

**THE INFLUENCE OF CAREGIVER INTERACTION ON CHILDREN'S GESTURE
AND SPEECH EXPRESSION: A CASE STUDY**

**A INFLUÊNCIA DA INTERAÇÃO DO CUIDADOR NA EXPRESSÃO DE GESTO E
FALA DA CRIANÇA: UM ESTUDO DE CASO**

**LA INFLUENCIA DE LA INTERACCIÓN DEL CUIDADOR EN LA EXPRESIÓN DE
GESTOS Y HABLA DEL NIÑO: UN ESTUDIO DE CASO**



Milena Magalhães GOMES¹
e-mail: milenamagalhaes.adv@gmail.com



Tatiana Ramalho BARBOSA²
e-mail: tatiramalho@hotmail.com

How to reference this paper:

GOMES, M. M.; BARBOSA, T. R. The influence of caregiver interaction on children's gesture and speech expression: a case study. **Rev. EntreLínguas**, Araraquara, v. 11, n. 00, e025006, 2025. e-ISSN: 2447-3529. DOI: 10.29051/el.v11i00.20143



| Submitted: 03/02/2025
| Revisions required: 06/03/2025
| Approved: 04/04/2025
| Published: 10/08/2025

Editor: Prof. Dr. Rosangela Sanches da Silveira Gileno
Deputy Executive Editor: Prof. Dr. José Anderson Santos Cruz

¹ Master's student in Linguistics, PROLING/UFPB, João Pessoa – PB – Brazil.

² PhD student in Linguistics, PROLING/UFPB, João Pessoa – PB – Brazil.

ABSTRACT: Children's language encompasses an inseparable mix of gestures, looks, verbal expressions, and vocal production. In this context, the question is how the interactional environment contributes to language acquisition development. This study aims to analyze joint attention scenes between father and daughter to identify the caregiver's role in the development of a child's gestures and speech. The methodology is based on the construction of the Multimodal Envelope, which corresponds to the look-gesture-vocal production set, through which a video is analyzed in which the father encourages his daughter's storytelling. As a result, we obtained an overview of how the child's speech and gesture support is strengthened by the father's actions, exposing the daughter to interaction contexts in which she is the main interlocutor. It is expected that this work can contribute to future investigations on the importance of the caregiver in scenes of joint attention during the process of language acquisition.

KEYWORDS: Joint attention. Cognitive development. Multimodality of language.

RESUMO: A linguagem infantil abrange uma mistura indissociável de gestos, olhares, expressões faciais e produção vocal. Nesse contexto, questiona-se como o meio interacional contribui para o desenvolvimento da aquisição da linguagem. Assim, este trabalho objetiva analisar cenas de atenção conjunta entre pai e filha, para identificar o papel do cuidador no desenvolvimento de gestos e falas da criança. A metodologia utilizada baseia-se na elaboração do Envelope Multimodal, correspondendo ao conjunto olhar-gesto-produção vocal, através do qual se analisa um vídeo em que o pai estimula a contação de história pela filha. Como resultado, obtivemos um panorama de como o suporte de falas e gestos da criança é fortalecido pela atuação do pai que expõe a filha a contextos de interação em que ela ocupa o papel de principal interlocutora. Espera-se que este trabalho contribua para investigações futuras sobre a importância do cuidador nas cenas de atenção conjunta durante o processo de aquisição da linguagem.

PALAVRAS-CHAVE: Atenção conjunta. Desenvolvimento cognitivo. Multimodalidade da língua.

RESUMEN: El lenguaje infantil abarca una mezcla indisoluble de gestos, miradas, expresiones faciales y producción vocal. En este contexto, se plantea la pregunta de cómo el entorno interactivo contribuye al desarrollo de la adquisición del lenguaje. Así, este trabajo tiene como objetivo analizar escenas de atención conjunta entre padre e hija, para identificar el papel del cuidador en el desarrollo de los gestos y el habla del niño. La metodología utilizada se basa en la elaboración del Sobre Multimodal, que corresponde al conjunto mirada-gesto-producción vocal, a través del cual se analiza un video en el que el padre estimula la narración de historias por parte de la hija. Como resultado, obtenemos un panorama de cómo el apoyo a las palabras y gestos de la niña se ve fortalecido por la actuación del padre, quien expone a la hija a contextos de interacción en los que ella ocupa el papel de principal interlocutora. Se espera que este trabajo contribuya a investigaciones futuras sobre la importancia del cuidador en las escenas de atención conjunta durante el proceso de adquisición del lenguaje.

PALABRAS CLAVE: Atención conjunta. Desarrollo cognitivo. Multimodalidad del lenguaje.

Introduction

Human beings have a biologically inherited ability to live within a culture, but biological inheritance does not excuse the context of the individual's interaction with the environment, which is crucial for the formation of cognition and all the characteristics of the species (Tomasello, 2019). In other words, the ability to engage and participate culturally is not just a product of the environment or society, but is rooted in human biology. For example, the ability to learn a language, play chess, or participate in social or religious rituals are skills that emerge naturally in humans due to our biological constitution in interaction with the environment.

Thus, in order to understand the process of human language acquisition and production, it is necessary to observe it as a product of cognition, through scenes of joint attention, through the convergence of language acquisition and cognitive development activities.

In this context, this research aims to analyze scenes of joint attention between father and daughter, in order to identify the role of the caregiver in the development of the child's gestures and speech. The methodology used is based on the creation of a Multimodal Envelope, which corresponds to the gaze-gesture-vocal production set, through which a 2 min 37 s video was analyzed, in which the caregiver (father) stimulates the storytelling being conducted by his daughter, aged approximately 36 months, who directs her speech to an object, a teddy bear.

To what extent is child language improved with the support of speech and gestures from an attentive caregiver, who stimulates the imagination and promotes the development of language acquisition? And what is the contribution of the interactional environment in which the child is inserted? This research seeks to answer these questions by means of a case study, supported by theoretical contributions on joint attention and language acquisition.

Starting from a sociointeractionist perspective, from Vygotsky (1978), through Langacker (2000) and Tomasello (2024), the aim of this study was to demonstrate how children get involved in events of joint intentionality (Tomasello, 2024), which include moments of joint attention, from which they begin to understand other people as intentional agents, just like themselves.

In the video analyzed for this study, we can see the deliberate intention of the caregiver to promote the development of his daughter's language, narration and argumentation, precisely because he is aware that she is still in the critical period of language acquisition and, therefore, her interaction with the environment plays a crucial role.

In order to structure the discussion proposed in this article in detail, we have divided the content into four parts, including this introductory section. The next topic presents the theoretical framework that guided the study. The methodology used and the discussion of the results are described below. Finally, the final considerations are presented, with the findings generated by the research.

Theoretical foundation

Historical and ontogenetic processes of human language acquisition

With regard to the historical processes and cognitive formation of linguistic competence, Tomasello (2024) provides evidence in his study that our species passes ontogenetically through two stages: collaboration and cultural aspects. The author reports that he started from research begun by Vygotsky (1978) with children and went on to experiment with chimpanzees to conclude that these two aspects are uniquely human: collaboration, since children engage in collaborative activities from an early age; and cultural artifacts, which influence how children learn norms and begin to interact in group activities.

In this direction, Sapir (1980) also points to speech as an acquired and cultural function, different from other activities, such as walking, which, for the author, is an organic and instinctive function. According to Sapir (1980, p. 12, our translation), “speaking is a historical inheritance of the group, the product of prolonged social use,” even if people do not realize the cognitive effort that goes into developing language.

It is through the processes of collaboration and culture that the hypothesis of use-based acquisition, defended by authors such as Tomasello (2019) and Langacker (2000), is established. According to this theory, knowledge of the various cognitive domains (such as playing chess, counting, dressing, etc.) is acquired through use, *i.e.*, through interaction with the environment. This includes language, in the sociointeractionist view. It is worth noting that, although they describe aspects of use and social interaction in detail, authors such as Tomasello (2019, 2024) and Vygotsky (1978, 2005) do not rule out the biological component or the cognitive mechanisms present in humans, considering them to be “the result of some type or mode of cultural transmission unique to the [human] species” (Tomasello, 2019, p. 5, our

translation). In other words, the sociointeractionist view also embraces biological characteristics, but assumes them as part of the process.

In this way, according to Tomasello (2019, p. 7, emphasis added, our translation), human learning happens because each individual recognizes others as co-specific, that is, “as beings *equal to him*, with mental and intentional lives equal to his.” Therefore, when children begin to perceive other people as “like them,” they are able to imagine themselves “‘in the mental shoes’ of someone else, so that they not only learn *from* the other, but *through* the other.” Thus, what differentiates human beings from non-human primates, and what has been responsible for the cultural evolution we have experienced to this day, is precisely the ability to pass on to other members of the species countless achievements in pursuit of a common good.

This human characteristic is called by Tomasello (2019, 2024) the “ratchet effect,” a process of historical cumulative evolution, transmitted by the social and in a reliable way, which happens through learning by imitation. Another point mentioned by the author is the active instruction of individuals in a given society, who pass on their experiences in such a way that it is possible to see the accumulation of knowledge acquired over the generations.

Therefore, the physical and mental structure of our ancestors, along with the ways in which we perceive, behave and think, are inherited biologically. However, these characteristics only manifest and develop in interaction with the environment around us. This study advocates a theory of language acquisition based on use, according to Tomasello (2019), which encourages the observation of speech production and comprehension to explain the sociocognitive bases of acquisition, based on empirical evidence of speech between children and adults.

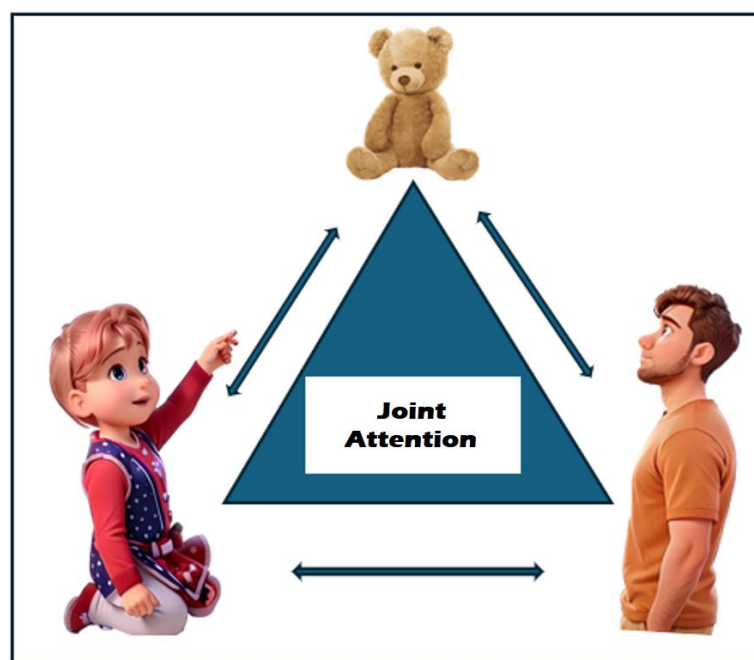
For Tomasello (2019), learning takes place in three basic ways: by imitation, by instruction and by collaboration. We know that these three gestures are of crucial importance in the development of cognition and especially language in children. It is in this sense that the author advocates joint attention, a topic that we will deal with later.

Joint attention and language acquisition

Scenes of joint attention are components of the cultural environment, in which interaction favors the recognition of conspecifics and the interpretation of their intentions. These are concrete instances of interaction between the child and the caregiver, which act as the central gear to promote the process of language acquisition. They revolve around the sharing of attention, context, linguistic and conceptual systems, intentions and intersubjectivity, involving comprehension operations between the child and the adult, considering the cognitive operations necessary for effective communication.

Joint attention occurs between the perceptual world and the linguistic world as one of the primary sociocognitive foundations of language acquisition, forming part of the socially shared reality and favoring the intersubjective context in which the perception of the symbol and the interpretation of the intentions of those involved occur. “The term *joint attention* is often used to characterize this whole complex of skills and social interactions” (Tomasello, 2019, p. 85-86, our translation). It emerges between nine and twelve months, when the baby exhibits triadic behavior, which involves interaction with objects and people, resulting in the triangle: child, adult, object or event, as illustrated in Figure 1.

Figure 1 – Joint attention



Source: Own elaboration based on Tomasello (2019).

The illustration deals with the model of interaction that occurs at this stage in the development of human babies (between nine and twelve months), called “joint attention behaviors,” following the gestures of the caregiver and, in particular, pointing to the things to which they refer. This shift from dyadic (between two entities) to triadic (three entities) behavior is exactly what differentiates humans from other primates (Tomasello, 2019).

In this sense, the dialogic-discursive perspective, which is based on previous research by authors such as Vygotsky (2005) and Bruner (2004, 2007), endorses the point of view of interaction as an important component of the child’s language acquisition process, highlighting the importance of the caregiver and the presence of the play instrument in this interactional process. According to Del Ré, Hilario and Vieira (2021), the dialogic-discursive approach is understood as the study of language development (and cognition) from the point of view of the interaction between the child and the other.

This approach analyzes children’s speech through the interaction between them and their interlocutor, taking into account the different aspects that interfere in this relationship, such as: “discursive and situational contexts, dialogism (constitutive of all discourse), empirical interaction relations between subjects, the constitution of subjects in discourse and the role they assume during communication, etc.” (Del Ré; Hilario; Vieira, 2021, p. 13, our translation). Triadic behavior, therefore, intrinsic to human beings, is affected by the scenario of language use.

Thus, interaction is seen as the intersection of the dialogic and multimodal perspectives of language, because social interaction is the factor that “provides an environment for the child to enter language, and the dialogic perspective points to the child’s relationship with himself and with the world, that is, the construction of the idea of the other as his interlocutor, which is only possible because language is multimodal” (Holanda *et al.*, 2023, p. 5, our translation).

With regard to the stages of development, Tomasello (2019) describes the period of emergence of joint attention as the nine-month revolution. From then on, the scenes of interaction become increasingly complex, as a result of understanding the symbolization that events, objects and people acquire, as well as the discovery and use of language constructions that children learn, first by imitation, but then with their own intentionality.

For Tomasello (2019, p. 94-95, our translation), “intentional agents are ordinary beings who have goals and make active choices among the behavioral means available to achieve those

goals, which includes actively choosing what to pay attention to in pursuit of those goals.” In other words, intentionality acts as a catalyst for the development of speaking skills.

Table 1 describes the author's conceptual categorization of the scenes of children's experiences.

Table 1 – Conceptual analysis and categorization of scenes of experience

Approximate Age	Experiential Scene	Language
9 months	Joint attention scenes (non-symbolized)	–
14 months	Symbolized scenes (undifferentiated symbolization)	Holophrases
18 months	Separated scenes (differentiation of events and participants)	Pivot-type constructions
22 months	Syntactic scenes (symbolic marking of participants)	Isolated verb constructions
36 months	Categorized scenes (generalized symbolic marking of participant roles)	General verb constructions

Source: Tomasello (2019, p. 212).

Thus, there is a complex evolution to the acquisition of a natural language. And for the child's development to be complete, it is essential to live in the environment and interact with other individuals. In this way, scenes of joint attention, such as those described and analyzed in this study, are essential for understanding symbols, grasping the structures of the language, for perspective-taking and for constructing one's own intentional discourse.

The role of gestures in communication

For McNeill (2016, p. 4), gestures are the “intrinsic image of language.” In other words, they are movements used intentionally which, when accompanied by orality, enhance the meaning contained in the speech. In this way, gesture and speech function as a unified system, in which gestures are a component of speech and thought, since language is necessarily linked to the image (McNeill, 2005).

The main definitions of gestures were addressed by Kendon (1980), who went on to study their functions within contexts of interaction. This typology became known as what McNeill (1992) called “Kendon’s continuum” and considers five types: gesticulations³, filler gestures, emblems, pantomimes, and sign languages.

According to Silva and Faria (2022, p. 203, our translation), “gesticulation is a movement that carries a meaning related to the accompanying utterance and encompasses many variants and uses.” In other words, they are spontaneous movements, usually of the hands and arms, which are in line with spoken language.

The filling gestures occupy a gap in the utterance. They are similar to the meaning of gesticulation, but in filler gestures the speaker interrupts their speech and inserts a gesture, which will give meaning to the utterance. McNeill (1992, p. 37) gives the following example to explain this type of gesture: “the parents were fine, but the children were [gesture],” in which the gesture occupies the space of the word “bad.”

Emblems are gestures that follow a pattern and have a culturally established meaning. They are like a social code and do not need speech to produce meaning (McNeill, 1992). The classic example of this type of gesture is the “OK” sign with the hand; or the “thumbs up” sign, if we consider the Brazilian context.

Pantomimes are representations that tell a story without the presence of speech. Silva and Faria (2022, p. 205, our translation) describe the following as examples of pantomimes: “pretending to be in pain or pretending to be happy.” In turn, sign languages are gestures that represent the lexicon of a linguistic system of a given community, for example, the Brazilian Sign Language (Libras) and other linguistic codes from different countries (Silva; Faria, 2022, p. 205).

In a study carried out by McNeill and Levy (1982, *apud* McNeill, 1992), gestures—specifically gesticulations and filler gestures—are further categorized according to the following dimensions: iconic, metaphorical, deictic and rhythmic.

According to McNeill (2005), movements that reproduce some action, or some concrete object or entity, are considered iconic. Similar to iconic gestures, metaphorical gestures also represent something; however, they can signify any entity with abstract content, such as an idea or a memory. Silva and Faria (2022, p. 207-208, our translation) exemplify this category with “the configuration of raised hands with the palm upwards, as if the speaker wanted to

³ Considering the translation of speech-framed gesture in Cavalcanti (2018).

demonstrate through this gesture that he did not understand anything about the lesson he was attending.” The shoulders can also be raised to emphasize the speaker’s intention of doubt.

Deictics are pointing gestures, with the function of indicating objects or events, which can be concrete or abstract. McNeill (1992) points out that it is common in contexts of interaction for the action of pointing to indicate a present object; however, they can indicate something that is not physically visible.

Finally, rhythmic gestures are so called because they work like the beats of a musical rhythm. These are movements without semantic content, short and quick, with the hands or fingers, up and down, or back and forth. Its function is to mark a part of the sentence that is to be emphasized within the general context (McNeill, 1992).

Based on the concepts presented about the ontogenesis of language acquisition, as well as joint attention and the various ways of categorizing gestures in different contexts of human interaction, the following topic sets out the methodology for analyzing the video with the dialogical scene between father and daughter.

Methodology and Discussion of Results

The data is taken from a recording made using a cell phone, during a father-daughter course interaction, without the child realizing they were being filmed. This is a female child, approximately 36 months old, interacting with her father, in which he holds the teddy bear and interacts with his daughter as if it were the toy he was talking about. The video lasts 2 min 37 s, and is presented here in a segmented form, in three parts, in which we describe the temporal clippings of speech and description of the actions of both interaction partners in a storytelling.

In the following communicative situations, we look at the scenes that have been described separately, in order to analyze the gestures that the child makes while telling a story. In this sense, not only speech, but also gestures (and even looks) are part of what Nóbrega (2010, p. 17, our translation) calls the “compositional planes of a linguistic multimodality envelope.” The following are scenes of joint attention in which three components of interaction are taken into account: gaze, gesture and vocal production. The Multimodal Envelope is considered the set of these components, as shown in Table 2.

Table 2 – Multimodal Envelope: part 1

Multimodal envelope		
Composition plans:	FATHER:	CHILD:
Look	00:01 The father sits behind the child while holding two teddy bears in his hands, facing the child. He looks at the teddy bears and strikes up a conversation.	00:01 The child responds by looking at both teddy bears.
		00:02 The child sits on the father's lap, with his back to him and facing the teddy bears the father is holding.
Gestures		00:01 Puts his right hand on his chest as he introduces himself (deictic).
	00:02 The father holds two teddy bears in his hands, facing the child, staging a conversation with the child. He gestures by swinging both teddy bears during his speech (gesturing).	
Vocal production		00:01 // My name is Isabela//
	00:02 // Isabela, tell me how the world was created. // How was the first day? // How was it? //	
		00:09 // On the first day, God created man from the dust of the earth. //
	00:13 //That was on the sixth day. // But what happened after He created man? // What happened? //	
		00:17 // He created from the dust of the earth. //
	00: 19 // Then what? //	
	00: 20 // Hmm //	00:20 // He blew in his nose // And... and... // Do you know what his name was? //
	00:27 // What was it? //	
		00:28 // Adam //
	00:29 // Ah! // What a beautiful name! // What else? //	
	00:32 //Ah! //	00:32 // God created a woman. // What... what... what... // Do you know what the woman's name was? //
	00:39 // I don't. // What was her name? //	
		00:41 // Eve. //
	00:42 // What a beautiful name. // Then what? // What happened? //	

Source: Survey data (2024).

In the first part, the daughter sits on her father's lap, with her back to him. Sitting behind the child, the father holds two teddy bears in his hands, facing the child. Throughout the video, the interaction takes place as follows: the father speaks in a caricatured voice, pretending to be the teddy bears, and the daughter responds by addressing them. In the first few seconds of the scene, the child begins her speech by introducing herself ("my name is Isabela"), while placing her hand on her chest, showing a deictic gesture of pointing to herself.

The father then gestures by swinging the teddy bears, simulating a conversation with the child. Based on the typology defined in Kendon's *continuum* (McNeill, 1992), these spontaneous movements during speech are classified as gesticulation, through which the father is "bringing to life" the teddy bears.

From the second 00:09, the child begins to tell the story, with the encouragement of the father, who interacts at every moment, showing surprise ("Ah!"), praising ("Oh, what a beautiful name! What else?") or asking ("But how did it happen after He created man? What happened?"). You can see that the father chooses the child as the main interlocutor, creating the "scenario" for the child to tell the story, guiding the child and showing interest in the child's speech.

At 00:49, the child begins to narrate the appearance of a snake ("Then, all of a sudden, a snake appeared!"), while at the same time making the gesture of raising his arms to the height of his head and opening his hands, projecting them forward, in order to imitate a snake. We describe this gesture as iconic, because it presents an image (the snake) through the arms and hands, representing a concrete entity, as McNeill (2005) points out.

Later on, two gestures are described which we categorize as gesticulation, following the model proposed in Kendon's *continuum* (McNeill, 1992): in the first (00:52), the father moves the two teddy bears away, making an expression of fright; and in the second (01:13), the child widens his eyes and arches his eyebrows, while reproducing his speech ("of the knowledge of good and evil"), reinforcing, through gesture, the idea of danger that is contained in the context.

With regard to the gazes, it can be seen that the child sometimes looks away from the toys next to them and looks back at the teddy bears, maintaining a dialog with them. The fact that the child pays attention to what the teddy bears "say" and responds by addressing them corroborates the importance of using playful practices to develop children's orality (Escarião, 2019).

Table 3 – Multimodal Envelope: part 2

Multimodal envelope		
Composition plans:	FATHER:	CHILD:
Look	00:45–1:27 The father continues behind the child, holding the teddy bears in front of him and looking at them.	00:45–01:27 The child looks at the toys next to them and then looks at the teddy bears again, continuing the dialog with them.
Gesture		00:49 Raises his arms above his head and opens his hands, projecting them forward to imitate a snake (iconic).
	00:52 The father pushes the two bears away, making a startled expression (gesturing).	
		01:13 He widens his eyes and arches his eyebrows (gesturing).
Vocal production	00:45 // Hmm... //	00:45 // The sna... // Eve was in paradise. //
		00:49 // Then, suddenly, a snake appeared! //
	00:52 // Hmm... // Then what? //	
		00:54 // Then he told Eve: // Eve, how about eating some of this fruit? //
	01:01 // Then what? //	
		01:03 // Then she ate it. //
	01:06 // Ahn...// But God had told her not to eat it, hadn't He? //	
		01:08 // He'd told her not to eat it. //
	01:11 // And what fruit was that? //	
		01:13 // Of the knowledge of good and evil. //
	01:15 // Hm... // Wow, I can't believe it. // It was... //	
		01:18 // Bad. //
	01:19 // And then. // The next day, what happened? //	
		01:22 // They were hidden from God. //
	01:26 // Then what? //	
	01:27 // Hmm //	01:27 // There, God was... walking through paradise. // Ouch... //

Source: Survey data (2024).

Table 4 – Multimodal Envelope: part 3

Multimodal envelope		
Composition plans:	FATHER:	CHILD:
Look	01:34 The father remains sitting behind the child, holding both teddy bears, facing the child, and looking at them.	01:34 The child looks at the toys next to him, looks straight ahead, and looks again at the teddy bears in his father's hands.
Gesture	01:34 The father keeps the conversation going with the teddy bears in his hands, swaying, as if they were talking (gesticulation).	
		02:36 Both hands together, clapping (emblematic)
Vocal production	01:34 // Who said that? //	
		01:35 // That you were naked? //
	01:37 // What did you do? //	
		01:40 // Y... you ate the fruit... I forbade you... not to eat it? //
	01:43 // And then. // Adam said what?	
		01:46 // Yes. // It was... it was my girl...friend that gave it to me. //
	01:50 // Jeez! // And what did he say to Eve? //	
		01:54 // No. // It was the snake that deceived me, and I ate it. //
	01:58 // Hmm // So, what happened? // What else? // What did God say to the snake? //	
		02:04 // In the end, the snake was going to crawl through the earth.
	02:10 // Ahn... // Crawl through the earth. // What else? //	
		02:12 // And eat of the dust of the earth. //
	02:14 Jeez! // Then what? //	
		02:15 // Ouch... // God put a fiery sword to prevent anyone from entering paradise.
	02:20 // Ah... // Got it. //	
	02:26 // And then... // go on. //	02:26 // Ouch. // Oh, they had lots of children. // Called Seven.
	02:32 // Seven? // Did they have a son called Seven?	
		02:34 // Yes. //
	02:36 // Hmm... //	02:36 // It's over //

Source: Survey data (2024).

In the third part, the father-daughter-teddy bear triad continues to follow the same dynamic: the father sits behind his daughter, holding the two teddy bears in his hands and facing her; the child looks at the toys next to her, looks straight ahead and looks back at the teddy bears in her father's hands, sometimes keeping her attention mainly on the teddy bears.

At 01:34, the father speaks while shaking the teddy bears as if they were talking, again characterizing gesticulation, as these are movements that are incorporated into the speech of "the teddy bears," since, in the children's imagination, they are the ones who are talking.

At the end of the video, at 02:36, the child finishes the story ("it's over") and puts both hands together, clapping, showing that she is satisfied with the work she has done. We consider clapping to be an emblematic gesture, even when accompanied by speech, because it has a culturally codified meaning and is widely recognized as a way of showing approval.

Based on the analysis described, and with a view to answering the questions raised at the beginning of this paper, it can be seen that children's language can be improved by relying on the caregiver's encouragement. In the study, it was observed that throughout the scene of interaction between father and daughter, the support of the father's words and gestures, through the play object, increased the interaction, adding encouragement for the story to continue. This means that episodes of joint attention can contribute to language development.

Consequently, it is possible to infer from the analysis that the interactional environment in which the child is inserted influences their development, to the extent that the whole produced by the child—gaze, gesture and vocal production—is strengthened through the participation of the caregiver. In the father-daughter-teddy bear triad, the father brings his daughter into the role of main interlocutor, validating her participation in the interaction and providing support so that she can communicate with confidence.

This continuous and progressive involvement of the daughter with the father can be understood, in the field of human learning, as Tomasello's (2019) conception of the process of the child's perception as an equal to the other, when they see the other as an equal being. In this sense, interaction in contexts of joint attention enables children to gradually recognize the other as an intentional agent, contributing to the formation of intersubjectivity, as well as enabling them to assimilate communicative patterns and participate more actively in interactions.

Thus, in the multimodal envelope presented, it can be seen that gestures function as a complementary entity to speech, bringing a range of meanings and adding emotion, an essential

factor in storytelling. We can also see that the playful element drives the dialog and brings the interlocutors even closer together.

Final considerations

By checking scenes of joint attention between father and daughter, this research aimed to identify the role of the caregiver in the child's expression of gestures and speech. To this end, a multimodal envelope analysis was carried out of a 2 min 37 s video in which the father encourages his daughter, aged around 36 months, to tell a story. The video being analyzed was described, highlighting the issues pertinent to the process of interaction between father and daughter. The presence of a third component (a teddy bear in the father's hands), which plays a fundamental role in the dynamics of the video, results in the triangle: child, adult, object, contemplated in Tomasello's studies (2019, 2023).

We can see the child's engagement in the activity by noticing that, throughout the filming, she directs her speech to the toy, without looking at her father at any time. Furthermore, considering multimodality as an intrinsic component of the language acquisition process, the constant presence of gestures—gestural, deictic, iconic, emblematic—and the development of speech seen in the video lead the discussion to the point where language is established through a matrix made up of a set of gestures and vocal productions (McNeill, 1985).

As a result of the analysis carried out, it can be seen that the support of speech and gestures is strengthened by the actions of a caregiver who seeks to stimulate the imagination and expose the child to contexts of interaction in which they play the role of the main interlocutor, participating in the dialog in equivalence with the adult. In this way, the caregiver not only helps develop language skills, but also fosters the child's imagination and autonomy.

The use of toys to make up the triad, bringing playfulness to the interaction, must also be taken into account for involvement during the activity. As Escarião (2019, p. 172) points out, children “enter the world of language, experiment with language, have the chance to expand their linguistic repertoire, as well as experience the essence of genres, which is conversation, dialog.” In this sense, it is essential for good language development that they have the freedom to communicate in their own way when interacting with adults or other children.

Finally, the aim of this work is to contribute to the construction of knowledge and the universality of empirical studies of language acquisition, especially in its multimodal aspect.

This highlights the importance of continuing research into matters related to the multimodality of gesture and speech and joint attention, considering different contexts and interactional practices.

REFERENCES

BRUNER, J. **Le développement de l'enfant: savoir faire, savoir dire**. Paris: PUF, 2004.

BRUNER, J. **Como as crianças aprendem a falar**. Tradução de Joana Chaves. Lisboa: Horizontes Pedagógicos, 2007.

DEL RÉ, A.; HILARIO, R. N.; VIEIRA, A. J. A linguagem da criança na concepção dialógico-discursiva: retrospectiva e desafios teórico-metodológicos para o campo de Aquisição da Linguagem. **Bakhtiniana**, São Paulo, v. 16, n. 1, p. 12-38, 2021. DOI: 10.1590/2176-457348071. Available from: <https://www.scielo.br/j/bak/a/dRS98pVJT4mJdmcc7JvkjyB/>. Accessed in: 4 Apr. 2023.

ESCARIÃO, A. D. **Oralidade em práticas lúdicas na Educação Infantil**. 2019. 218 f. Tese (Doutorado em Linguística) – Universidade Federal da Paraíba, João Pessoa, 2019.

HOLANDA, D. X.; CAVALCANTE, M. C. B.; SILVA, E. E. O.; SILVA, D. M.; LIMA, V. P. Sob um olhar dialógico-multimodal: a matriz gesto e fala como palco de entrada da criança na linguagem. **Revista do GELNE**, Natal, v. 25, n. 3, e32212, 2023. DOI: 10.21680/1517-7874.2023v25n3ID32212. Available from: <https://periodicos.ufpb.br/gelne/article/view/32212>. Accessed in: 26 Mar. 2025.

KENDON, A. Gesticulation and speech: Two aspects of the process of utterance. In: KEY, M. R. (ed.). **The relation between verbal and nonverbal communication**. Haia: De Gruyter Mouton, 1980. p. 207-227.

LANGACKER, R. A dynamic usage-based model. In: BARLOW, M.; KEMMER, S. (ed.). **Usage-based models of Language**. Califórnia: SLI Publications, 2000.

MCNEILL, D. So you think gestures are nonverbal? **Psychological Review**, v. 92, n. 3, p. 350-371, 1985. Available from: http://www.communicationcache.com/uploads/1/0/8/8/10887248/so_you_think_gestures_are_nonverbal.pdf. Accessed in :26 Mar. 2025.

MCNEILL, D. **Hand and mind: what gestures reveal about thought**. Chicago: University of Chicago Press, 1992.

MCNEILL, D. **Gesture and thought**. Chicago: University of Chicago, 2005.

MCNEILL, D. **Why we gesture: the surprising role of hand movements in communication**. New York: Cambridge University Press, 2016.

NÓBREGA, P. V. A. **Dialogia mãe-bebê: a emergência do envelope multimodal em contextos de atenção conjunta**. 2010. 168 f. Dissertation (Master's Degree in Linguistics and Education) – Universidade Federal da Paraíba, João Pessoa, 2010. Available from: <https://repositorio.ufpb.br/jspui/handle/tede/6511>. Accessed in: 26 Mar. 2025.

SAPIR, E. **A linguagem: introdução ao estudo da fala**. 2. ed. São Paulo: Perspectiva, 1980.

SILVA, F. A.; FARIA, E. M. B. Considerações sobre gestos da fala: contínuo de Kendon, dimensões e sistema integrado. **Acta Semiótica et Lingvística**, v. 27, n. 1, 2022. DOI: 10.22478/ufpb.2446-7006.46v27n1.62399. Available from: <https://periodicos.ufpb.br/ojs2/index.php/actas/article/view/62399>. Accessed in: 20 Mar. 2025.

TOMASELLO, M. **Origens culturais da aquisição do conhecimento humano**. 2. ed. São Paulo: Martins Fontes, 2019.

TOMASELLO, Michael. Entrevista com Michael Tomasello. [Entrevista cedida a] BARBOSA, T. R.; SILVA, H. M. L.; OLIVEIRA, R. C.; GOUVEIA, R. T. **Revista da ABRALIN**, v. 23, n. 2, p. 877-893, 2024. DOI: 10.25189/rabralin.v23i2.2202. Available from: <https://revista.abralin.org/index.php/abralin/article/view/2202>. Accessed in: 26 mar. 2025.

VYGOTSKY, L. S. **Pensamento e linguagem**. São Paulo: Martins Fontes, 2005.

VYGOTSKY, L. S. **A formação social da mente**. São Paulo: Martins Fontes, 1978.

CRediT Author Statement

- ☐ **Acknowledgements:** I would like to thank the journal's editorial team for their prompt and excellent service.
 - ☐ **Funding:** There is no support from institutions.
 - ☐ **Conflicts of interest:** There are no conflicts of interest.
 - ☐ **Ethical approval:** The work respected ethics during the research. It did not have to go through an ethics committee.
 - ☐ **Data and material availability:** All the data of interest in the work is in the work itself, as well as the bibliographical references. The data and materials used in the work are from the author's own collection.
 - ☐ **Authors' contributions:** Author 1 – Milena Magalhães Gomes: video collection from the author's collection; data analysis and interpretation, bibliographical research and writing of the text; Author 2 – Tatiana Ramalho Barbosa: data analysis and interpretation, bibliographical research and writing of the text.
-

Processing and editing: Editora Ibero-Americana de Educação
Proofreading, formatting, standardization and translation

