

**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**

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



Milena Magalhães GOMES<sup>1</sup>  
e-mail: milenamagalhaes.adv@gmail.com



Tatiana Ramalho BARBOSA<sup>2</sup>  
e-mail: tatiramalho@hotmail.com

**Como referenciar este artigo:**

GOMES, M. M.; BARBOSA, T. R. A influência da interação do cuidador na expressão de gesto e fala da criança: um estudo de caso. **Rev. EntreLínguas**, Araraquara, v. 11, n. 00, e025006, 2025. e-ISSN: 2447-3529. DOI: 10.29051/el.v11i00.20143



| Submetido em: 03/02/2025  
| Revisões requeridas em: 06/03/2025  
| Aprovado em: 04/04/2025  
| Publicado em: 10/08/2025

**Editora:** Profa. Dra. Rosangela Sanches da Silveira Gileno

**Editor Adjunto Executivo:** Prof. Dr. José Anderson Santos Cruz

<sup>1</sup> Mestranda em Linguística, PROLING/UFPB, João Pessoa – PB – Brasil.

<sup>2</sup> Doutoranda em Linguística, PROLING/UFPB, João Pessoa – PB – Brasil.

**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.

**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.

## Introdução

Os seres humanos possuem uma capacidade biologicamente herdada de viver dentro de uma cultura, mas a herança biológica não escusa o contexto de interação do indivíduo com o meio, que se configura como crucial para a formação da cognição e de todas as características da espécie (Tomasello, 2019). Ou seja, a habilidade de engajar-se e participar culturalmente não é apenas um produto do ambiente ou da sociedade, mas está enraizada na biologia humana. Por exemplo, a capacidade de aprender uma linguagem, jogar xadrez ou participar de rituais sociais, ou religiosos, são habilidades que emergem naturalmente em humanos devido à nossa constituição biológica em interação com o meio.

Dessa forma, visando compreender o processo de aquisição e produção da linguagem humana, é necessário observá-la como um produto da cognição, através de cenas de atenção conjunta, mediante a convergência das atividades de aquisição linguística e de desenvolvimento cognitivo.

Diante desse contexto, esta pesquisa tem como objetivo analisar cenas de atenção conjunta entre pai e filha, a fim de identificar o papel do cuidador no desenvolvimento de gestos e falas da criança. A metodologia utilizada baseia-se na criação de um Envelope Multimodal, o qual corresponde ao conjunto olhar–gesto–produção vocal, através do qual foi feita a análise de um vídeo de 2 min 37 s, onde o cuidador (pai) estimula a contação de história que está sendo conduzida por sua filha, de aproximadamente 36 meses de idade, que direciona sua fala a um objeto, um ursinho de pelúcia.

Até que ponto a linguagem infantil é melhorada com o suporte de falas e gestos de um cuidador atencioso, que estimule a imaginação e promova o desenvolvimento da aquisição da linguagem? E qual é a contribuição do meio interacional em que a criança está inserida? Esta pesquisa busca responder a essas questões, por meio do estudo de caso e amparada pelas contribuições teóricas sobre atenção conjunta e aquisição da linguagem.

Partindo da perspectiva sociointeracionista, desde Vygotsky (1978), passando por Langacker (2000) e Tomasello (2024), buscou-se, com este estudo, demonstrar como crianças se envolvem em eventos de intencionalidade conjunta (Tomasello, 2024), que incluem momentos de atenção conjunta, a partir dos quais começam a entender as outras pessoas como agentes intencionais, iguais a elas próprias.

No vídeo analisado para este estudo, percebe-se a intenção deliberada do cuidador em promover o desenvolvimento da linguagem, da narração e da argumentação de sua filha,

exatamente por ter consciência de que ela ainda está no período crítico de aquisição da linguagem e, portanto, sua interação com o meio assume papel crucial.

Com vistas a estruturar detalhadamente a discussão proposta neste artigo, dividimos o conteúdo em quatro partes, incluindo esta seção introdutória. O próximo tópico traz o arcabouço teórico que norteou o estudo. Na sequência, são descritas a metodologia empregada e a discussão dos resultados encontrados. Por fim, são expostas as considerações finais, com as constatações geradas pela pesquisa.

## **Fundamentação teórica**

### **Processos históricos e ontogenéticos da aquisição da linguagem humana**

No que diz respeito aos processos históricos e de formação cognitiva da competência linguística, Tomasello (2024) traz evidências, em seu estudo, de que a nossa espécie passa ontogeneticamente por duas etapas: colaboração e aspectos culturais. O autor relata que partiu de pesquisas iniciadas por Vygotsky (1978) com crianças e passou a fazer experimentos com chimpanzés para concluir que esses dois aspectos são unicamente humanos: a colaboração, visto que, desde pequenas, as crianças se engajam em atividades colaborativas; e os artefatos culturais, que influenciam como as crianças aprendem normas e começam a interagir em atividades em grupo.

Nessa direção, Sapir (1980) também aponta a fala como uma função adquirida e cultural, diferente de outras atividades, como andar, que, para o autor, é uma função orgânica e instintiva. Segundo Sapir (1980, p. 12), “falar é uma herança histórica do grupo, produto de um uso social prolongado”, ainda que as pessoas não se deem conta do esforço cognitivo que despendem para desenvolverem a linguagem.

É através dos processos de colaboração e cultura que se estabelece a hipótese da aquisição baseada no uso, defendida por autores como Tomasello (2019) e Langacker (2000). De acordo com essa teoria, os conhecimentos dos diversos domínios cognitivos (tais como jogar xadrez, contar, vestir-se etc.) são adquiridos pelo uso, ou seja, através da interação com o meio. Isso inclui a linguagem, na visão sociointeracionista. Vale ressaltar que, embora descrevam detalhadamente aspectos do uso e da interação social, autores como Tomasello (2019, 2024) e Vygotsky (1978, 2005) não descartam o componente biológico ou os mecanismos cognitivos

presentes nos humanos, considerando-os como “resultado de algum tipo ou modo de transmissão cultural únicos da espécie [humana]” (Tomasello, 2019, p. 5). Ou seja, a visão sociointeracionista também abarca as características biológicas, porém as assume como parte do processo.

Desse modo, segundo Tomasello (2019, p. 7, grifos do autor), a aprendizagem humana acontece pelo fato de cada indivíduo reconhecer os outros como coespecíficos, ou seja, “como seres *iguais a ele*, com vidas mentais e intencionais iguais às dele”. Portanto, quando a criança passa a perceber outras pessoas como “iguais a ela”, ela é capaz de se imaginar “‘na pele mental’ de outra pessoa, de modo que não só aprende *do* outro, mas *através* do outro”. Assim, o que diferencia os seres humanos dos primatas não humanos, e que foi responsável pela evolução cultural que vivemos até hoje, é exatamente a capacidade de transmitir aos demais membros da espécie inúmeros feitos em busca de um bem comum.

Essa característica humana é chamada por Tomasello (2019, 2024) de “efeito catraca”, um processo de evolução cumulativa histórica, transmitida pelo social e de forma confiável, que acontece através da aprendizagem por imitação. Outro ponto mencionado pelo autor é a instrução ativa por parte dos indivíduos de uma determinada sociedade, ao repassarem suas experiências de maneira que é possível perceber a acumulação do conhecimento adquirido ao longo das gerações.

Sendo assim, a estrutura física e mental dos nossos antepassados, juntamente com as formas como percebemos, comportamo-nos e pensamos, são herdadas biologicamente. No entanto essas características só se manifestam e se desenvolvem em interação com o ambiente ao nosso redor. Defende-se, neste estudo, uma teoria da aquisição da linguagem baseada no uso, conforme Tomasello (2019), que propicia a observação da produção e da compreensão da fala para explicar as bases sociocognitivas da aquisição, em evidências empíricas de fala entre a criança e o adulto.

Para Tomasello (2019), a aprendizagem ocorre a partir de três tipos básicos: por imitação, por instrução e por colaboração. Sabemos que esses três gestos assumem crucial importância no desenvolvimento da cognição e, especialmente, da linguagem em crianças. É nesse sentido que o autor defende a atenção conjunta, tema que passaremos a tratar adiante.

## A atenção conjunta e a aquisição da linguagem

As cenas de atenção conjunta são componentes do ambiente cultural, em que a interação favorece o reconhecimento dos coespecíficos e a interpretação de suas intenções. São instâncias concretas de interação entre a criança e o cuidador, que funcionam como a engrenagem central para promover o processo de aquisição da linguagem. Elas giram em torno do compartilhamento da atenção, do contexto, dos sistemas linguístico e conceitual, das intenções e da intersubjetividade, envolvendo operações de compreensão entre a criança e o adulto, considerando as operações cognitivas necessárias para fins de comunicação efetiva.

A atenção conjunta ocorre entre o mundo perceptual e o mundo linguístico como um dos fundamentos sociocognitivos primários da aquisição da linguagem, fazendo parte da realidade socialmente compartilhada e favorecendo o contexto intersubjetivo, em que ocorre a percepção do símbolo e a interpretação das intenções dos envolvidos. “O termo *atenção conjunta* costuma ser usado para caracterizar todo esse complexo de habilidades e interações sociais” (Tomasello, 2019, p. 85-86). Sua emergência acontece entre os nove e os doze meses, quando o bebê apresenta comportamento triádico, que envolve interação com objetos e pessoas, resultando no triângulo: criança, adulto, objeto ou evento, conforme ilustra a Figura 1.

**Figura 1** – Atenção conjunta



Fonte: elaboração Própria baseada em Tomasello (2019).



A ilustração aborda o modelo de interação que ocorre nessa fase do desenvolvimento dos bebês humanos (entre os nove e doze meses), chamado de “comportamentos de atenção conjunta”, acompanhando os gestos do cuidador e, em especial, apontando para as coisas a que se referem. Essa mudança de comportamentos diádicos (entre dois entes) para triádicos (três entidades) é exatamente o que diferencia o ser humano de outros primatas (Tomasello, 2019).

Nesse sentido, a perspectiva dialógico-discursiva, que se baseia em pesquisas precedentes de autores como Vygotsky (2005) e Bruner (2004, 2007), endossa o ponto de vista da interação como importante componente do processo de aquisição da linguagem pela criança, destacando a importância do cuidador e, ainda, da presença do instrumento lúdico nesse processo interacional. Segundo Del Ré, Hilario e Vieira (2021), por abordagem dialógico-discursiva entende-se o estudo do desenvolvimento da linguagem (e da cognição) sob o ponto de vista da interação entre a criança e o outro.

Por essa abordagem, analisa-se a fala da criança por meio da interação entre ela e o seu interlocutor, levando em consideração os diferentes aspectos que interferem nessa relação, como: “os contextos discursivos e situacionais, a dialogia (constitutiva de todo discurso), as relações de interação empírica entre sujeitos, a constituição dos sujeitos no discurso e o papel que assumem durante a comunicação, etc.” (Del Ré; Hilario; Vieira, 2021, p. 13). O comportamento triádico, portanto, intrínseco aos seres humanos, é afetado pelo cenário de uso da linguagem.

Assim, a interação é vista como o momento de interseção da perspectiva dialógica e multimodal da linguagem, isso porque a interação social é o fator que “propicia um ambiente para a entrada da criança na linguagem, e a perspectiva dialógica aponta para a relação da criança consigo e com o mundo, ou seja, a construção da ideia do outro como seu interlocutor, a qual só é possível porque a linguagem é multimodal” (Holanda *et al.*, 2023, p. 5).

Com relação às fases do desenvolvimento, Tomasello (2019) descreve o período da emergência da atenção conjunta como a revolução dos nove meses. A partir de então, as cenas de interação passam a ser cada vez mais complexas, em decorrência da compreensão da simbolização que eventos, objetos e pessoas adquirem, bem como da descoberta e uso de construções de linguagem que as crianças aprendem, primeiramente por imitação, mas, em seguida, com intencionalidade própria.

Para Tomasello (2019, p. 94-95), “agentes intencionais são seres comuns que têm objetivos e fazem escolhas ativas entre os meios comportamentais disponíveis para atingir

aqueles objetivos, o que inclui escolher ativamente a que se vai prestar atenção na busca desses objetivos”. Ou seja, a intencionalidade atua como um catalisador para o desenvolvimento das habilidades de fala.

A Figura 2 descreve a categorização conceitual do autor sobre as cenas de experiências vividas pelas crianças.

**Quadro 1** – Análise e categorização conceitual de cenas de experiência

<b>Idade aproximada</b>	<b>Cena experiencial</b>	<b>Linguagem</b>
9 meses	Cenas de atenção conjunta (não-simbolizadas)	–
14 meses	Cenas simbolizadas (simbolização indiferenciada)	Holófrases
18 meses	Cenas separadas (diferenciação de eventos e participantes)	Construções de tipo pivô
22 meses	Cenas sintáticas (marcação simbólica de participantes)	Construções verbais insuladas
36 meses	Cenas categorizadas (marcação simbólica generalizada de papeis de participantes)	Construções verbais gerais

Fonte: Tomasello (2019, p. 212)

Destarte, percebe-se uma evolução complexa para a aquisição de uma língua natural. E, para que o desenvolvimento da criança se dê de forma completa, é inelutável a convivência com o meio e a interação com outros indivíduos. Dessa forma, cenas de atenção conjunta, como as descritas e analisadas neste estudo, são essenciais para a compreensão dos símbolos, a apreensão de estruturas próprias da língua, para a perspectivação e para a construção de discurso próprio e intencional.



## O papel dos gestos na comunicação

Para McNeill (2016, p. 4, tradução nossa), os gestos são a “imagem intrínseca da linguagem”. Ou seja, são movimentos utilizados intencionalmente que, quando acompanhados pela oralidade, potencializam o significado contido na fala. Dessa forma, gesto e fala funcionam como um sistema unificado, em que os gestos são um componente da fala e do pensamento, uma vez que a linguagem está necessariamente atrelada à imagem (McNeill, 2005).

As principais definições dos gestos foram abordadas por Kendon (1980), que passou a estudar as suas funções dentro dos contextos de interação. Essa tipologia ficou conhecida pelo que McNeill (1992) chamou de “continuum de Kendon” e considera cinco tipos: gesticulações, gestos preenchedores<sup>3</sup>, emblemas, pantomimas, e línguas de sinais.

Segundo descrevem Silva e Faria (2022, p. 203), “a gesticulação é um movimento que carrega um significado relacionado ao enunciado que o acompanha e abrange muitas variantes e usos”. Ou seja, são movimentos espontâneos, normalmente das mãos e dos braços, que estão em consonância com a linguagem oral.

Os gestos preenchedores ocupam uma lacuna no enunciado. São parecidos com o sentido da gesticulação, porém, nos gestos preenchedores, o locutor interrompe a sua fala e encaixa um gesto, que dará sentido ao enunciado. McNeill (1992, p. 37, tradução nossa) traz o seguinte exemplo, para explicar esse tipo de gesto: “os pais estavam bem, mas as crianças estavam [gesto]”, em que o gesto ocupa o espaço da palavra “mal”.

Os emblemas são gestos que seguem um padrão e que possuem um significado culturalmente estabelecido. São como um código social e não necessitam da fala para produzir sentido (McNeill, 1992). O exemplo clássico para esse tipo de gesto é o sinal de “OK” com a mão; ou o sinal de “joia”, se considerarmos o contexto brasileiro.

Pantomimas são representações que contam uma história sem a presença da fala. Silva e Faria (2022, p. 205) descrevem como exemplos de pantomimas: “fingir estar sentindo alguma dor ou simular estar alegre”. Por sua vez, as línguas de sinais são os gestos que representam o léxico de um sistema linguístico de determinada comunidade, por exemplo, a Língua Brasileira de Sinais (Libras) e outros códigos linguísticos de diferentes países (Silva; Faria, 2022, p. 205).

<sup>3</sup> Considerando a tradução de *speech-framed gesture* em Cavalcanti (2018).

Em um estudo realizado por McNeill e Levy (1982, *apud* McNeill, 1992), os gestos — especificamente gesticulações e gestos preenchedores — são enquadrados ainda a partir das seguintes dimensões: icônicos, metafóricos, dêiticos e ritmados.

São considerados icônicos, na percepção de McNeill (2005), os movimentos que reproduzem alguma ação, ou algum objeto ou entidade concreta. Semelhantes aos gestos icônicos, os metafóricos também representam algo; porém podem significar qualquer entidade de conteúdo abstrato, como uma ideia ou uma memória. Silva e Faria (2022, p. 207-208) exemplificam essa categoria com “a configuração das mãos levantadas com a palma para cima, como se o falante quisesse demonstrar através desse gesto que não entendeu nada da aula a que assistiu”. Os ombros também podem ser levantados para enfatizar a intenção de dúvida do interlocutor.

Os dêiticos são os gestos de apontar, com a função de indicar objetos ou eventos, que podem ser concretos ou abstratos. McNeill (1992) frisa que é comum que, nos contextos de interação, a ação de apontar esteja indicando um objeto presente; no entanto podem indicar algo que não está visível fisicamente.

Por último, os gestos ritmados são assim chamados porque funcionam como as batidas de um ritmo musical. São movimentos sem conteúdo semântico, curtos e rápidos, com as mãos ou com os dedos, para cima e para baixo, ou para frente e para trás. A sua função é marcar uma parte da sentença que se quer ressaltar dentro do contexto geral (McNeill, 1992).

Partindo dos conceitos apresentados acerca da ontogênese da aquisição da linguagem, bem como da atenção conjunta e das variadas formas de categorização dos gestos, nos diferentes contextos de interação humana, expõe-se, no tópico a seguir, a metodologia de análise do vídeo com a cena dialógica entre pai e filha.

## **Metodologia e Discussão dos Resultados**

Os dados são retirados de gravação feita através de um celular, durante a interação pai–filha–ursinho, sem que a criança percebesse que estava sendo filmada. Trata-se de uma criança do sexo feminino, de aproximadamente 36 meses, em contexto de interação com o pai, em que ele segura o ursinho e interage com a filha como se fosse o brinquedo que estivesse falando. O vídeo tem duração de 2 m 37 s, e é apresentado aqui de forma segmentada, em três partes, nas

quais descrevemos os recortes temporais de falas e descrição das ações de ambos os parceiros da interação, em uma contação de história.

Nas situações comunicativas a seguir, observamos as cenas que foram descritas separadamente, a fim de analisarmos os gestos que a criança realiza enquanto conta uma história. Nesse sentido, não apenas a fala, mas também os gestos (e até o olhar) fazem parte do que Nóbrega (2010, p. 17) chama de “planos de composição de um envelope da multimodalidade linguística”. A seguir, são descritas as cenas de atenção conjunta em que três componentes da interação são levados em conta: o olhar, o gesto e a produção vocal. Considera-se que o Envelope Multimodal é o conjunto desses componentes, conforme Quadro 2.

Na primeira parte, a filha se acomoda no colo do pai, de costas para ele. Sentado atrás da criança, o pai segura dois ursinhos de pelúcia com as mãos, virados de frente para a criança. Durante todo o vídeo, a interação acontece da seguinte forma: o pai fala com uma voz caricata, simulando ser os ursinhos, e a filha responde dirigindo-se a eles. Nos primeiros segundos da cena, a criança inicia a sua fala apresentando-se (“meu nome é Isabela”), enquanto coloca a mão no peito, manifestando um gesto dêitico de apontar para si.

Na sequência, o pai gesticula balançando os ursinhos, simulando uma conversa com a criança. Com base na tipologia definida no *continuum* de Kendon (McNeill, 1992), esses movimentos espontâneos, durante a fala, são classificados como gesticulação, através da qual o pai estaria “dando vida” aos ursinhos.

A partir do segundo 00:09, a criança começa a contar a história, com o incentivo do pai, que a todo momento interage, demonstrando surpresa (“Ham!”), elogiando (“Ai, que nome lindo! Que mais?”) ou perguntando (“Mas como é que foi depois que ele criou o homem? O que aconteceu?”). É possível perceber que o pai elege a criança como principal interlocutora, ao criar o “cenário” para que ela conte a história, guiando-a e manifestando interesse na fala da criança.

**Quadro 2 – Envelope Multimodal: parte 1**

Envelope multimodal		
Planos de composição:	PAI:	CRIANÇA:
Olhar	00:01 O pai se senta atrás da criança, enquanto segura dois ursinhos de pelúcia nas mãos, virados de frente para a criança. Ele olha para os ursinhos e encena uma conversa.	00:01 A criança responde olhando para os dois ursinhos.
		00:02 A criança se acomoda no colo do pai, de costas para ele e de frente para os ursinhos que o pai está segurando.
Gestos		00:01 Coloca a mão direita no peito enquanto se apresenta (dêitico).
	00:02 O pai segura dois ursinhos de pelúcia nas mãos, virados de frente para a criança, encenando uma conversa com a criança. Gesticula balançando os dois ursinhos, durante a sua fala (gesticulação).	
Produção vocal		00:01 // Meu nome é Isabela//
	00:02 // Isabela, conta pra mim como é que foi a criação do mundo. // Como é que foi, no primeiro dia? // Como é que foi? //	
		00:09 // No primeiro dia, Deus criou o homem do pó da terra. //
	00:13 //Isso foi no sexto dia. // Mas como é que foi depois que ele criou o homem? // O que aconteceu? //	
		00:17 // Ele criou do pó da terra. //
	00: 19 // E aí? //	
	00: 20 // Hum //	00:20 // Ele soprou no nariz dele // E... e... // Sabe qual era o nome dele? //
	00:27 // Qual era? //	
		00:28 // Adão. //
	00:29 // Ham! // Ai que nome lindo! // Que mais? //	
	00:32 //Ham! //	00:32 // Deus criou uma mulher. // Que... que... que... // Sabe qual era o nome da mulher? //
	00:39 // Eu não. // Qual era o nome dela? //	
		00:41 // Eva. //
	00:42 // Ai que nome lindo. // E aí? // O que aconteceu? //	

Fonte: dados da pesquisa (2024).

**Quadro 3** – Envelope Multimodal: parte 2

Envelope multimodal		
Planos de composição:	PAI:	CRIANÇA:
Olhar	00:45–1:27 O pai continua atrás da criança, segurando os ursinhos na sua frente e olhando para eles.	00:45–01:27 A criança olha para os brinquedos que estão ao lado, e volta a olhar para os ursinhos, mantendo o diálogo com eles.
Gesto		00:49 Levanta os braços na altura da cabeça e abre as mãos, projetando-as para frente, para imitar uma serpente (icônico).
	00:52 O pai afasta os dois ursinhos, fazendo expressão de susto (gesticulação).	
		01:13 Arregala os olhos e arqueia as sobrancelhas (gesticulação).
Produção vocal	00:45 // Hum... //	00:45 // A serp... // Eva tava no paraíso. //
		00:49 // Aí, de repente, apareceu uma serpente! //
	00:52 // Hmmm... // E aí? //	
		00:54 // Aí, falou pra Eva: // Eva, que tal você comer desse fruto aqui? //
	01:01 // E aí? //	
		01:03 // Aí, ela comeu. //
	01:06 // Hmmm... // Mas Deus tinha falado pra ela não comer, foi? //	
		01:08 // Tinha falado pra ela não comer. //
	01:11 // E que fruto era esse? //	
		01:13 // Do conhecimento do bem e do mal. //
	01:15 // Hmm... // Eita, não acredito. // Que coisa mais... //	
		01:18 // Feia. //
	01:19 // E aí, depois. // No outro dia, o que aconteceu? //	
		01:22 // Eles ficaram escondidos de Deus. //
	01:26 // E aí? //	
	01:27 // Hum //	01:27 // Aí, Deus ta... tava caminhando pelo paraíso. // Ai... //

Fonte: dados da pesquisa (2024).

Em 00:49, a criança começa a narrar a aparição de uma serpente (“Aí, de repente, apareceu uma serpente!”), ao mesmo tempo em que faz o gesto de levantar os braços na altura da cabeça e abrir as mãos, projetando-as para frente, no intuito de imitar uma cobra. Descrevemos esse gesto como icônico, por apresentar uma imagem (a serpente) através dos braços e mãos, representando uma entidade concreta, conforme aduz McNeill (2005).

Mais adiante, são descritos dois gestos que categorizamos como gesticulação, seguindo o modelo proposto no *continuum* de Kendon (McNeill, 1992): no primeiro (00:52), o pai afasta os dois ursinhos, fazendo expressão de susto; e no segundo (01:13), a criança arregala os olhos e arqueia as sobrancelhas, enquanto reproduz a sua fala (“do conhecimento do bem e do mal”), reforçando, através do gesto, a ideia de perigo que está contida no contexto.

Com relação aos olhares, nota-se que a criança desvia o olhar algumas vezes na direção dos brinquedos que estão ao seu lado e volta a olhar para os ursinhos, mantendo o diálogo com eles. O fato de a criança prestar atenção ao que os ursinhos “falam” e responder dirigindo-se a eles corrobora a importância da utilização de práticas lúdicas para o desenvolvimento da oralidade infantil (Escarião, 2019).

Na terceira parte, a tríade pai-filha-ursinho permanece seguindo a mesma dinâmica: o pai sentado atrás da filha, segurando os dois ursinhos de pelúcia nas mãos e virados de frente para ela; a criança olha para os brinquedos que estão ao lado, olha para a frente e volta a olhar para os ursinhos nas mãos do pai, algumas vezes, mantendo a atenção principalmente nos ursinhos.

Em 01:34, o pai fala enquanto balança os ursinhos como se eles estivessem falando, caracterizando novamente a gesticulação, por se tratar de movimentos que se incorporam à fala “dos ursinhos”, já que, no imaginário infantil, são eles que estão falando.

Ao final do vídeo, em 02:36, a criança encerra a história (“cabou”) e junta as duas mãos, batendo palmas, demonstrando que ficou satisfeita com o trabalho realizado. Consideramos as palmas como um gesto emblemático, mesmo vindo acompanhadas de uma fala, porque possuem um significado culturalmente codificado e são amplamente reconhecidas como uma forma de demonstrar aprovação.

**Quadro 4 – Envelope Multimodal: parte 3**

Envelope multimodal		
Planos de composição:	PAI:	CRIANÇA:
Olhar	01:34 O pai permanece sentado atrás da criança, segurando os dois ursinhos de pelúcia, virados de frente para a criança, e olhando para eles.	01:34 A criança olha para os brinquedos que estão ao lado, olha para a frente, e volta a olhar para os ursinhos nas mãos do pai.
Gesto	01:34 O pai mantém o diálogo com os ursinhos nas mãos, balançando, como se eles estivessem falando (gesticulação).	
		02:36 Junta das duas mãos, batendo palmas (emblemático).
Produção vocal	01:34 // Quem disse quê? //	
		01:35 // Que vocês estavam nus? //
	01:37 // Vocês por acaso fizeram o quê? //	
		01:40 // Co...comeram do fruto... proibi... vocês não comerem? //
	01:43 // E aí... // Adão falou o quê? //	
		01:46 // Sim. // Foi... foi a minha na...morada que me deu. //
	01:50 // Eita! // E ele falou o que pra Eva? //	
		01:54 // Não. // Foi a serpente que me enganou, e eu comi. //
	01:58 // Hummm // E aí, o que aconteceu? // Que mais? // Deus falou o que pra serpente? //	
		02:04 // Por esse fim, a serpente ia rastejar pela terra.
	02:10 // Ham... // Rastejar pela terra. // Que mais? //	
		02:12 // E comer do pó da terra. //
	02:14 // Eita! // E aí? //	
		02:15 // Ai... // Deus colocou uma espada de fogo pra... pra ninguém entrar no paraíso.
	02:20 // Ah... // Entendi. //	
	02:26 // E depois... // vai. //	02:26 // Ai. // Ai... eles tiveram muitos filhos. // Chamado Sete.
	02:32 // Sete? // Tiveram um filho chamado Sete? //	
		02:34 // Sim. //
	02:36 // Hum... //	02:36 // Cabou. //

Fonte: dados da pesquisa (2024).



A partir da análise descrita, e visando responder aos questionamentos levantados no início deste trabalho, é possível verificar que a linguagem infantil pode ser melhorada apoiando-se no estímulo do cuidador. No estudo, observa-se que, ao longo de toda a cena de interação entre pai e filha, o suporte de falas e gestos do pai, através do objeto lúdico, incrementou a interação, acrescentando um encorajamento para a continuação da estória. Isso significa que episódios de atenção conjunta são capazes de contribuir para o desenvolvimento da linguagem.

Por consequência, é possível inferir, com base na análise, que o meio interacional em que a criança está inserida influencia no seu desenvolvimento, na medida em que o conjunto por ela produzido — olhar, gesto e produção vocal — é fortalecido através da participação do cuidador. Na tríade pai-filha-ursinho, o pai traz a filha para o papel de interlocutora principal, validando a sua participação na interação e atribuindo suporte para que ela consiga realizar a comunicação com confiança.

Esse envolvimento contínuo e progressivo da filha com o pai pode ser entendido, na seara da aprendizagem humana, como a concepção de Tomasello (2019) sobre o processo de percepção da criança como um igual ao outro, quando ela vê o outro como um ser igual. Nesse sentido, a interação em contextos de atenção conjunta possibilita que a criança gradualmente passe a reconhecer o outro como um agente intencional, contribuindo para a formação da intersubjetividade, além de possibilitar que ela assimile padrões comunicativos e participe de maneira mais ativa nas interações.

Destarte, no envelope multimodal apresentado, verifica-se que a gestualidade funciona como uma entidade complementar à fala, trazendo amplitude de sentidos e acrescentando a emoção, fator essencial na contação de histórias. Percebe-se, ainda, que o elemento lúdico impulsiona o diálogo e aproxima ainda mais os interlocutores.

## **Considerações finais**

A partir da verificação de cenas de atenção conjunta entre pai e filha, esta pesquisa objetivou identificar o papel do cuidador na expressão de gestos e falas da criança. Para esse fim, metodologicamente, foi realizada a análise, através do envelope multimodal, de um vídeo de 2 m 37 s em que o pai estimula a filha, com idade aproximada de 36 meses, a contar uma história. O vídeo em análise foi descrito, evidenciando as questões pertinentes ao processo de

interação entre pai e filha. A presença de um terceiro componente (ursinho nas mãos do pai), que passa a ocupar papel fundamental na dinâmica do vídeo, resulta no triângulo: criança, adulto, objeto, contemplado nos estudos de Tomasello (2019, 2024).

Constatamos o engajamento da criança na atividade, ao perceber que ela, durante toda a filmagem, direciona sua fala ao brinquedo, sem olhar para o pai em nenhum momento. Além disso, considerando a multimodalidade como componente intrínseco do processo de aquisição da linguagem, a presença constante dos gestos — de gestualidade, dêiticos, icônicos, emblemáticos — e o desenvolvimento da fala verificados no vídeo conduzem a discussão ao ponto em que a língua se estabelece através de uma matriz composta pelo conjunto de gestos e produções vocais (McNeill, 1985).

Verifica-se, como consequência da análise realizada, que o suporte de falas e gestos é fortalecido pela atuação de um cuidador que busca estimular a imaginação e expor a criança a contextos de interação em que ela ocupe um papel de principal interlocutora, participando do diálogo em equivalência com o adulto. Dessa forma, o cuidador não apenas auxilia no desenvolvimento de habilidades linguísticas, mas também fomenta a imaginação e a autonomia da criança.

A utilização do brinquedo para compor a tríade, trazendo o lúdico para a interação, também deve ser levada em consideração para o envolvimento durante a atividade. Conforme pontua Escarião (2019, p. 172), as crianças “entram no mundo da linguagem, experimentam a língua, têm a possibilidade de ampliar o repertório linguístico, além de experimentar a essência dos gêneros que é a conversa, o diálogo”. Nesse sentido, é essencial para o bom desenvolvimento da linguagem que elas tenham a liberdade de se comunicar do seu jeito na interação com adultos ou outras crianças.

Por fim, pretende-se, com este trabalho, contribuir para a construção do conhecimento e a universalidade dos estudos empíricos de aquisição da linguagem, especialmente no seu aspecto multimodal. Desse modo, ressalta-se a importância da continuidade das pesquisas em matérias relacionadas à multimodalidade de gesto e fala e à atenção conjunta, considerando diferentes contextos e práticas interacionais.

## REFERÊNCIAS

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. Disponível em: <https://www.scielo.br/j/bak/a/dRS98pVJT4mJdmcc7JvkjyB/>. Acesso em: 4 abr. 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. Disponível em: <https://periodicos.ufrn.br/gelne/article/view/32212>. Acesso em: 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. Disponível em: [http://www.communicationcache.com/uploads/1/0/8/8/10887248/so\\_you\\_think\\_gestures\\_are\\_nonverbal.pdf](http://www.communicationcache.com/uploads/1/0/8/8/10887248/so_you_think_gestures_are_nonverbal.pdf). Acesso em: 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. Nova York: Cambridge University Press, 2016.

NÓBREGA, P. V. **Dialogia mãe-bebê**: a emergência do envelope multimodal em contextos de atenção conjunta. 2010. 168 f. Dissertação (Mestrado em Linguística e ensino) – Universidade Federal da Paraíba, João Pessoa, 2010.

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. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/actas/article/view/62399>. Acesso em: 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. Disponível em: <https://revista.abralin.org/index.php/abralin/article/view/2202>. Acesso em: 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***

---

- ☐ **Reconhecimentos:** Gostaria de agradecer à equipe editora da revista pelos serviços prestados com rapidez e excelência.
  - ☐ **Financiamento:** Não há fomento de instituições.
  - ☐ **Conflitos de interesse:** Não há conflitos de interesse.
  - ☐ **Aprovação ética:** O trabalho respeitou a ética durante a pesquisa. Não precisou passar por comitê de ética.
  - ☐ **Disponibilidade de dados e material:** Todos os dados de interesse do trabalho estão no próprio trabalho, além das referências bibliográficas. Os dados e materiais utilizados no trabalho são do acervo próprio do autor.
  - ☐ **Contribuições dos autores:** Autor 1 – Milena Magalhães Gomes: coleta de vídeo do acervo do autor; análise e interpretação dos dados, pesquisas bibliográficas e redação do texto; Autor 2 – Tatiana Ramalho Barbosa: análise e interpretação dos dados, pesquisas bibliográficas e redação do texto.
- 

**Processamento e editoração: Editora Ibero-Americana de Educação**  
Revisão, formatação, normalização e tradução

