
- Jakobson, R. (1950). Linguistics and poetics, en T. A. Sebeok (ed.), *Style in Language*, Cambridge: MIT Press.
- Jakobson, R. (1972). Motor signs for 'Yes' and 'No'. *Language in Society*, 1(1), 91-96.
- Jashiashvili, K. (2017). *Comunicación no verbal y su aplicación a la enseñanza de español como lengua extranjera. Estudio comparativo de gestos españoles*, Tesis doctoral, Universidad de Cádiz.
- Jelec, A. y Jaworska, D. (2014). Thoughts on the table: Gesture as a tool for thinking in blind and visually impaired children. *Yearbook of the Poznan Linguistic Meeting*, 1 (1), 73-88.
- Jefferson, G. (1972). Side sequences, en Sudnow, D. (Ed.), *Studies in Social Interaction* (294-338). New York: The Free Press.
- Jefferson, G. (1973), A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences. *Semiotica*, 9, 47-96.

- Jefferson, G. (1984). On stepwise transition from talk about a trouble to inappropriately nextpositioned matters, en M. J. Atkinson y J. Heritage (Eds.), *Structures of Social Action: Studies in Conversation Analysis*, (191–222). Cambridge: Cambridge University Press.
- Jefferson, G. (Ed.)(1992). *Harvey Sacks' Lectures on Conversation*, Oxford: Blackwell.
- Jefferson, G. (2004). Glossary of transcript symbols with an Introduction, en G. H. Lerner (Ed.), *Conversation Analysis: Studies from the first generation*. Philadelphia: John Benjamins, 13-23.
- Jewitt, C., Bezemer, J. y O'Halloran, K. (2016). *Introducing Multimodality*. Routledge.
- Jiménez-Bravo, M. y Marrero, V. (2020). Multimodal Perception of Prominence in Spontaneous Speech: A Methodological Proposal using Mixed Models and AIC. *Speech Communication*, 124, 28-45.
- Jehoul, A., Brône, G. y Feyaerts, K. (2017). The shrug as marker of obviousness: Corpus evidence from Dutch face-to-face conversations. *Linguistics Vanguard*, 3(s1), 20160082. <https://doi.org/10.1515/lingvan-2016-0082>
- Jewitt, C., Bezemer, J. y O'Halloran, K. (2016). *Introducing Multimodality* (1ª ed.). Routledge. <https://doi.org/10.4324/9781315638027>
- Jokinen, K., Nishida, M., Yamamoto, S. (2009). Eye gaze experiments for conversation monitoring, en *Proceedings of the 3rd International Universal Communications Symposium (IUCS'09)*. ACM Press, New York.
- Jokinen, K. y Vanhasalo, M. (2009). Stand-up Gestures – Annotation for Communication Management, en *Proceedings of the NODALIDA 2009 Workshop Multimodal Communication: from Human Behaviour to Computational Models. Odense, Denmark, May 2009* (15-20).
- Jokinen, K., Furukawa, H., Nishida, M. y Yamamoto, S. (2013). Gaze and turn-taking behavior in casual conversational interactions. *ACM Transactions on Interactive Intelligent Systems*, 3 (2), artículo 12. <http://dx.doi.org/10.1145/2499474.2499481>
- Jørgensen, A. M (2007). COLA: un corpus oral de lenguaje adolescente. *Anejos de Oralía*, 3/1, 225-234.
- Kamunen, A. (2018). Open Hand Prone as a resource in multimodal claims to interruption: Stopping a co-participant's turn-at-talk. *Gesture*, 17(2), 291-321.
- Kamunen, A. (2019) How to Disengage: Suspension, Body Torque, and Repair. *Research on Language and Social Interaction*, 52(3), 1-21.
- Kelly, S. D., Barr, D., Breckinridge Church, R. y Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, 40, 577-592.
- Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica*, 26, 22-63.
- Kendon, A. (1970). Movement coordination in social interaction: Some examples described. *Acta Psychologica*, 32, 101-125.
- Kendon, A. (1972). Some relationships between body motion and speech. An analysis of an example, en A. Siegman y B. Pope (Eds.), *Studies in Dyadic Communication*

- (177-210). Elmsford, New York: Pergamon Press. doi: 10.1016/B978-0-08-015867-9.50013-7
- Kendon, A. (1973). The role of visible behavior in the organization of social interaction, en M. Cranach, e I. Vine (Eds.). *Social communication and movement*. Londres: Academic Press.
- Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance, en M. R. Key (Ed.), *The Relation between Verbal and Nonverbal Communication* (207-227). The Hague: Mouton.
- Kendon, A. (1981). Geography of gesture. *Semiotica*, 37, 129-163.
- Kendon, A. (1983). Gesture and Speech: How they interact, en J. M. Wiemann y R. Harrison (Eds.), *Sage Annual Reviews of Communication: Nonverbal Interaction*, 11 (13-46). Beverly Hills, CA: Sage Publications.
- Kendon, A. (1988). How gestures can become like words, en F. Poyatos (Ed.) *Crosscultural Perspectives in Nonverbal Communication* (131-141). Toronto, CJ: Hogrefe).
- Kendon, A. (1990). Movement coordination in social interaction: some examples described, en A. Kendon, *Conducting Interaction. Patterns of Behavior in Focused Encounters*. Cambridge: Cambridge University Press.
- Kendon, A. (1994). Do gestures communicate? A review. *Research on Language and Social Interaction*, 27(3), 175-200.
- Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of Pragmatics*, 23, 247-279. doi: 10.1016/0378-2166(94)00037-F
- Kendon, A. (2002). Some uses of the head shake. *Gesture*, 2 (2), 147-182. doi: 10.1075/gest.2.2.03ken
- Kendon, A. (2004). *Gesture. Visible action as utterance*. Cambridge: Cambridge University Press.
- Kendon, A. (2008). Some reflections on the relationship between 'gesture' and 'sign', *Gesture*, 8(3), 348-366.
- Kendon, A. (2013). Exploring the utterance roles of visible bodily action: A personal account, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (7-28). Berlin: De Gruyter Mouton.
- Kendon, A., (2017). Pragmatic functions of gestures: Some observations on the history of their study and their nature. *Gesture*, 16, 157-75.
- Kerbrat-Orecchioni, C. (1990), *Les interactions verbales. Tome I*, Armand Colin.
- Kerbrat-Orecchioni, C. (1992). *Les interactions verbales II*, Paris, Armand Colin.
- Kerbrat-Orecchioni, C. (1996). *La conversation*, Paris, Seuil.
- Kidwell, M. (1997). Demonstrating Reciprocity: Knowledge Displays as a Resource for the Unaddressed Participant. *Issues in Applied Linguistics*, 8 (2), 85-96.
- Kidwell, M. (2006), 'Calm down!': The role of gaze in the interactional management of hysteria by the police. *Discourse Studies*, 8(6), 745-770.

- Kidwell, M. (2013). Framing, grounding, and coordinating conversational interaction: Posture, gaze, facial expression, and movement in space. en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (100-112). Berlin: De Gruyter Mouton.
- Kimbara, I. (2006). On gestural mimicry. *Gesture*, 6, 39-61. doi: 10.1075/gest.6.1.03kim
- Kipp, M., Neff, M., Albrecht, I. (2007). An annotation scheme for conversational gestures: how to economically capture timing and form, en *Proceedings of the Workshop on "Multimodal Corpora" at LREC 2007*, 325–339 DOI: 10.1007/s10579-007-9053-5
- Kita, S. (2003). *Pointing where language, culture, and cognition meet*. Mahwah, NJ: Lawrence Erlbaum.
- Knapp, M.L. (1972). *Nonverbal Communication in Human Interaction*. New York: Holt, Rinehart y Winston.
- Knapp, M.L. (1980). *Essentials of nonverbal communication*. Nueva York, Holt, Rinehart & Winston.
- Kok, K., Bergmann, K., Cienki, A. y Kopp, S. (2016). Mapping out the multifunctionality of speakers' gestures. *Gesture* 15:1, 37-59. doi 10.1075/gest.15.1.02kok
- Kousidis, S., Pfeiffer, T., Malisz, Z., Wagner, P., Schlangen, D. (2012). Evaluating a minimally invasive laboratory architecture for recording multimodal conversational data, en *Proceedings of the Interdisciplinary Workshop on Feedback Behaviours in Dialogue* (39–42). Stevenson, WA, USA.
- Koutsombogera, M., Vogel, C. (2018). Modeling Collaborative Multimodal Behavior in Group Dialogues: The MULTISIMO Corpus, en *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. *European Language Resources Association (ELRA)*.
- Kress, G. (2009). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.
- Labov, W. (1983). *Modelos sociolingüísticos*. Cátedra.
- Ladewig, S. H. (2011). Putting the cyclic gesture on a cognitive basis. *CogniTextes*, 6. <http://cognitextes.revues.org/406>.
- Ladewig, S. H. (2012). *Syntactic and semantic integration of gestures into speech: Structural, cognitive, and conceptual aspects*. Tesis doctoral, European University Viadrina, Frankfurt (Oder).
- Ladewig, S. H. (2014a). Recurrent gestures, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 2: An International Handbook on Multimodality in Human Interaction* (1558–1575). Berlin: De Gruyter Mouton.
- Ladewig, S. H. (2014b). The cyclic gesture, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (1605-1618). Berlin: De Gruyter Mouton.
- Ladewig, S. y Bressemer, J. (2013). A linguistic perspective on the notation of gesture phases, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf

- (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (1060-1078). Berlin: De Gruyter Mouton.
- Ladewig, S. H. (2020). *Integrating Gestures. The Dimension of Multimodality in Cognitive Grammar*. Berlin, Boston: De Gruyter Mouton.
- Lakoff, R. T. (1973). The logic of politeness; or, minding your p's and q's. *Papers from the Regional Meeting, Chicago Linguistic Society, IX*, 292-305.
- Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.
- Leech, G. (1983). *Principles of pragmatics*, Londres, Longman.
- Lerner, G. H. (1992). Collectivities in action: Establishing the relevance of conjoined participation in conversation. *Text*, 13(2), 213-245.
- Lerner, G. H. (2002). Turn-Sharing: The Choral Co-Production of Talk-in-Interaction, en *The Language of Turn and Sequence* (223-256). OUP, Oxford.
- Lerner, G. H. (2003). Selecting next speaker: The context-sensitive operation of a context-free organization, *Language in Society*, 32, 177-201.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.
- Levy, E. T. y McNeill, D. (1992). Speech, gesture, and discourse, *Discourse Processes*, 15(3), 277-301.
- Li, Xiaoting (2014). *Multimodality, Interaction and Turn-taking in Mandarin Conversation*. John Benjamins. DOI: 10.1075/scld.3
- Llopis Cardona, A. (2016). Significado y funciones en los marcadores discursivos. *Verba: Anuario galego de filoloxia*, 43, 231-268.
- Llopis Cardona, A. (2020). Funciones, posición y unidades discursivas en *no sé* y *yo qué sé*, en Messias, Martí y Fuentes, *Aportaciones desde el español y el portugués a los marcadores discursivos* (249-271). Sevilla: Universidad de Sevilla.
- Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture*, 7(2), 179-214.
- López-Oziblo, R. (2016). Reflexiones sobre la producción de gestos en estudiantes de lenguas extranjeras. *Didáctica. Lengua y Literatura*, 28, 165-180.
- Lopez-Oziblo, R. (2019). *Marcando secuencias o disfluencias con gestos. Congreso de Pragmática y Discurso Oral*, Salamanca: Spain.
- Lyons, J. (1972). Human Language, en R. A. Hinde (Ed.), *Non-Verbal Communication*. Cambridge: Cambridge University Press.
- Malisz, Z., Włodarczak, M., Buschmeier, H. Skubisz, J., Kopp, S., Wagner, P. (2016). The ALICO corpus: analysing the active listener. *Lang Resources & Evaluation*, 50, 411-442. <https://doi.org/10.1007/s10579-016-9355-6>
- Mandelbaum, J. (1987). Couples sharing stories. *Communication Quarterly* 35(2), 144-170.
- Manero, E., Amigot, L., Olza, I. (2023). Fraseología del desacuerdo en un corpus multimodal de televisión: un estudio multinivel. *Círculo de Lingüística Aplicada a la Comunicación*, vol. sept.

- Martín Zorraquino, M.A. y Portolés, J. (1999), Los marcadores del discurso, en I. Bosque y V. Demonte (Dir.), *Gramática descriptiva de la Lengua española*. Vol. 3. (4051-4213). Madrid: Espasa.
- Martinell, E. y Forment Fernández, M. (1998). Aspectos semiótico-gestuales en algunas unidades fraseológicas del español, en Wotjak (Ed.), *Estudios de fraseología y fraseología del español actual* (225-239). Frankfurt a.M.: Vervuert.
- Martinell, E. y H. Ueda (Eds.) (1998). *Diccionario de gestos españoles*. Recuperado el 20 de junio de 2023 de <https://lecture.ecc.u-tokyo.ac.jp/~cueda/gakusyu/gestos/index.html>
- Matsumoto D. y Hwang H. C. (2013). Cultural Similarities and Differences in Emblematic Gestures. *Journal of Nonverbal Behaviour*, 37(1), 1-27.
- Maynard, S. K. (1989). Japanese conversation: Self-contextualization through structure and interactional management. Norwood, NJ: Ablex Publishing.
- Mchoul, A. (1978). The organization of turns at formal talk in the classroom. *Language in society*, 7, 183-213.
- McClave, E. Z. (1994). Gestural beats: The rhythm hypothesis. *Journal of Psycholinguistic Research*, 23(1), 45-66.
- McClave, E. Z. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, 32, 855-878.
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, 92, 350-371. doi: 10.1037/0033-295X.92.3.350
- McNeill, D. (1992). *Hand and Mind. What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press.
- McNeill, D. (2000). Introduction, D. McNeill (Ed.), *Language and Gesture* (1-10). Cambridge: Cambridge University Press.
- McNeill, D. (2005). *Gesture and Thought*. Chicago: University of Chicago Press.
- McNeill, D. (2014). Gesture – speech unity: Phylogenesis, ontogenesis, and microgenesis. *Language, Interaction and Acquisition*, 5 (2), 137-184. doi: 10.1075/lia.5.2.01mcn
- McNeill, D. y Levy, E. (1982). Conceptual representations in language activity and gesture, en R. J. Jarvella y W. Klein (Eds.), *Speech, Place and Action: Studies in Deixis and Related Topics*, 271-295. Chichester: Wiley.
- McNeill, D. y Levy, E. (1993). Cohesion and gesture. *Discourse Processes*, 16(4), 363-386, DOI: 10.1080/01638539309544845
- McNeill, D., Cassell, J. y Levy, E. (1993). Abstract deixis. *Semiotica* 95(1-2), 5-20. <https://doi.org/10.1515/semi.1993.95.1-2.5>
- Melinger, A. y W.J.M. Levelt (2004). Gesture and the communicative intention of the speaker. *Gesture* 4(2), 119-141.
- Meo-Zilio, G. y Mejía, S. (1980). *Diccionario de Gestos: España e Hispanoamérica*. Bogotá: Instituto Caro y Cuervo.
- Moeschler (1985). *Argumentation et conversation. Éléments pour une analyse pragmatique du discours*. Genève: Hatier-Crédif.

- Mondada, L. (2006). Participants' online analysis and multimodal practices: projecting the end of the turn and the closing of the sequence. *Discourse Studies*, 8, 117-129.
- Mondada, L. (2007). Multimodal resources for turn-taking: pointing and the emergence of possible next speakers. *Discourse Studies*, 9(2), 194-225.
- Mondada, L. (2009). Video recording practices and the reflexive constitution of the interactional order: some systematic uses of the split-screen technique. *Human Studies*, 32, 1, 67-99.
- Mondada, L. (2012). The conversation analytic approach to data collection, en J. Sidnell, y T. Stivers (Eds.). *The Handbook of Conversation Analysis* (32-56). Wiley-Blackwell.
- Mondada, L. (2013a). Conversation analysis: Talk and bodily resources for the organization of social interaction, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (218-226). Berlin: De Gruyter Mouton.
- Mondada, L. (2013b). Multimodal interaction, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (577-589). Berlin: De Gruyter Mouton.
- Mondada, L. (2013c). Video as a tool in the social sciences, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (982-991). Berlin: De Gruyter Mouton.
- Mondada, L. (2014). Pointing, talk, and the bodies. Reference and joint attention as embodied interactional achievements, en M. Seyfeddinipur y M. Gullberg (Eds.) *Conversation to Visible Action as Utterance: Essays in honor of Adam Kendon*. John Benjamins Publishing Company.
- Mondada, L. (2015). Multimodal completions, en A. Deppermann y S. Günthner (Eds.), *Temporality in Interaction*. Benjamins, Amsterdam.
- Mondada, L. (2018) Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality. *Research on Language and Social Interaction*, 51(1), 85-106.
- Monterubbianesi, M. G. (2013) La comunicación no verbal en el aula: una propuesta de clasificación y análisis de los gestos para las clases de E/LE. *MarcoELE: Revista de Didáctica Español Lengua Extranjera*, 16.
- Montes, R. G. (1994). Relaciones entre expresiones verbales y no verbales en la organización del discurso, en *Estudios de Lingüística Aplicada*, 19/20, 253-272.
- Montes Miró, R. G. (2002). "Haciendo a un lado": gestos de desconfirmación en el habla mexicana. *Iztapalapa* 53, 248-267.
- Montes, R. G. (2009). Autorreferencialidad y autoseñalamiento, en *Actas Virtuales del IV Coloquio de Investigadores en Estudios del Discurso*, 219-227.
- Moore, E., y Llompart, J. (2017). Recoger, transcribir, analizar y presentar datos interaccionales plurilingües, en E. Moore y M. Dooly (Eds.), *Enfoques cualitativos*

para la investigación en educación plurilingüe (418-433).
<https://doi.org/10.14705/rpnet.2017.emmd2016.639>

- Morel, M. A. (2010a). Structure coénonciative du texte oral dialogué: intonation, syntaxe, regard et geste, en S.L. Florea, C. Papahagi, L. Pop, A. Curea (Eds.), *Directions actuelles en linguistique du texte [Actes du Colloque international de Cluj 2008]* II 9–22. Cluj-Napoca: Casa Cartii de Stiinta.
- Morel, M. A. (2010b). Déflexivité et décondensation dans le dialogue oral en français: marqueurs grammaticaux, intonation, regard et geste, en D. Bottineau y L. Beggioni (Dir.) *La déflexivité* [Langages 178], 115–131. Paris: Larousse/Armand Colin.
- Morel, M. A. y Rialland A. (1992). Emboîtement, autonomies, ruptures dans l'intonation française, *Travaux Linguistiques du CERLICO*, 5.
- Morel, M. A. y Danon-Boileau, L. (1998). *Grammaire de l'intonation. L'exemple du français*. Paris: Ophrys.
- Morel, M. A. y Vladimirska, E. (2014). Intonation and gesture in the segmentation of speech units. The discursive marker vraiment: integration, focalisation, formulation, en Pons Bordería, S. (ed), *Discourse Segmentation in Romance Languages* (185-218). Amsterdam / Philadelphia: John Benjamins.
- Moreno García, C. y Tuts, M. (2006). Mírame que te estoy hablando, en *Actas XVII Congreso Internacional de ASELE* (1193-1209). Publicaciones Universidad de la Rioja.
- Morris, D. (1994). *Bodytalk, A world Guide to Gestures*. London: Jonathan Cape.
- Morris, D., Collett, P., Marsh, P. y O'Shaughnessy, M. (1979). *Gestures. Their Origins and Distributions*. New York: Stein and Day.
- Müller, C. (1998). *Redebegleitende Gesten. Kulturgeschichte–Theorie–Sprachvergleich*. Berlin: Arno Spitz.
- Müller, C. (2004). Forms and uses of the Palm Up Open Hand: A case of a gesture family?, en Cornelia Müller & Roland Posner (Eds.), *The semantics and pragmatics of everyday gestures* (233-256). Berlin: Weidler Buchverlag.
- Muller, C. (2008). *Metaphors dead and alive, sleeping and waking: A dynamic view*. Chicago & London: University of Chicago Press. doi: 10.7208/chicago/9780226548265.001.0001
- Muller, C. (2009). Gesture and Language, en K. Malmkjaer (Ed.), *Routledge's Linguistics Encyclopedia* (214-217). Abington/New York: Routledge.
- Muller, C. (2010a). Mimesis und Gestik, en G. Koch, C. Voss y M. Vohler (Eds.), *Die Mimesis und ihre Künste* (149-187). Munchen: Wilhelm Fink.
- Muller, C. (2010b). Wie Gesten bedeuten. Eine kognitiv-linguistische und sequenzanalytische Perspektive, en *Sprache und Literatur*, 41(1): 37-68.
- Müller, C. (2013). Gestures as a medium of expression: The linguistic potential of gestures, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tesselndorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (202-217). Berlin: De Gruyter Mouton.
- Müller, C. (2017). How recurrent gestures mean. Conventionalized contexts-of-use and embodied motivation. *Gesture* 16(2), 276-303. doi 10.1075/gest.16.2.05mul

- Müller, C. (2018). Gesture and Sign: Cataclysmic Break or Dynamic Relations?. *Frontiers in Psychology* 9:1651. doi: 10.3389/fpsyg.2018.01651
- Müller, C. (en prensa). A Toolbox of Methods for Gesture Analysis, en A. Cienki (ed.) *Handbook of Gesture Studies*. Cambridge: Cambridge University Press.
- Müller, C. (en preparación). *Gesture and Language*.
- Müller, C. y Bohle, U. (2007). Das Fundament fokussierter Interaktion. Zur Vorbereitung und Herstellung von Interaktionsräumen durch körperliche Koordination, en R. Schmitt (ed.), *Koordination. Analysen zur multimodalen Kommunikation* (129-165). Tübingen: Gunter Narr.
- Müller, C., Bressemer, J., & Ladewig, S. H. (2013). Towards a grammar of gesture: A formbased view, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (707–733). Berlin: De Gruyter Mouton.
- Müller, C., Ladewig, S. H., & Bressemer, J. (2013). Gesture and speech from a linguistic point of view, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (55-81). Berlin: De Gruyter Mouton.
- Muñoz Trujillo, M. y Guerrero González, S. (2019). Gesticulación y funciones pragmáticas en narraciones conversacionales. *Oralia*, 22 (1), 103-126.
- Nascimento Dominique, N. (2005). Comunicación no verbal: algunas diferencias gestuales entre España y Brasil, *Linred: lingüística en la red*, 15/03/2005.
- Nascimento Dominique, N. (2012). *La comunicación sin palabras. Estudio comparativo de gestos usados en España y Brasil*. Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá.
- Navarretta, C., Ahlsen, E., Allwood J. Paggio, P. (2011). Creating Comparable Multimodal Corpora for Nordic Languages, en B. Sandford Pedersen, G. Nespore e I. Skadina (Eds.) *NODALIDA 2011 Conference Proceedings* (153-160).
- Navarretta, C., Paggio, P. (2013). Multimodal Behaviour and Interlocutor Identification in Political Debates, en I. Poggi, F. D’Errico, L. Vincze y A. Vinciarelli (Eds.), *Multimodal Communication in Political Speech. Shaping Minds and Social Action*, Vol. 7688, (99-113). Springer Science+Business Media.
- Navarretta, C. y Paggio, P. (2017). The Danish NOMCO corpus: multimodal interaction in first acquaintance conversations. *Lang Resources & Evaluation*, 51, 463-494.
- Narbona, A. (1988). Sintaxis coloquial: problemas y métodos. *LEA*, 10(1), 81-106.
- Nota, N., Trujillo, J. P., y Holler, J. (2022). Conversational eyebrow frowns facilitate question identification: An online VR study. *PsyArXiv*, 10.31234/osf.io/fcj8b. doi:10.31234/osf.io/fcj8b
- Oertel, C., Włodarczak, M., Edlund, J., Wagner, P., Gustafson, J. (2012) Gaze patterns in turn-taking. *Proc. Interspeech*, 2246-2249. DOI: 10.21437/Interspeech.2012-132
- Olza, I. (2022). Patrones multimodales de (des)alineación conversacional, en V. Pérez Béjar y M. Méndez Orense (Eds.), *Perspectivas integradas para el análisis de la oralidad*. Sevilla: Universidad de Sevilla.

- Oreström, B. (1983). *Turn-taking in English Conversation*. Lund, iberFörlang Lund.
- Owen, M. (1981). Conversation units and the use of *well...*, en P. Werth (Ed.), *Conversation and Discourse* (99-116). New York: St. Martin's Press.
- Özyürek, A. (2000). The influence of addressee location on spatial language and representational gestures of direction, en David McNeill (Ed.), *Language and Gesture* (64-83). Cambridge: Cambridge University Press.
- Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language*, 46, 688-704.
- Padilla, X. (2004). Del oyente receptor al oyente combatiente. *ELUA*, 18, 213-230.
- Padilla, X. (2020). El papel del mecanismo de reacción en la definición de las unidades conversacionales. Su contribución a la caracterización del dinamismo discursivo, en A. Cadedo y A. Hidalgo (Eds.) *Pragmática del español hablado. Hacia nuevos horizontes*, Universitat de València.
- Pagán, C., Valenzuela, J., Alcaraz, D., Olza, I. y Ramscar, M. (2020). Quantifying the speech-gesture relation with massive multimodal datasets: Informativity in time expressions, *PLoS ONE*, 15(6), e0233892.
- Paggio, P., Navarretta C. (2017). The Danish NOMCO corpus: multimodal interaction in first acquaintance conversations. *Lang Resources & Evaluation*, 51, 463-494. DOI 10.1007/s10579-016-9371-6
- Parrill, F. y Sweetser, E. (2004). What we mean by meaning: Conceptual integration in gesture analysis and transcription. *Gesture*, 4(2), 197-214.
- Pascual Aliaga, E. (2018): Análisis prosódico de las estructuras truncadas en la conversación coloquial española: funciones de formulación y atenuación, en A. García Ramón y M. A. Soler Bonafont (Eds.), *ELUA. Estudios de Lingüística*, Anexo IV (57-84).
- Payrató, Lluís (1993). A pragmatic view on autonomous gestures: A first repertoire of Catalan emblems. *Journal of Pragmatics*, 20(3), 193-216.
- Payrató, Ll. (1994). Emblema: quan el gest ho és tot. *Revista d'etnologia de Catalunya*, 4, 20-31.
- Payrató, Ll. (2003). What does 'the same gesture' mean? Emblematic gestures from some cognitive-linguistic theories, en M. Rector, I. Poggi y N. Trigo (Eds.): *Gestures. Meaning and Use*, 73-81. Porto: Universidade Fernando Pessoa.
- Payrató, Ll. (2004). Notes on pragmatic and social aspects of everyday gestures, en Müller, C. y Posner, R. (Eds.), *The Semantics and Pragmatics of Everyday Gestures*, 103-113. Berlin: Weidler.
- Payrató, Ll. (2009). Non-verbal communication, en J. Verschueren y J. O. Östman (Eds.), *Key notions for pragmatics* (163-194). Amsterdam & Philadelphia: John Benjamins.
- Payrató, Ll. (2014). Emblems or quotable gestures: Structures, categories, and functions, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 2: An International Handbook on Multimodality in Human Interaction* (1474-1481). Berlin: De Gruyter Mouton.

- Payrató, Ll. (2016). Multimodalidad y lengua oral: enseñando las fronteras del lenguaje, en *Textos de didáctica de la lengua y la literatura*. Barcelona, Año XXII (73), 43-49.
- Payrató, Ll. (2019). Pragmatics and mulimodality. A reflection on multimodal pragmastylistics, en R. Giora, M. Haugh (Coords.) *Doing pragmatics interculturally: cognitive, philosophical, and sociopragmatic perspectives*, 293-312.
- Payrató, Ll. y Clemente, I. (2020). *Gestures We Live by. The Pragmatics of Emblematic Gestures*. Boston / Berlin: De Gruyter Mouton.
- Payrató, L. y Teßendorf, S. (2014). Pragmatic gestures, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tesselndorf (Eds.), *Body - Language - Communication. Volume 2: An International Handbook on Multimodality in Human Interaction (1531-1539)*. Berlin: De Gruyter Mouton.
- Peiró-Marquez, L. e Ibarretxe-Atuñano, I. (2021) ¿Cómo incide la gestualidad en la codificación del movimiento?. *e-AESLA*, 7.
- Peräkylä, A. y Ruusuvuori, J. (2006). Facial expression in an assessment, en H. Knoblauch, B. Schnetzler, J. Raab y H.G. Soeffner (Eds.), *Video analysis: methodology and methods: qualitative audiovisual data analysis in sociology*. Peter Lang, Frankfurt am Main, 127-142.
- Ruusuvuori, J. y Peräkylä, A. (2009). Facial and verbal expressions in assessing stories and topics. *Research on Language and Social Interaction*, 42 (4), 377-394.
- Pfänder Pfänder, S. *Freiburg SofaTalks*. Recuperado el 20 de junio de 2023 de: http://moca.phil2.uni-freiburg.de/moca3_v3/index.php?vi=14
- Philips, S. (1976). Some sources of cultural variability in the regulation of talk. *Language in society*, 5, 81-95.
- Pike, K. L. y Pike, E. (1983). *Text and Tagmeme*. Londres, Frances Pinter.
- Poggi I., Magno Caldognetto E. (1996). A Score for the analysis of gesture in multimodal communication, en L. Messing (Ed.), *Proceedings of the Workshop on the Integration of Gesture in Speech (Applied Science and Engineering Labs., Newark and Wilmington, Delaware USA, October 7-8, 1996)*, (235-244).
- Poggi, I (2002). Symbolic gestures. The case of the Italian gestinary. *Gesture*, 2(1), 71-98.
- Poggi, I. (2004). The Italian gestinary. Meaning representation, ambiguity, and context., en C. Müller y R. Posner (Eds.), *The Semantics and Pragmatics of Everyday Gestures*, 73-88. Berlin: Weidler.
- Poggi, I (2013). Mind, hands, face, and body: a sketch of a goal and belief view of multimodal communication, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tesselndorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction (627-647)*. Berlin: De Gruyter Mouton.
- Pomerantz, A. (1984). Agreeing and disagreeing with assessments: some features found in preferred/ dispreferred turn shapes, en J.M. Atkinson y J. Heritage (Ed.) *Structures of Social Action (57-101)*. Cambridge: Cambridge University Press.

- Pons Bordería, S. (2014). Models of discourse segmentation in Romance languages: An overview, en Pons Bordería, S. (Ed.), *Discourse Segmentation in Romance Languages*. Amsterdam / Philadelphia: John Benjamins, 1-22.
- Pons Bordería, S. (2016). Cómo dividir una conversación en actos y subactos, en A. M. Bañón Hernández, M.^a M. Espejo Muriel, B. Herrero Muñoz-Cobo y J. L. López Cruces (Eds.), *Oralidad y análisis del discurso. Homenaje a Luis Cortés Rodríguez* (545-566). Almería: Editorial Universidad de Almería.
- Pons Bordería, S. (2018). The combination of discourse markers: Keys to untying a Gordian knot, en M. Borreguero Zuloaga, V. Atayan y S. Große (Eds.), *Models of Discourse Units in Romance Languages* (121-158), *Revue Romane* 53(1).
- Pons Bordería, S. (dir.) (2023): *Corpus Val.Es.Co 3.0*. <http://www.valesco.es>
- Pons Bordería, S. (2022). *Creación y análisis de corpus orales: saberes prácticos y reflexiones teóricas*. Berlin, Germany: Peter Lang Verlag.
- Pons, S. y Fischer, K. (2021). Using Discourse Segmentation to Account for the Polyfunctionality of Discourse Markers: The Case of Well. *Journal of Pragmatics*, 173, 101-118.
- Pons Bordería, S., Pardo Llibrer, A., Alemany Martínez, A. (en prensa): *La marcación discursiva en español: descripción y análisis estadístico desde el DPDE*. Editorial de la Universidad de Sevilla.
- Poyatos, F. (1970). Kinésica del español actual. *Hispania*, 53, 444-452.
- Poyatos, F. (1972). Paralenguaje y kinésica del personaje novelesco: nueva perspectiva en el análisis de la narración. *Revista de Occidente*, 113/114, 148-170; *Prohemio*, 3(2), 191-307.
- Poyatos, F. (1981). Gesture inventories: Fieldwork methodology and problems, en Adam Kendon (Ed.), *Nonverbal communication, interaction, and gesture. Selections from Semiotica*, (371-400). The Hague: Mouton. doi: 10.1515/9783110880021.371
- Poyatos, F. (1987) Nonverbal communication in simultaneous and consecutive interpretation: A theoretical model and new perspectives. *TextContext*, 2 (3), 73-108.
- Poyatos, F. (1992). Nonverbal communication in foreign-language teaching and learning: A theoretical and methodological approach, en A. Helbo (Ed.), *Evaluation and Language Teaching* (115-143). Berlin/Nueva York: Peter Lang Verlag.
- Poyatos, F. (1994). *Comunicación no verbal I. Cultura, lenguaje y conversación*. Istmo: Madrid.
- Poyatos, F. (1996). La lengua hablada como realidad verbal-no verbal: nuevas perspectivas, en A. Briz (Coord.), *Pragmática y gramática del español hablado. Actas del II Simposio de Español Hablado* (215-224). Valencia: Libros Pórtico
- Poyatos, F. (2003). La comunicación no verbal: Algunas de sus perspectivas de estudio e investigación. *Revista de Investigación Lingüística*, 2 (6), 67-83.
- Poyatos, F. (2018). Los estudios de comunicación no verbal como rama interdisciplinar de la lingüística, *Linred*, 16.
- Prieto, P. y Borrás-Comes, S.T. (2013). Prosody and gesture constrain the interpretation of double negation. *Lingua*, 131, 136-150.

- Prillwitz, S., Leven, R., Zienert, H., Hanke, T. y Henning, J. (1989). *HamNoSys Version 2.0 Hamburger Notationssystem für Gebärdensprachen: Eine Einführung*. Hamburg: Signum Verlag.
- Provine, R. R. (2004). Laughing, tickling, and the evolution of speech and self. *Current Directions in Psychological Science*, 13(6), 215–218. <http://doi.org/10.1111/j.0963-7214.2004.00311.x>
- Quintero García, D. (2005). Reflexiones acerca de la enseñanza de signos kinésicos, proxémicos y cronémicos en E/LE para estudiantes japoneses. *Cuadernos Canela* 17, 109-121.
- Rasenberg, M. Özyürek, A., Dingemanse, M. (2020). Alignment in Multimodal Interaction: An Integrative Framework. *Cognitive Science A Multidisciplinary Journal*, 44(11).
- Raymond, C. W. y Olguín, L. M. (2022). *Análisis de la Conversación: fundamentos, metodología y alcances*, Routledge. <https://doi.org/10.4324/9780429507274>
- Rauscher, J. (1986). Wann und mit welchen Gründen wird Proxemik zum Gegenstand der Semiotik, en K. D. Dutz, P. Schmitter (Eds.), *Geschichte und Geschichtsschreibung der Semiotik*. Fallstudien, (439-452). Münster: MAkS-Publikationen.
- Ricci Bitti, P. E. (2014). Facial expression and social interaction, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 2: An International Handbook on Multimodality in Human Interaction* (1342-1348). Berlin: De Gruyter Mouton.
- Roberts, S., Torreira, F., Levinson, S. C. (2015). The effects of processing and sequence organization on the timing of turn taking: a corpus study. *Frontiers in Psychology*, 6.
- Rodríguez González, M. A. (2003). Lenguaje de signos, Alicante: Biblioteca Virtual Miguel de Cervantes. <https://www.cervantesvirtual.com/nd/ark:/59851/bmc08633>
- Rodríguez Rosique, S. y J. L. Cifuentes Honrubia (Coords.): Corpus GestINF. Universidad de Alicante. Biblioteca Virtual Miguel de Cervantes. Recuperado el 20 de junio de 2023 de: https://www.cervantesvirtual.com/portales/proyecto_gestinf/corpus/
- Rodríguez Rosique, S. (2021). Creencia, activación y comunidades culturales: GestINF, o la pertinencia de un corpus basado en el conocimiento compartido. *ORALIA*, 24(1), 151-186.
- Rohrer, P. L., Vilà-Giménez, I., Florit-Pons, J., Esteve-Gibert, N., Ren, A., Shattuck-Hufnagel, S., & Prieto, P. (2020). The MultiModal MultiDimensional (M3D) labelling scheme for the annotation of audiovisual corpora. *Proceedings of the 7th Gesture and Speech in Interaction (GESPIN)*. KTH Speech, Music & Hearing and Språkbanken Tal. Stockholm, Sweden.
- Rossano, F. (2012). Gaze in Conversation, en J. Sidnell y T. Stivers (Eds.), *The Handbook of Conversation Analysis* (308-329). Chichester: John Wiley y Sons, Ltd.
- Rossini, N. (2012). *Reinterpreting gesture as language: Language "in action"*. IOS Press.
- Roulet et al. (1985). *L'articulation du discours en français contemporain*. Peter Lang.
- Roulet et al. (2001). *Un modèle et un instrument d'analyse de l'organisation du discours*. Peter Lang.

- Ruesch, R., Kees, W. (1956). *Nonverbal Communication*. University of California Press.
- Ruiz Fajardo, G. (2016). *Columbia Corpus of Spanish Conversations (Corpus de conversaciones para E/LE)*. Columbia University. Recuperado el 30 de enero de 2020 en: <https://edblogs.columbia.edu/corpusdeconversaciones/>
- Sacks, H. (1984), Notes on methodology. En Heritage, J., Atkinson K. M. (Eds.), *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press, 2–27.
- Sacks, H., Schegloff, E.A. y Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50 (4), 696-735.
- Sacks, H. y Schegloff, E. A. (2002). Home position, *Gesture*, 2(2), 133-146.
- Salameh Jiménez, S., Estellés Arguedas, M., Pons Bordería, S. (2018). Beyond the notion of periphery: An account of polyfunctional discourse markers within the Val.Es.Co. model of discourse segmentation, en Beeching, K., Ghezzi, C. y Molinelli, P. (Eds.), *Positioning the Self and Others: Linguistic perspectives*, 105-125.
- Sánchez Benítez, G. (2009). La comunicación no verbal, en *II Jornadas de formación de profesores de E/LE: estrategias de enseñanza y aprendizaje de español en China*, MarcoE/LE, 8, 1-6.
- Saussure, F. de (1916). *Curso de Lingüística General*. Madrid: Alianza, 1983. Edición de Tullio de Mauro.
- Schefflen, A. E. (1966). Natural history method in psychotherapy, en L. Gottschalk y A. Auerbach (Eds.), *Methods of research in psychotherapy* (263-86). New York: Appleton-Century-Crofts.
- Schefflen, A. E. (1973). *Communicational structure*. Bloomington: University of Indiana Press.
- Schegloff, E. A. (1972). Sequencing In Conversational Openings. In J. Fishman (Ed.), *Volume 2 Selected Studies and Applications* (91-125). Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110880434-006>
- Schegloff, E. A. (1979). Identification and recognition in telephone conversation openings, en G. Psathas, *Everyday Language: Studies in Ethnomethodology* (23-78). New York: Irvington Publishers.
- Schegloff, E.A. (1982). Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences, en D. Tannen (Ed.), *Analyzing Discourse: Text and Talk* (71-93). Georgetown University Press.
- Schegloff, E.A. (1984). On some gestures' relation to talk, en J. M. Atkinson y J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (266–296). Cambridge & Paris: Cambridge University Press & Éditions de la Maison des Sciences de l'Homme.
- Schegloff, E.A. (1987). Analyzing single episodes of interaction: An exercise in conversation analysis. *Social Psychology Quarterly*, 50, 101-114.
- Schegloff, E.A. (1991) Reflections on talk and social structure, en D. Boden y D. H. Zimmerman (Eds.), *Talk and Social Structure* (44-70). University of California.
- Schegloff, E.A. (1998). Body Torque, *Social Research*, 65 (3), 535-596.

- Schegloff, E. A. (1996). Turn organization: One intersection of grammar and interaction. en E. Ochs , E. A. Schegloff y S. A. Thompson (Eds.), *Interaction and grammar*, (52-133). Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CBO9780511620874.002>
- Schegloff, E.A. (2007). *Sequence Organization in Interaction. A Primer in Conversation Analysis*. Vol. 1. Cambridge: Cambridge University Press.
- Schegloff, E.A. y Sacks, H. (1973). Opening up closings. *Semiotica*, 8(4), 289-327.
- Scherer, K. P. (1977). Affektlaute und vokale Embleme, en R. Posner, H.P. Reinecke (eds.), *Zekenprozesse*. 199-214. Wiesbaden: Athenaion.
- Schmitt, R. (2005). Zur multimodalen Struktur von turn-taking, *Online-Zeitschrift zur verbalen Interaktion*, 6, 17-61.
- Schmidt, S. (2013). Proxémica y comunicación intercultural: la comunicación no verbal en la enseñanza de E/LE. Tesis doctoral. Universitat Autònoma de Barcelona. Facultat de Filosofia i Lletres.
- Schönherr, B. (2014), Categories and functions of posture, gaze, face, and body movements, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 2: An International Handbook on Multimodality in Human Interaction* (1333-1341). Berlin: De Gruyter Mouton.
- Scott, S. K., Lavan, N., Chen S. y McGettigan C. (2014). The social life of laughter. *Trends in Cognitive Sciences*, 18(12), 618–620.
<http://doi.org/10.1016/j.tics.2014.09.002>.
- Searle, John (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139173438
- Searle, J. R. (1972). *Les actes de langage*, Paris, Hermann.
- Searle, J. R. (1979). *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CBO9780511609213>
- Selting, M. (2005). Syntax and prosody as methods for the construction and identification of turn-constructive units in conversation, en A. Hakulinen y M. Selting (Eds.), *Syntax and Lexis in Conversation: Studies on the use of linguistic resources in talk-in-interaction*, (17-44). John Benjamins Publishing Company.
- Selting, M. (2013). 38. Verbal, vocal, and visual practices in conversational interaction. en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tessendorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (589-609). Berlin: De Gruyter Mouton.
- Selting, M., Auer, P., Barth-Weingarten, D. Bergmann, J., Bergmann, P. et al. (2011). A system for transcribing talk-in-interaction: GAT 2 translated and adapted for English. *Gesprächsforschung*, 12, 1-51.
- Shannon, E., y Weaver, W. (1949): *The Mathematical Theory of Communication*. Urbana, University of Illinois Press.
- Sherzer, J. (1973). *Verbal and Nonverbal Deixis: The Pointed Lip Gesture among the San Blas Cuna*. University of Texas, Austin Institute of Latin American Studies.

- Sherzer, J. (1991), The Brazilian Thumbs-Up Gesture, *Journal of Linguistic Anthropology*, 1(2), 189-197.
- Sidnell, J. (2006). Coordinating gesture, talk, and gaze in reenactments, *Research on Language and Social Interaction*, 39(4), 377-409.
- Sinclair, J. y Coulthard, M. (1975). *Towards an Analysis of Discourse: the English Used by Teachers and Pupils*. London: Oxford University Press.
- Sinclair, J. y Coulthard, M. (1992). Towards an analysis of discourse, en M. Coulthard (Ed.) *Advances in Spoken Discourse Analysis* (14-60). Routledge.
- Slama-Cazacu, T. (1976). Nonverbal components in message sequence: “Mixed syntax”, en W. C. McCormack y S. A. Wurm (Eds.), *Language and man: Anthropological issues* (217–227). The Hague: Mouton.
- Beck Nielsen, S. (2020). Silence gestures revisited. *Gesture* 19(2/3), 246-268. DOI: 10.1075/gest.20022.bec
- Sperber, D. y Wilson, D. (1986). *Relevance. Communication and Cognition*. Oxford, Blackwell.
- Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction*, 41(1), 31-57.
- Stivers, T. (2021). Is Conversation Built for Two? The Partitioning of Social Interaction. *Research on Language and Social Interaction*, 54(1), 1-19, DOI: 10.1080/08351813.2020.1864158
- Stivers, T. y Rossano, F. (2010a). A Scalar View of Response Relevance. *Research on Language in Social Interaction*, 43(1), 49-56.
- Stivers, T. y Rossano, F. (2010b). Mobilizing Response. *Research on Language y Social Interaction*, 43(1). 3–31.
- Stivers, T y Sidnell, J. (2005). Introduction: Multimodal interaction. *Semiotica* 156(1/4): 1–20.
- Stivers, T., Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoyman, G., Rossano, F., de Ruiter, J.P. Yoon, K. y Levinson, S.C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587-92.
- Stokoe, W. C. (1960). Sign Language Structure: An Outline of the Visual Communication System of the American Deaf. *Studies in Linguistics, Occasional Papers*, 8. Buffalo, NY: University of Buffalo.
- Streeck, J. (1993) Gesture as communication I: Its coordination with gaze and speech. *Communication Monographs*, 60, 275–299.
- Streeck, Jürgen (1994). ‘Speech-handling’: *The metaphorical representation of speech in gestures. A cross-cultural study*. Unpublished manuscript, Austin, TX.
- Streeck, J. (2002). Grammars, words, and embodied meanings: On the uses and evolution of so and like. *Journal of Communication*, 52(3): 581-596.
- Streeck, J. (2005). Pragmatic aspects of gesture, en J. Mey (Ed.), *Encyclopedia of Language and Linguistics*, 5 (71-76). Oxford: Elsevier.

- Streeck, J. (2009a). *Gesturecraft: The manufacture of meaning*. Amsterdam y Philadelphia: John Benjamins. doi: 10.1075/gs.2
- Streeck, J. (2009b). Forward-gesturing. *Discourse Processes*, 46(2), 161-179.
- Streeck, J. (2013). Interaction and the living body. *Journal of Pragmatics*, 46(1), 69-90.
- Streeck, J. (2014). Mutual gaze and recognition. Revisiting Kendon's "Gaze direction in two-person conversation", en M. Seyfeddinipur y M. Gullberg (Eds.) *Conversation to Visible Action as Utterance: Essays in honor of Adam Kendon*. John Benjamins Publishing Company.
- Streeck, J. (2017). *Self-making man. A day of action, life, and language*. Cambridge & New York: Cambridge University Press.
- Streeck, J. (2021). The emancipation of gestures. *Interactional Linguistics*, 1:1, 90-122.
- Streeck, J., y Hartge, U. (1992). Previews: Gestures at the transition place, en P. Auer y A. di Luzio (Eds.), *The contextualization of language* (138-158). Amsterdam: Benjamins.
- Stubbs, M. (1983). *Discourse Analysis*. Oxford: Basil Blackwell.
- Stubbs, M. (1987). *Análisis del discurso. Análisis sociolingüístico del lenguaje natural*. Alianza Editorial.
- Szatrowski, P. (2000). Relation between gaze, head nodding and aizuti "back channel" at a Japanese company meeting. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, 26 (283-294). Berkeley, CA: Berkeley Linguistics Society.
- Szatrowski, P. (2003). Gaze, head nodding, and *aizuchi* in information presenting activities, en P. M. Clancy (Ed.), *Japanese/Korean Linguistics*, 11, 119-132. Stanford, CA: CSLI.
- Tabacaru, S. (2019). *A Multimodal Study of Sarcasm in Interactional Humor*. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110629446>
- Tabensky, A. (2004) La repetición del gesto en el discurso interactivo. *Signos* 37, 55, 41-58.
- Talmy, L. (1985). Lexicalization Patterns: Semantic Structure in Lexical Forms, en T. Shopen (Ed.), *Language Typology and Syntactic Description. Vol. III.: Grammatical Categories and the Lexicon* (57-149). Cambridge: Cambridge University Press.
- Talmy, L. (1991). Path to Realization: A Typology of Event Conflation, en *Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society* (480-519). Berkeley: University of California Press.
- Ten Have, P. (1999). *Doing Conversation Analysis. A practical guide*. Introducing Qualitative Methods Series. London: Sage Publications.
- Teßendorf, S. (2013). Emblems, quotable gestures, or conventionalized body movements, en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tesselendorf (Eds.), *Body - Language - Communication. Volume 1: An International Handbook on Multimodality in Human Interaction* (82-100). Berlin: De Gruyter Mouton.
- Teßendorf, S. (2014). Pragmatic and metaphoric – combining functional with cognitive approaches in the analysis of the "brushing aside gesture", en C. Müller, E. Fricke, A. Cienki, D. McNeill, S. Ladewig, S. Tesselendorf (Eds.), *Body - Language -*

- Communication. Volume 2: An International Handbook on Multimodality in Human Interaction* (1540-1558). Berlin: De Gruyter Mouton.
- Tiittula, L. (2007). Blickorganisation in der side-by-side-Positionierung am Beispiel eines Geschäftsgesprächs, en Reinhold Schmitt (Ed.), *Koordination. Analysen zur multimodalen Kommunikation* (225-261). Tübingen: Gunter Narr.
- Torreira, F. & Ernestus, M. (2012). Weakening of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish. *Phonetica*, 69, 124-148. DOI: 10.1159/000 343635
- Trager, G. L. (1958). Paralanguage: a first approximation, *Studies in Linguistics*, 13 (1), 1-12.
- Traugott, E. C. (1988). Pragmatic strengthening and grammaticalization, en *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society* (406–416). Berkeley, CA: Berkeley Linguistics Society.
- Ueda, H. (1998). Semántica de los gestos españoles. *Lingüística Hispánica*, 20. España: Universidad de Estudios Extranjeros de Kansai.
- Urbanik, P. y Svennevig, J. (2021). Action-Depicting Gestures and Morphosyntax: The Function of Gestures Speech Alignment in the Conversational Turn. *Frontiers in Psychology*, 12:689292.
- Vajrabhaya, P. y Pederson, E. (2018). Teasing apart listener-sensitivity. The role of interaction. *Gesture* 17(1), 65-97. doi.org/10.1075/gest.00011.vaj
- van Son R.J.J.H., Wesseling, W., Sanders, E., van den Heuvel, H. (2008). The IFADV corpus: A free dialog video corpus. *International Conference on Language Resources and Evaluation, LREC 2008, 26 May - 1 June 2008*, Marrakech, Morocco.
- Vertegaal, R., Slagter, R., Van Deer Veer G. y Nijholt, A. (2001). Eye gaze patterns in conversations: There is more the conversational agents than meets the eyes, en *Proceedings of the ACM CHI Human Factors in Computing Systems Conference (CHI'01)*.
- Wharton, T. (2009). *Pragmatics and non-verbal communication*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511635649
- Vladimirska, E. (2008). Vraiment: Identité sémantique et variations discursives, en *Estudos Linguísticos / Linguistic Studies* 2 (275–286). Lisboa: Colibri. Faculdade de Letras de Lisboa/CLUNL.
- Wehling, E. (2017). Discourse management gestures. *Gesture*, 16(2), 245-276.
- Wells, B., & Macfarlane, S. (1998). Prosody as an Interactional Resource: Turn-projection and Overlap. *Language and Speech*, 41(3-4), 265-294.
- Gaviño Rodríguez, V., Centeno de Guirotane, M., Salieta Olvera, P., Camarena Molina, I. (2011). *Diccionario de gestos españoles*.
<https://www.coloquial.es/es/diccionario-de-gestos-espanoles/>
- Wiener, M., Devoe, S., Rubinow, S. y Geller, J. (1972). Nonverbal behavior and nonverbal communication. *Psychological Review*, 79(3), 185-214. <https://doi.org/10.1037/h0032710>
- Wilson, T.P., J.M. Wiemann y D.H. Zimmerman (1984). Models of turn taking in conversation interaction. *Journal of Language and Social Psychology*, 3, 159-183

- Włodarczak, M., Buschmeier, H., Malisz, Z., Kopp, S., Wagner, P. (2012). Listener head gestures and verbal feedback expressions in a distraction task, en *Proceedings of the Interdisciplinary Workshop on Feedback Behaviours in Dialogue* (93-96). Stevenson, WA, USA.
- Wundt, W. (1973 [1900]). *The Language of Gestures*. The Hague: Mouton.
- vom Lehn, D., Heath, C., Hindmarsh, J. (2005). Examining exhibits: interaction in museums and galleries. *Communication and Cognition*, 38(3/4), 229-247.
- Yagüe (2004). ELAO y E/LE. Hablando por los codos: enseñar gestos en las clases de español. *redE/LE*, 1.
- Yániz Ciriza, J. (2022). *Acuerdo y desacuerdo: un análisis multimodal*. Trabajo fin de máster, Universidad de Murcia.
- Yngve, V. (1970). On getting a word in edgewise. *Papers from the 6th regional meeting of the Chicago Linguistic Society*, 567-78.

Anexo I

	Acceso	Lengua	Nº de conversaciones	Duración de cada conversación	Duración total	Nº de hablantes por conversación	Lugar de grabación	Equipo de grabación	Género. Temática libre o guiada.	Transcripción	Secreta	
corpus NOMCO (Multimodal Corpus Analysis in the Nordic Countries) (Navarreta et al. 2011).	Requiere solicitud	Sueco	39	8-10 min.	5 horas	2	Laboratorio		Encuentro entre desconocidos	Oral ortográfica y anotación de gestos MUMIN	No	
		Danés	12	5 min.	1 hora	2	Laboratorio		Encuentro entre desconocidos	Oral ortográfica y anotación de gestos MUMIN	No	
		Sueco (subcorpus of the Gothenburg Spoken Language Corpus (GSLC) (Allwood et al., 2000))	82		122 horas	Entre 2 y 12				discusiones concertadas y naturales, reuniones formales e informales y cenas	Oral ortografica (Gothenburg Transcription Standard, Modified Standard Orthography) y anotación de gestos MUMIN	No
		estonio	2	30 min.	1 hora	3				rol play	Oral ortografica y anotación de gestos MUMIN	No
		finlandés					4			Conversación durante juego de cartas. Libre.	Oral ortografica y anotación de gestos MUMIN	No

	Acceso	Lengua	Nº de conversaciones	Duración de cada conversación	Duración total	Nº de hablantes por conversación	Lugar de grabación	Equipo de grabación	Género. Temática libre o guiada.	Transcripción	Secreta
corpus IFA Dialog Video (van Son et al. 2008)	Libre	Holandés	23	15 min.	6 horas	2	Laboratorio	2 cámaras, micrófonos de cabeza		Oral ortográfica y anotación de miradas	No
corpus DF2F	Requiere solicitud	Holandés	10	1 hora	10 horas	2	Laboratorio	3 cámaras, micrófonos de cabeza, eye-tracker (POR FASES)	Conversación. Libre.		No
El corpus CID (Corpus d'interactions dialogales/Corpus of Interactional Data) (Bertrand et al. 2008)	Libre	Francés	8	1 hora	8 horas	2 (miembros del laboratorio o estudiantes de doctorado)	Laboratorio	1 cámara, micrófonos de cabeza	En una parte, tema sugerido, pero no obligado; en otra parte, tareas orientadas.	Anotación de gestos adaptada de Kipp (2006) y el esquema de codificación MUMIN (Allwood et al., 2005).	No
corpus CLAPI (Corpus de Langue Parlée en Interaction)	Libre	Francés	69	variable, de minutos a horas	67 horas	variable	Espacios cotidianos	1 o 2 cámaras	Diferentes contextos: profesional, privado, institucional, comercial, médico, aula, etc.	Oral	No
Active Listening Corpus (ALICO)	Requiere solicitud	alemán	50	6:36 min. de media	5 horas 31 minutos	2	Laboratorio	3 cámaras, 2 micrófonos de cabeza, 1 micrófono para la escena	Conversación orientada: roles de hablante y de oyente predefinidos	Anotación de gestos de cabeza y feedback (Malisz et al. 2016)	No

	Acceso	Lengua	Nº de conversaciones	Duración de cada conversación	Duración total	Nº de hablantes por conversación	Lugar de grabación	Equipo de grabación	Género. Temática libre o guiada.	Transcripción	Secreta
AliLeGra (Feyaerts et al. 2012) de KU Leuven y la Karl-Franzens-Universität de Graz	Requiere solicitud	holandés	21	20 minutos	8 horas	2	Laboratorio	3 cámaras, biosensores	Conversación. Grabación secreta previa a falsa prueba		Sí
Insight Interaction corpus (KU Leuven)	Requiere solicitud	holandés	15	25 minutos	6 horas 15 minutos	2	Laboratorio	head-mounted scene cameras and eye-trackers	datos controlados experimentalmente y conversaciones espontáneas	Oral y anotación de la mirada y los gestos. Mirada monitorizada por eye-tracker	No
MULTISIMO (Koutsombogera y Vogel 2018)	Requiere solicitud	inglés	18		3 horas	2 (+ 1 facilitador humano)	Laboratorio	4 cámaras	Tarea: dos participantes colaboran entre sí para resolver un cuestionario	Transcripción de habla, la estructura del diálogo, los comentarios del facilitador, anotación del tipo de risa y anotación parcial de gestos.	No

Anexo II



VNIVERSITAT
D VALÈNCIA

AUTORIZACIÓN PARA EMPLEAR LA GRABACIÓN Y LA TRANSCRIPCIÓN DEL MATERIAL CON FINES INVESTIGADORES EN LINGÜÍSTICA

A. Autorización previa a la grabación

Dña./D. _____ con documento de identificación o
pasaporte número _____

DECLARO

- 1) que se me ha informado de que voy a ser grabado/a de forma secreta en las próximas semanas;
- 2) que, posteriormente a la grabación, podré ver y escuchar el contenido de mi grabación;
- 3) que, en caso de no estar de acuerdo, puedo ejercer mi derecho a retirar la grabación.

A los efectos oportunos, firmo la presente autorización en _____
a _____ de 20____.

Fdo. _____

B. Autorización posterior a la grabación

Dña./D. _____ con documento de identificación o
pasaporte número _____

DECLARO

- 1) que se me ha informado de que he sido grabado/a secretamente y he visto y escuchado el contenido de mi grabación;
- 2) que se me ha informado de que puedo ejercer mi derecho a retirar la grabación.

Y, por tanto, AUTORIZO al uso de la grabación de su contenido para fines estrictamente de investigación.

A los efectos oportunos, firmo la presente autorización, en _____ a _____ de
20____.

Fdo. _____

1. Datos personales

Los datos personales obtenidos mediante el presente formulario se incorporarán a los sistemas de información de la Universitat de València – Estudi General (links.uv.es/lopd/dpo) en el marco del Proyecto FFI2016-77842-P, Unidades discursivas para una descripción sistemática de los marcadores del discurso en español (UDEMADIS), financiado por el MINECO, la AEI y los fondos FEDER, dirigido por el doctor Salvador Pons Bordería.

La información objeto de tratamiento será utilizada para el desarrollo de funciones docentes y académicas propias de la Universitat de València como la investigación, la creación, desarrollo, transmisión y crítica de la ciencia, de la técnica y de la cultura y la difusión, la valorización y la transferencia del conocimiento.

En concreto, estas grabaciones formarán parte del corpus de la tesis de Sandra Cabanes Pérez. Dicho corpus está compuesto por un conjunto de conversaciones coloquiales obtenidas en las provincias de Alicante y Valencia. Con el objetivo de analizar estas muestras de habla, se recogen conversaciones informales espontáneas (coloquiales) reales grabadas en lugares cotidianos para los hablantes, en una situación de familiaridad o amistad para su posterior análisis lingüístico.

La Universitat de València se compromete a que cualquier divulgación pública de los resultados obtenidos con motivo de la investigación se realizará anonimizando debidamente los datos utilizados, de modo que los sujetos de la investigación no resultarán identificados o identificables.

La base jurídica del tratamiento es el consentimiento del afectado/a y se prevé la conservación de los datos personales durante cinco años. Transcurrido ese periodo, los datos se conservarán debidamente disociados para garantizar el anonimato.

2. Registro de imagen o sonido

En el marco del desarrollo de la actividad se obtendrán registros de audio y vídeo. Ud. Autoriza a la Universitat de València al uso, edición, difusión y explotación de estos registros exclusivamente para fines de investigación. En caso de utilización, se asegurará que el afectado/a nunca sea identificado por su nombre ni mediante información verbal que le haga identificable.

Todo ello con la única salvedad y limitación de aquellas utilidades o aplicaciones que pudieran atentar a los derechos garantizados en la Ley Orgánica 1/1982, de 5 de mayo, de Protección Civil al Derecho al Honor, la Intimidad Personal y familiar y a la Propia Imagen, así como del pleno respeto de las previsiones específicas del art. 4 de la Ley Orgánica 1/1996, de 15 de enero, de protección jurídica del menor.

3. Publicación

Los resultados del proyecto son susceptibles de publicación. En caso de tal utilización, se asegurará que Ud. nunca sea identificado/a por su nombre, apellidos, ni mediante información verbal que le haga identificable.

4. Ejercicio de derechos

Las autorizaciones concedidas en este documento podrán ser revocadas mediante la presentación del oportuno escrito. La revocación comportará la retirada de la información de los sistemas de la Universitat de València en un plazo prudencial de tiempo en función de la disponibilidad de recursos.

Puede obtener más información acerca de sus derechos en: links.uv.es/lopd/derechos

Y en prueba de conformidad, firmo el presente documento en el lugar y la fecha indicados en el encabezamiento.

Nombre y apellidos	Nombre y apellidos Firma PADRE / MADRE / TUTOR <i>Rellenar solo en caso de menores o personas con incapacidad legal.</i>
--------------------	--

Anexo III

Instrucciones para ganchos:

1. Los hablantes que participen en la conversación, hablen más o hablen menos, deben haber firmado la primera autorización para ser grabados al menos una semana antes de la fecha de grabación. La grabación debe ser secreta. No les dé detalles sobre el momento o la situación en la que se realizará ni la cámara que utilizará, tan solo que se les requiere una grabación en vídeo para un estudio sobre situaciones cotidianas.
2. Avíselos en el momento de la firma de que no vuelvan a mencionar nada sobre la grabación hasta el momento de la segunda firma, de esta manera se evitará que durante la grabación hagan comentarios sobre la posible grabación o elucubren si están siendo grabados en ese momento.
3. La grabación de menores de edad es ilegal, por lo que no pueden participar en la conversación. Si fortuitamente aparece un menor en la imagen en segundo plano y no se mueve, será posible ocultarlo en la edición. Si el menor habla en una conversación ajena y se distinguen sus palabras la conversación queda invalidada.
4. Si fortuitamente participa un hablante que no había firmado la autorización, la conversación queda invalidada. Sin embargo, es válida si la aparición de esa persona es breve; por ejemplo, un camarero que les atiende en un bar o un conocido que pasa a saludar y se va en un minuto.
5. Coloque la cámara enfocando hacia el lugar en el que va a tener lugar la conversación sin que los demás hablantes sean conscientes. La mejor opción es colocarla antes de que entren a la habitación. Si usted llega después, coloque la cámara con disimulo, dando la espalda y fingiendo una tarea (como mirar el móvil).
6. En la grabación debe verse la cara de todos los hablantes, bien de frente, bien ligeramente de perfil, pero no pueden situarse de espaldas a la cámara. En este sentido, a veces colocar la cámara después de conocer la disposición de los hablantes es mejor. Tenga en cuenta que a mayor distancia el plano es más amplio. Si cree que va a ser imposible que se vea a todos los hablantes, desista; pruebe de nuevo en su próximo encuentro.
7. Se busca que cada conversación dure una media de 20 minutos, (10 y 30 también es posible). Igualmente, no es necesario ajustar la duración del encuentro a la grabación, puede dejar la cámara grabando hasta que se agote la batería y recogerla con disimulo más tarde.
8. Hay que evitar música, televisión o ruido de ambiente, pero sin que se descubra la grabación por procurar las condiciones. Si no es posible, debe procurarse que el volumen del ruido no sea tan alto como para que la cámara no capte las palabras de los hablantes. Es posible utilizar la grabadora del móvil como herramienta secundaria, después se alinearán ambos audios.
9. La conversación debe ser libre, no semidirigida. Es decir, una grabación se tiene que descartar si el gancho orienta la marcha de la conversación proponiendo temas de conversación, interpelando a los oyentes para que hablen, etc. Hay que tener en cuenta que en ocasiones suceden silencios en la conversación (por ejemplo, una cena entre dos informantes con la televisión puesta), pero no hay que ponerse nervioso por ello. Es preferible dejar que la conversación marche con naturalidad y usted mismo olvide que está siendo grabado. Será la investigadora la que decida qué material es válido o no.
10. No puede haber cortes en la grabación. Por ejemplo, no se puede grabar cinco minutos, pausar la grabación y reanudarla después.

11. No son válidas las conversaciones que no tengan un fin interpersonal. Es decir, una reunión de trabajo, una lección o una compra-venta no son válidas porque tienen un fin transaccional. Lo dicho no quita que pueda existir algún diálogo breve en el que haya un desvío de estos parámetros (por ejemplo: pedir algo al camarero en una terraza).
12. Si uno de los hablantes descubre que están siendo grabados, la grabación queda invalidada. Informe a la investigadora para decidir si se podrá repetir la grabación en un futuro o no.
13. Terminada la grabación, no informe a los hablantes todavía. Es posible que la grabación no sea válida y se tenga que repetir otro día. Envíe el material a la investigadora para que decida si es apta.
14. En el caso de que la grabación sea apta, se les dará la oportunidad a los hablantes de ver el vídeo antes de la segunda firma.
15. En total, serán 3 firmas por hablante (incluido el gancho): una para autorizar ser grabados, otra para ceder el vídeo concreto y otra para ceder sus datos personales.

