

# Sincronia gesto-fala na emergência da fluência infantil

**Marianne Carvalho Bezerra Cavalcante**

Universidade Federal da Paraíba (UFPB), João Pessoa, Paraíba, Brasil  
marianne.cavalcante@gmail.com

**Andressa Toscano Moura de Caldas Barros de Almeida**

Universidade Federal da Paraíba (UFPB), João Pessoa, Paraíba, Brasil  
andressatoscano@yahoo.com

**Paulo Vinícius Ávila-Nóbrega**

Universidade Estadual da Paraíba (UEPB), João Pessoa, Paraíba, Brasil  
pvletras@yahoo.com.br

**Paula Michely Soares da Silva**

Universidade Federal da Paraíba (UFPB), João Pessoa, Paraíba, Brasil  
paula-michely@hotmail.com

**DOI:** <http://dx.doi.org/10.21165/el.v45i2.984>

## Resumo

Apresentaremos e discutiremos neste trabalho o conceito de multimodalidade inserido nas pesquisas em aquisição da linguagem. Entendemos, portanto, a multimodalidade como uma via de observação e análise da língua(gem) através de elementos multimodais que compõem esta (língua)gem, isto é, a fala, o gesto e o olhar (CAVALCANTE, 2009; ÁVILA-NÓBREGA, 2010; BARROS, 2012; CAVALCANTE et al. 2005). Tomamos como base teórica para as análises multimodais a classificação dos gestos proposta por McNeill (1985, 1992); a tipologia prosódica de Barros (2012) e a discussão sobre envelope multimodal de Ávila-Nóbrega e Cavalcante (2012a). Além disso, discutiremos como dispomos nossos dados para análise no *software* ELAN e apresentaremos dados de uma díade mãe-criança analisados sob esta perspectiva.

**Palavras-chave:** multimodalidade; aquisição da linguagem; gesto; sincronia; fluência.

## Gesture-Speech Synchrony in the Emergence of Children's Fluency

**Abstract:** This paper presents and discusses the concept of multimodality in language acquisition research. Multimodality is understood as a means of language analysis and observation that uses multimodal elements to form this language: speech, gesturing and gaze (CAVALCANTE, 2009; ÁVILA-NÓBREGA, 2010, BARROS, 2014; CAVALCANTE et al., 2015). The theoretical framework employed comprehends McNeill's (1985, 1992) gesture classification, Barros' (2012) prosodic typology and Avila-Nóbrega & Cavalcante's (2012) discussion about multimodal envelope. Furthermore, this paper presents a set of data including a mother-child dyad, and discusses how the software named as ELAN was used to organize and analyze the data.

**Keywords:** multimodality; language acquisition; gesture; synchrony; fluency.

## Introdução

Neste artigo, apresentaremos dados de aquisição da linguagem analisados por um viés multimodal. Diante disto, é necessário destacar que entendemos a linguagem como sendo multimodal, ou seja, é através das modalidades de uso da língua (fala, gesto, olhar) coatuando na produção linguística que teremos interação e consequentemente comunicação. Assim, são muitos os trabalhos em aquisição que têm se debruçado sob uma perspectiva multimodal de língua (CAVALCANTE, 1994; ÁVILA-NÓBREGA, 2010; BARROS, 2012; COSTA FILHO, 2011, FONTE, 2010), tendo como objetos de estudo a atenção conjunta, os gestos e a prosódia. Diante disso, faremos a seguir um apanhado teórico de cada um desses elementos multimodais.

## O gesto como elemento multimodal

Antes de começar esta seção, precisamos delimitar o que chamamos de gesto. McNeill (1985) e McNeill et al. (2002) dizem que a palavra gesto recobre uma multiplicidade de movimentos comunicativos, principalmente, mas não sempre, os de mãos e braços. Em um sentido mais amplo, o gesto inclui não só movimento de mãos, mas também expressão facial e troca de olhares (QUEK et al., 2006). Para este trabalho, privilegiaremos os movimentos de mãos e braços e assumimos o risco de não ter uma análise mais rica e completa por deixar outros fatores corpóreos de fora, embora assumamos que expressões faciais, troca de olhar, e movimentos de cabeça componham, também, uma gama variada de gestos.

Tendo dito isto, faremos agora um apanhado do que se tem na literatura a respeito dos gestos, começando com a nomenclatura proposta por Kendon (1982) e organizada por McNeill (1992, 2000) no que ele chama de *continuum* de Kendon. São eles:

GESTICULAÇÃO – GESTOS PREENCHEDORES<sup>1</sup> – EMBLEMAS – PANTOMIMAS – SINAIS

A gesticulação é descrita como aqueles gestos que acompanham o fluxo de fala. Sendo o tipo mais frequente no uso diário e cobrindo uma gama de usos e variedades, é produzido principalmente com os braços e as mãos, mas não é restrito a essas partes do corpo, a cabeça pode ser usada e as pernas e pés também podem aparecer como um gesto. A gesticulação combina tanto universais quanto traços linguísticos específicos de uma comunidade.

Os gestos preenchedores são descritos por McNeill (1992) como parte da sentença. O termo “speech-framed gestures” sugere um gesto que ocupa um lugar na sentença, preenchendo um espaço gramatical, ao invés de acompanhar o fluxo de fala, como a gesticulação. Já os emblemas são os sinais convencionalizados, tais como a mão fechada com o polegar para cima ou o sinal de OK. Estes são específicos da cultura, têm formas e significado padrão e variam de lugar para lugar.

As pantomimas são gestos ou seqüências de gestos que narram uma história, simulam uma ação ou objeto, produzidos sem fala. E os sinais, por sua vez, são palavras em uma língua de sinais como LIBRAS. As línguas de sinais têm sua própria estrutura linguística, incluindo padrões gramaticais, morfológicos, etc.

---

<sup>1</sup> Tradução nossa.

Temos, portanto, quatro instâncias no *continuum*: gesto-fala, gesto-propriedades linguísticas, gesto-convenções (se o gesto está presente ou não em dada cultura) e gesto-caráter semiótico (se o significado é determinado pelas partes ou pelo todo). Como podemos ver no quadro a seguir:

**Tabela 1.** Continuum de Kendon (MCNEILL, 2000)

	GESTICULAÇÃO	GESTOS PREENCHEDORES	EMBLEMAS	PANTOMIMAS	SINAIS
Gesto-fala	Presença obrigatória de fala	Presença obrigatória de fala	Presença opcional de fala	Ausência obrigatória de fala	Ausência obrigatória de fala
Gesto-propriedades linguísticas	Ausência de propriedades linguísticas	Presença de propriedades linguísticas	Presença de algumas propriedades linguísticas	Ausência de propriedades linguísticas	Presença de propriedades linguísticas
Gesto-convenções	Não convencional	Não convencional	Parcialmente convencionais	Não convencional	Totalmente convencional
Gesto-caráter semiótico	Global e sintético	Global e analítico	Segmentado e sintético	Global e analítico	Segmentado e analítico

Ao mover da esquerda para direita pelo *continuum* vemos que a obrigatoriedade de presença de fala diminui da gesticulação para os sinais. Assim, a gesticulação é obrigatoriamente acompanhada da fala, mas não tem propriedades linguísticas. Gestos preenchedores também têm presença de fala obrigatória, mas se relacionam com a fala de uma forma diferente, sequencialmente, ao invés de coocorrendo e têm papel linguístico específico. Os sinais são obrigatoriamente ausentes de fala, mas têm propriedades essenciais de uma língua.

A nomenclatura para gestos não termina no *continuum* de Kendon para McNeill. Em seus inúmeros trabalhos (1985, 1992, 2000, 2002) ele também expõe uma nomenclatura sua (gestos icônicos, metafóricos, dêiticos e *beats*<sup>2</sup>) para alguns gestos que estariam delimitados como gesticulação ou preenchedores no *continuum*.

Os gestos icônicos apresentam imagens de entidades e/ou objetos concretos. Tido então, como um símbolo referencial. Já os metafóricos não estão limitados à descrição de eventos concretos. Eles também podem apresentar conteúdo abstrato. Em um gesto metafórico, um significado abstrato é descrito como se tivesse forma ou ocupasse um espaço. Os dêiticos são prototípicos e o mais conhecido é o apontar, que identifica um objeto/entidade em discussão. Os *beats* são assim chamados porque a mão parece estar batendo ritmadamente, servindo como marcador da fala. É interessante destacar que essas dimensões são “fluidas” e podem ser achadas misturadas num mesmo gesto. Assim, *beats* combinam com apontar e muitos gestos icônicos são também dêiticos.

Tendo definido nossos dois objetos privilegiados para esta análise (prosódia e gesto) nossa pergunta agora é: como é possível unir gesto e fala numa mesma matriz de significação? É claro que eles pertencem a modalidades diferentes de expressão, mas estão ligados em alguns níveis e trabalham juntos para apresentar as mesmas unidades semânticas. As duas modalidades não são redundantes, mas sim coexpressivas, o que significa que elas dividem a mesma origem semântica, mas são capazes de expressar informações diferentes (MCNEILL et al., 2002).

<sup>2</sup> A não tradução deste se dá pela falta de um termo que possua a mesma carga semântica, pois a ideia que se tem é a de ritmo, batida, musicalidade.

McNeill (1985, 1992) diz que gesto e fala estão ligados a uma mesma matriz de significação e que não podem se dissociar. Nossa outra pergunta então é: por que não? Goldin-Meadow (2006) responde que quando o gesto é produzido isoladamente e assume todo o fardo da comunicação, ele toma forma de língua (como é o caso dos sinais no *continuum* de Kendon). Entretanto, quando o gesto é produzido em conjunto, dividindo o fardo da comunicação com a fala, ele toma uma forma não-segmentada, imagística, transmitindo informação que não é dita (na fala).

### **Breve histórico sobre os gestos**

Nesta seção faremos um breve levantamento sobre a história da comunicação não-verbal discutidos por Mark Knapp e Judith Hall (1999, p. 36-39). Este breve panorama nos apresentará o contexto de pesquisas em que Adam Kendon esteve inserido, o que pode ajudar-nos a entender melhor os pensamentos à época.

O interesse pela gestualidade emerge posteriormente à Segunda Guerra Mundial, o que não significa que antes desse período não houvesse interessados nesses estudos ou a produção de trabalhos a esse respeito. Por exemplo, a obra *Institutio Oratoria*, de Quintiliano, é considerada importante fonte a respeito dos estudos sobre os gestos e foi escrita no século I.

O foco na comunicação não-verbal nunca foi interesse de uma única área. Podemos encontrar menções a movimentos corporais, ou até mesmo pesquisas aplicadas em obras como, por exemplo, *A expressão das emoções em homens e animais*, de Charles Darwin, datada de 1872. Além da biologia, outros estudiosos da antropologia, linguística, filosofia, psiquiatria, psicologia, sociologia, ciência da estética e outras manifestaram seus interesses em compreender os sentidos que o corpo e seus movimentos podem apresentar.

Durante o início do século XX, foram feitas análises isoladas da voz, da aparência física, da vestimenta e da face. Pouca atenção era dada à proxêmica, ao meio ambiente e à cinestesia e menos ainda ao comportamento ocular e ao toque.

Em meados de 1950 houve um aumento significativo no número desses trabalhos. Antropólogos como Ray. L. Birdwhistell e Edward T. Hall foram responsáveis por aplicar alguns postulados da linguística aos fenômenos gestuais atribuindo novos rótulos aos estudos do corpo e do espaço e por iniciar um programa de investigação para cada área. Freud e outros terapeutas já haviam se interessado, antes dos anos 50, por sinais do corpo, das mãos e do rosto, mas a obra *Nonverbal communication: notes on the visual perception of human relations*, do psiquiatra Jurgen Ruesch e do fotógrafo Weldon Kees provavelmente foi a primeira a usar em seu título a expressão “comunicação não-verbal”. Essa publicação forneceu elementos teóricos adicionais quanto a origens, uso e codificação do comportamento corporal e mais uma extensa documentação visual do papel dos ambientes na comunicação.

Nos anos 1960, houve uma avalanche nos estudos na área e partes específicas do corpo eram objeto de longos programas de pesquisa. Um artigo de Paul Ekman e Wallace Friesen sobre origens, uso e codificação do comportamento não-verbal parece ter sido a peça teórica clássica desse período. O artigo em questão distinguia cinco áreas de estudo que abrangem a maior parte da atual pesquisa desses autores: emblemas, ilustradores, demonstrações de emoção, reguladores e adaptadores.

Na década de 1970, houve uma explosão de vendas de livros que influenciavam o público americano com conclusões errôneas, segundo Knapp e Hall (1999, p. 38), a respeito de como fazer uma venda, como obter um parceiro sexual, perceber fingimento, afirmar superioridade etc. Os leitores, com essas publicações, ficavam com a impressão de que a leitura de sinais corporais e faciais era a chave do sucesso em qualquer encontro humano e que nos tornaríamos totalmente transparentes se os nossos sinais fossem desvendados. No entanto, apesar do modismo americano desses estudos que se apresentavam no auge, estudiosos de renome procuraram desenvolver pesquisas confiáveis a respeito do comportamento corporal, ocular, facial, da pupila com o intuito de fazer uma espécie de compêndio.

Já por volta de 1980 alguns estudiosos continuaram a particularizar suas teses, enquanto outros se concentravam na identificação das maneiras pelas quais uma variedade de signos não-verbais atua em conjunto a fim de atingir objetivos comuns, como, por exemplo, levar alguém a fazer algo por você, mostrar afeto, mentir etc. Tornara-se claro que não poderíamos compreender inteiramente o papel dos sinais não-verbais na obtenção desses objetivos se não observássemos o comportamento verbal concomitante a esses sinais e tentássemos desenvolver teorias sobre como diversos sinais verbais e não-verbais interagem no processo interativo.

Sendo assim, após anos separando e isolando esses sinais microscopicamente, os estudiosos gradualmente perceberam que esses elementos funcionariam em um exercício conjunto. As pesquisas sobre a comunicação não-verbal seguem sendo modificadas das seguintes maneiras:

1. do estudo das situações não-interativas às interativas;
2. do estudo de uma pessoa a ambos os interlocutores;
3. do estudo de um único ponto do tempo a mudanças ao longo do tempo;
4. de estudo de comportamentos únicos a estudos multicomportamentais.

O novo perfil dos estudos sobre os gestos começa a amadurecer e articular outras discussões, inclusive unificando outros elementos da interação aos gestos. Nos Estados Unidos, encontramos publicações de alguns autores como David McNeill, que discute a relação multimodal da matriz gesto-fala, Charles Goodwin e seus trabalhos sobre gestos e afasia, Cynthia Butcher e Susan Goldin-Meadow com seus estudos sobre a produção de palavras com movimentos das mãos e Adam Kendon ao discutir o sistema de funcionamento dos gestos e da fala.

Na Europa, temos trabalhos na Suécia de John Laver e Janet Beck, que pesquisa a relação de elementos como voz, postura e gestos, na França podemos encontrar trabalhos de Isabelle Gaitella e Jacques Boyer a respeito da relação da voz e dos gestos.

Um forte grupo também tem se fundamentado em estudos multimodais e multissensoriais levando em consideração a interação entre adultos, entre adultos e computadores e entre adultos e crianças. O *The MARCS Institute* é desenvolvido na Universidade de Sidney e tem como coordenadores Denis Burnhan, Kate Stevens, Chris Daves, André van Schaik e Simeon Simoff.

## **Matriz gesto-fala em aquisição de linguagem**

Desde os anos de 1970, cientistas do desenvolvimento têm investigado associações entre desenvolvimento linguístico e alguns aspectos de atividades manuais, especialmente gestos comunicativos e simbólicos (BATES; DICK, 2002). Dessa forma, o estatuto do(s) gesto(s) em aquisição de linguagem vem sendo discutido por diversos autores (CAVALCANTE; 1994; ÁVILA-NÓBREGA, 2010; GOLDIN-MEADOW, 2006, 2009; FONTE; BARROS; CAVALCANTE; SOARES, 2014; CAVALCANTE; BARROS; SILVA; ÁVILA-NÓBREGA, 2015) e tem sido pauta de pesquisas atuais.

Para Goldin-Meadow (2009), gestos são atos comunicativos livres para tomar formas que a fala não pode assumir, ou, para crianças nos primeiros estágios de aquisição, formas que elas ainda não conseguem articular no discurso. Em um momento do desenvolvimento em que as crianças são limitadas ao que conseguem dizer, os gestos oferecem um caminho adicional de expressão, expandindo a gama de ideias que elas são capazes de expressar.

Sabemos que as crianças exploram a modalidade manual desde muito cedo e, assim sendo, os gestos fornecem um caminho para as primeiras palavras. É interessante perceber que as crianças raramente combinam gestos com outros gestos, mas frequentemente combinam gestos com palavras antes mesmo de começarem a combinar palavras com palavras (GOLDIN-MEADOW; MORFORD, 1985).

Butcher e Goldin-Meadow (2000) observaram três meninas e três meninos longitudinalmente durante a transição do estágio de uma palavra para o de duas palavras. A ideia era explorar se gestos comunicativos formam um único sistema em crianças, assim como formam em adultos. O que acharam foi que, inicialmente, os gestos tendiam a ser produzidos sem palavras, e, em raras ocasiões, quando produzidos com palavras, a fala era sem sentido e não sincronizada com o gesto. No entanto, as duas características que definem a integração gesto-fala em adultos, a coerência semântica e a temporalidade sincrônica foram achadas na comunicação infantil no momento e antes do período de duas palavras. Essa pesquisa vem colaborar com a nossa hipótese de que desde cedo gesto e fala estão unidos numa mesma matriz de significação.

Portanto, tomamos a matriz gestuo-vocal como nosso objeto de análise, por coatuar na linguagem engajando a criança na dialogia e guiando-a para a aquisição. Propomos então uma tipologia prosódica, considerando quatro momentos do funcionamento da fala na trajetória linguística infantil: balbucio, jargão, primeiras palavras, holófrases e blocos de enunciado (BARROS, 2012).

O balbucio é definido como a produção de sílabas que têm, tipicamente, o formato consoante vogal, por exemplo [ma, da, ba]. Tais sílabas são muitas vezes repetitivas e ritmadas (LOCKE, 1995). Neste artigo, consideramos o balbucio como uma produção vocal inicial da trajetória linguística infantil.

Os jargões são definidos como longas seqüências de sílabas que contêm padrões de acento e entonação variados e variáveis, que aparecem na fala infantil em torno dos 12 ou 13 meses de idade. Essas produções vocais soam como enunciados completos que carregam conteúdo de afirmações ou perguntas, ocorrendo, muitas vezes, concomitantes a palavras reais. No entanto, os jargões não apresentam conteúdo linguístico ou estrutura gramatical (DROMI, 2002).

Scarpa (2007) considera jargão quando o contorno entonacional se estende a uma cadeia de sílabas ou um longo fragmento composto por sílabas ininteligíveis. O balbucio tardio evolui para jargão quando a entonação é considerada mais madura e os contornos são preenchidos por sílabas tipicamente do balbucio, mas reconhecíveis como intenção comunicativa pelos adultos, que sempre lhe atribuem significado de uma frase ou sentença.

O terceiro momento do funcionamento da fala definido nessa proposta é o das primeiras palavras reconhecíveis. Assim, consideramos nesse período as produções infantis contendo enunciados de uma palavra, que já nem são balbucios nem puramente jargões, visto que as consideramos reconhecíveis na língua adulta e interpretáveis pelo interlocutor.

As produções das primeiras palavras da criança indicam mudança nos contornos entonacionais, que foram constituídos por mais variações de altura. Essas variações de altura são reconhecidas como diferentes intenções comunicativas: pergunta, afirmação, pedido etc. e são recorrentes e produtivas e não mais ou menos efêmeras como os balbucios e jargões. Nessas primeiras palavras, encontramos sequências mais curtas em comparação às do jargão e um padrão silábico reconhecido como fazendo parte de um léxico primitivo.

Definimos o período que chamamos de blocos de enunciado como o momento em que a criança alterna a produção de holófrases com enunciados completos. Notamos em nossos dados que nesse momento do processo aquisicional a criança já é capaz de fazer pedidos, perguntas e produzir respostas mais longas com significado completo, superando os enunciados holofrásticos. Notamos que a partir de um ano e meio a criança começa a arriscar-se nesses enunciados mais longos, juntando duas ou mais holófrases.

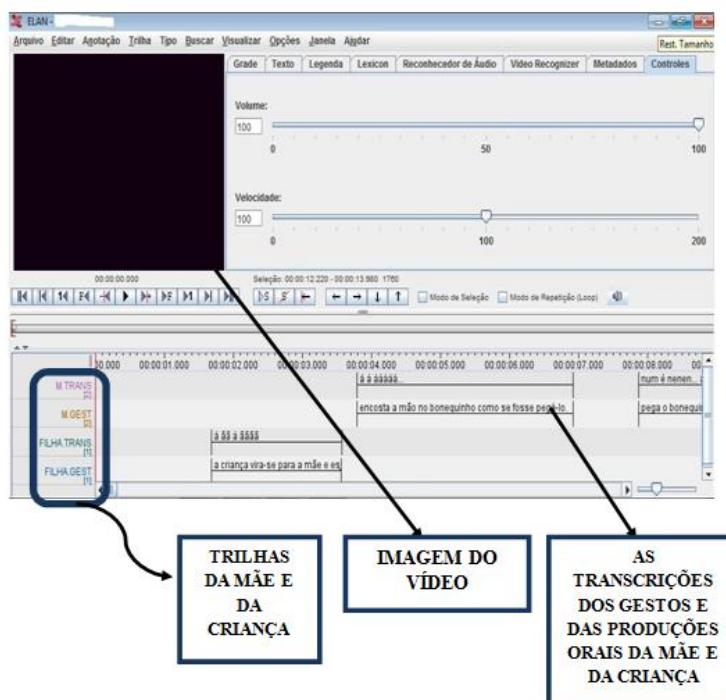
## **Metodologia de transcrição e análise numa proposta multimodal**

Para observarmos os aspectos multimodais, utilizamos para nossas transcrições e análise dos dados o programa ELAN (*Eudico Linguistic Annotator*), que possibilita a criação de anotações, edição, visualização e busca de anotações através de dados de vídeo e áudio simultaneamente. Além disso, ele permite a transcrição e anotações das análises em linhas denominadas de trilhas. A criação dessas trilhas e suas nomeações são determinadas pelo pesquisador/transcritor. Essas trilhas permitem fazer anotações de determinado registro no tempo exato, e se for necessário fazer alguma alteração isso poderá ser feito sem nenhuma perda de anotações anteriores ou subsequentes.

O ELAN é um *software* que traz vantagens para qualquer pessoa que necessite utilizá-lo, sendo ainda um *software* gratuito e que não apresenta dificuldades no manuseio, além de apresentar vantagens no processo de transcrição e organização de dados. Proporciona ainda ao pesquisador a elaboração de novos quadros com mesclas das trilhas. As trilhas geradas pelo programa a partir das transcrições podem ser mescladas, criando novos quadros, denominadas neste trabalho “quadros de mesclas”. As trilhas a serem mescladas são escolhidas de acordo com a necessidade/foco da análise e permitem a visualização de componentes multimodais de maneira concomitante.

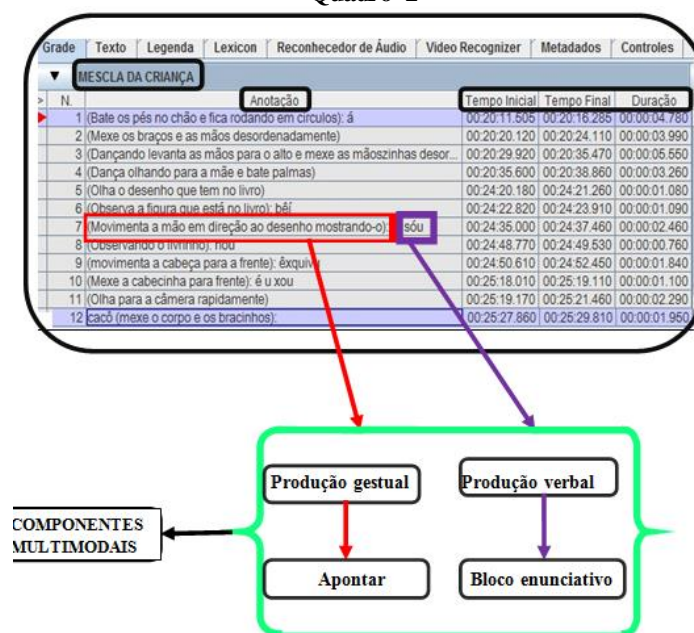
Além disso, procuramos ilustrar via ELAN os momentos em que os gestos e a produção verbal estão presentes nos momentos de interação, corroborando assim com a base teórica deste trabalho. Vejamos abaixo a ilustração do programa ELAN (quadro 1).

Quadro 1



Para materialização da transcrição multimodal – o envelope multimodal – exemplificamos com a criação de trilhas: trilha da produção gestual da mãe “Gest.MÃE”, trilha da produção verbal materna intitulada no programa e no nosso trabalho como “Fala. MÃE”, criamos também a trilha das produções gestuais da criança “Gest. BEBÊ” e trilha da produção vocal da criança intitulada como “Fala.BEBÊ”.

Quadro 2



Os componentes multimodais (quadro 2 acima) organizados em trilhas compõem aquilo que denominamos de envelope multimodal, quando ocorre a mescla das trilhas – gestos, produção vocal/oral, olhar, prosódia (ÁVILA NOBREGA; CAVALCANTE, 2012a; 2012b).



## Sincronia gesto fala: uma perspectiva de análise

O estabelecimento de interação colaborativa se constrói a partir da percepção mútua das ações visíveis do outro. Assim, movimentos faciais, sensibilidade à orientação e direção do olhar, aspectos de orientação e de postura do comportamento interacional possibilitam uma extrema rapidez e sutileza na interação entre pares, que caracterizam a interação humana (KENDON, 2009).

Neste sentido, todo enunciado linguístico emprega, de forma integrada, padrões de vocalização e entonação, pausas e ritmicidades, que se manifestam não só de forma audível, mas cineticamente também, e sempre, como uma parte desta, existem movimentos dos olhos, das pálpebras, das sobrancelhas, bem como da boca, e os padrões de ação por parte da cabeça (KENDON, 2009). Tal integração foi historicamente deixada à margem dos estudos linguísticos e num movimento de fins da década de 90 do século XX, vem se materializando a partir de trabalhos de autores como Kendon<sup>3</sup> (1981, 1992, 2004, 2009), McNeill (1985, 2000, 2008), dentre outros.

Partindo das macrocategorias gestuais discutidas por Kendon (1981), a saber: gesticulação, pantomima e emblemas, já mostradas no *continuum* de Kendon acima, McNeill apresenta 04 tipologias gestuais: os gestos icônicos, metafóricos, dêiticos e *beats*<sup>4</sup>. Os *gestos icônicos* estão estreitamente ligados ao discurso, servindo para ilustrar o que está sendo dito, por exemplo, quando uma pessoa demonstra um objeto físico usando as mãos para mostrar seu tamanho; os *gestos metafóricos* são parecidos em sua superfície com os gestos icônicos, contudo, possuem a particularidade de referirem expressões abstratas; os *gestos dêiticos* são os demonstrativos ou direcionais, geralmente acompanham as palavras como “aqui”, “lá”, “isto”, “eu” e “você”, pode ser representado pelos movimentos de apontar; os *gestos ritmados (beats)* são nomeados assim porque aparecem como o tempo da batida musical; as mãos se movem no mesmo ritmo da pulsação da fala (cf. MCNEILL, 1992, p. 15-18).

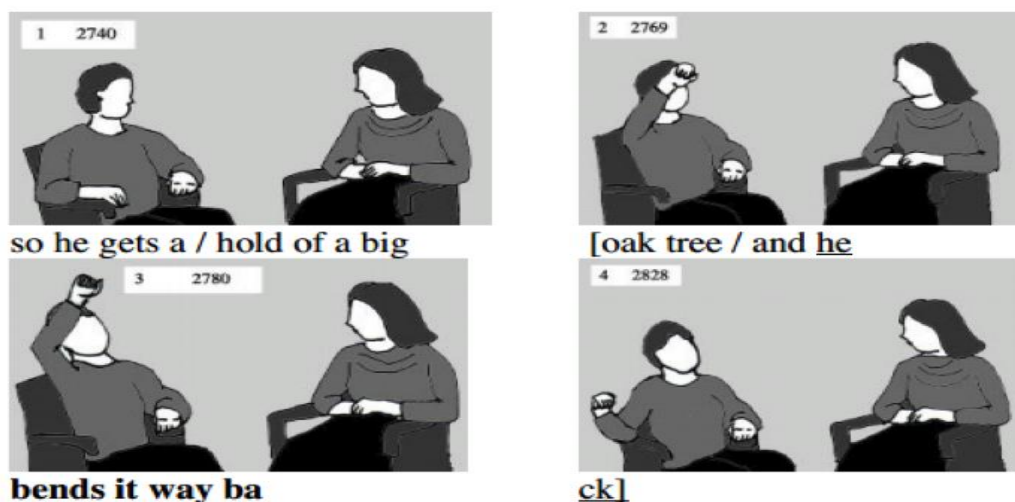
Além dessa tipologia, os gestos passam por uma série de fases, cada uma com sua própria posição e função no gesto: as *unidades gestuais*, *frases gestuais* e *fases gestuais*. Uma *unidade gestual* é o intervalo entre descansos sucessivos dos membros. A *frase gestual* consiste de até 5 fases gestuais: preparação, curso (*stroke*), retração, pausa pré-curso e pós-curso. As *fases gestuais* organizam-se em função do curso (cf. MCNEILL, 1992, p. 15-18). A seguir, uma ilustração da frase gestual:

---

<sup>3</sup> Para uma discussão acerca do *continuum* de Kendon e sua classificação ver Cavalcante et al. (2015).

<sup>4</sup> A não tradução deste termo se dá pela falta de um termo que possua a mesma carga semântica, pois a ideia que se tem é a de ritmo, batida, musicalidade.

so he gets a / hold of a big [oak tree / and he bends it way back]



Legenda: 1. pré-preparação; 2. sustentação pré-curso; 3. Meio do curso; 4. final do curso e início da sustentação pós-curso (na metade de “back”)

A combinação síncrona da fala com o gesto pode ser chamada de *ponto de saliência*<sup>5</sup> (GP), tal como proposto por McNeill (1992). É chamado GP porque caracteriza os picos de congruência de uma unidade de “GESTO-FALA” e sua estruturação cognitiva<sup>6</sup>, aqui neste artigo consideraremos como momentos de sincronia, em que coincide a produção do gesto e da produção vocal, que será caracterizada adiante na análise.

Na produção do discurso há gestos com características recorrentes, o que chamamos *captações/repetições*. Através das *captações* usadas pelo falante, percebemos as *combinações de unidades de discurso*. O *mimentismo* é uma dessas características, constitui-se como o recorte/repetição do gesto do interlocutor na mesma *fase gestual* em que comparecem unidades gestuais (enunciados discursivos) consecutivas ou co-ocorrentes (MCNEILL, 1992).

<sup>5</sup> *Growth point*.

<sup>6</sup> Há uma longa discussão acerca da estruturação cognitiva dos GP que não será objeto deste artigo.

## Exemplos ilustrativos de dados multimodais com sincronia gesto-fala<sup>7</sup>

**Cena 1:** Mãe e criança (C – 11 meses e 13 dias) interagindo no quarto da criança. C está sentada na cama e volta-se para a pesquisadora que está filmando e faz um comentário divertido, “como se” narrasse algo



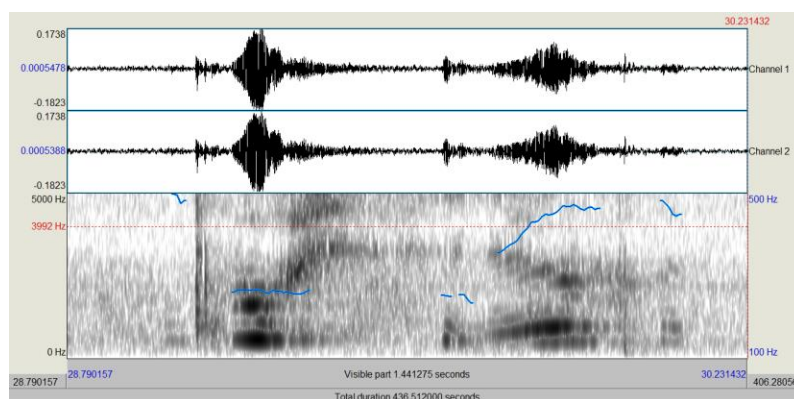
Fase 1



Fase 2



Fase 3



**Espectrograma 1**

Nesta cena, é possível perceber a matriz multimodal propiciando uma fala fluente<sup>8</sup>. A criança eleva ambos os braços para cima e para baixo e produz uma fala jargonizada “**ebuze::embuze::**” (apresentada no espectrograma), com a presença de contornos típicos de narrativa tom médio/alto, depois alto com leve queda. Em sincronia, inclusive, com a gesticulação braçal, que traz a presença de gestos metafóricos. Além disso, durante a produção vocal não há presença de hesitações e/ou pausas, constituindo-se assim tal como Scarpa (1988) define, com blocos prosódicos organizados.

A unidade gestual, composta de 03 fases gestuais, apresentada na cena 1, pode ser assim descrita:

<sup>7</sup> Esses dados ilustrativos foram publicados em Cavalcante; Barros; Ávila Nóbrega; Silva (2015). São resultados de filmagens longitudinais naturalísticas de interação mãe-criança ao longo de 24 meses de vida da criança, com intervalos de 15 dias entre as filmagens. As transcrições e análises foram feitas com o uso dos *softwares* Praat e Elan e compõem o banco de dados do LAFE (Laboratório de Aquisição da Fala e da Escrita) da UFPB.

<sup>8</sup> “Em relação à fluência, adotamos a perspectiva de Scarpa (1985) que concebe os trechos de fala fluente como os já conhecidos, arrumados, que vêm em bloco e os disfluente como aqueles instáveis, com tentativas pouco frutíferas de segmentação de blocos prosódicos. Destacamos o papel da gesticulação como uma pista importante da fluência infantil. Concebemos que a mesma se estrutura em estágios anteriores ao reconhecimento ao uso dos blocos de enunciados, principalmente na presença dos jargões articulados à gesticulação como mostraremos em nossas análises” (CAVALCANTE et al., 2015).

Fase 1 – preparação: braços para baixo, laterais ao corpo, braço esquerdo levemente elevado, face com expressão de esboço de sorriso com lábios semicerrados e olhar levemente desviado para direita;

Fase 2 – curso: dois braços suspensos, laterais ao rosto, com mãos semiespalmadas na altura da face, mão esquerda um pouco mais alta do que a direita; lábios entreabertos na produção do jargão “**ebuze::embuze::**”, face em sorriso, olhar lateral para direita;

Fases 3 – retração, pausa: braços em movimento descente, voltando a lateral do corpo na altura da cintura, braço esquerdo mais baixo do que o direito que vai em movimento descendente mais lento do que o esquerdo, tronco e cabeça direcionados para direita, olhar direcionado para direita, face com esboço de sorriso, lábios semicerrados.

A sincronia gesto-fala ocorre na fase 2, no momento de produção do jargão “**ebuze::embuze::**”, aliado ao gesto metafórico descrito na fase 2.

Cena 2: Mãe e criança (C- 22 meses e 10 dias) interagindo na sala de visitas da casa. Enquanto conversam, a mãe propõe que cantem músicas infantis e C começa a cantar a música de Chapeuzinho Vermelho.



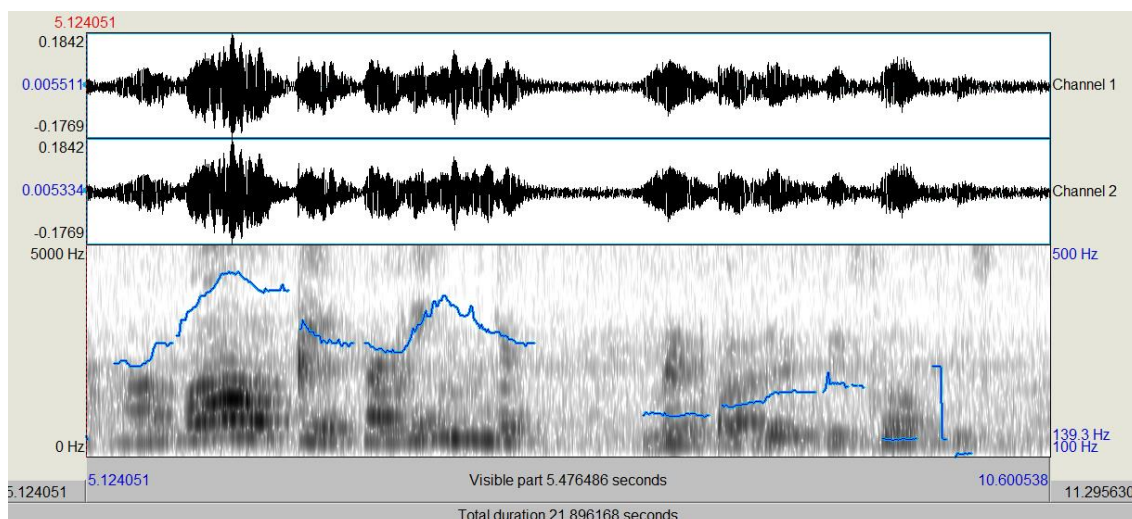
Fase 1



Fase 2



Fase 3



Espectrograma 2

Nesta cena, é interessante destacar, novamente, a matriz multimodal contribuindo para fluência, onde gesticulação – braços elevados ilustrando o ritmo da música da Chapeuzinho Vermelho, com a presença dos gestos denominados *beats*. A produção musical compõe-se de um bloco prosódico melódico da cantiga popular sustentando pelo fragmento melódico “pela estrada afora eu vou bem sozinha, levar esses doces para vovozinha...”. Na cena, a criança eleva os braços e acompanha a melodia produzida pela mãe, enquanto esboça a produção de jargão/holófrase – “**ãn::zinha**”, para se referir ao trecho “para vovozinha...”.

A unidade gestual de C, composta de 04 fases gestuais, apresentada na cena 2 pode ser assim descrita:

Fase 1 – preparação: braços elevados semifletidos, laterais ao corpo na altura do abdômen próximos um do outro, braço esquerdo levemente elevado em relação ao direito, mãos semifechadas, quase juntas, face tensa com olhar voltado para os braços e mãos que estão à sua frente;

Fase 2 – curso: dois braços elevados semifletidos suspensos e abertos, perpendiculares ao corpo, postura ereta, mãos semifechadas, olhar direcionado para frente, face neutra;

Fase 3 – meio do curso: braços elevados semifletidos juntos à frente do rosto, perpendiculares ao corpo, postura ereta, mãos semifechadas, mãos juntas, tocando uma na outra, olhar direcionado para frente, face neutra, enquanto movimenta os lábios e produz o jargão/holófrase “**ãn::zinha**”;

Fase 4 – retração: braços em movimento descendente, voltando a lateral do corpo na altura da cintura, braço esquerdo mais baixo do que o direito.

A sincronia gesto-fala ocorre na fase 3, no momento de produção do jargão “**ãn::zinha**”, aliado ao *beat* descrito nesta fase.

Cena 3: Mãe e criança (C – 24 meses e 03 dias) estão no corredor, perto da cozinha. É a hora do lanche de C e a mãe está ajudando para que ela coma as bananas que estão no prato. Ela tem dificuldades em garfar as bananas e escolhe as que quer pegar. Quando a mãe vai ajudá-la, ela aponta qual banana ela quer.



Fase 1

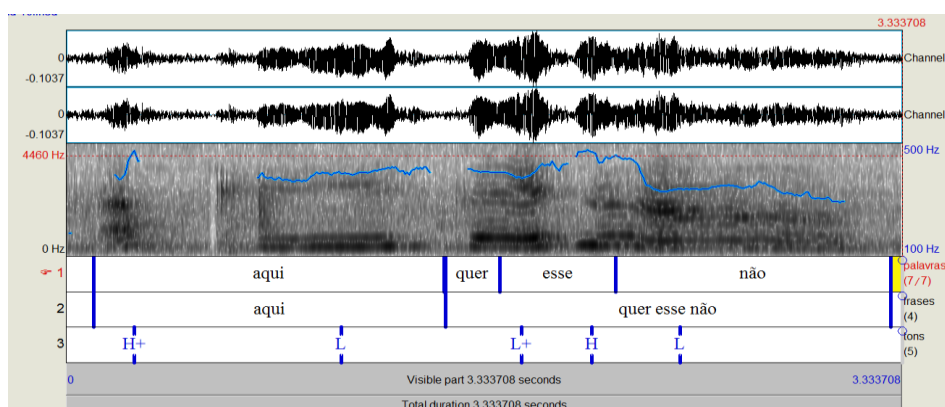


Fase 2



Fase 3

Na cena mostrada acima, gesto e fala formam uma unidade semântica coerente, pois a criança quer um pedaço de banana específico e aponta para ele a fim de obtê-lo enquanto produz “**aqui/ quer esse não**”, indicando à mãe qual ela quer e olhando para o objeto em discussão. Vejamos a produção no Praat:



### Espectrograma 3

Essa produção da criança se caracteriza como um bloco de enunciado, pois não temos uma ou mais palavras isoladas, e sim um contínuo prosódico, além de um todo

semântico. Percebemos que nesse recorte, diferente do primeiro, ela já consegue formar um enunciado completo, então, não podemos deixar de destacar que a matriz gesto-fala está ainda mais consolidada e servindo a um propósito comunicativo.

A unidade gestual de C, composta de 04 fases gestuais, apresentada na cena 2 pode ser assim descrita:

Fase 1 – preparação: braço direito semifletido com mão pegando o garfo entregue pela mãe, tocando nas rodela de banana, dedo indicador para frente, olhar direcionado para as rodela de banana que o garfo toca, face tensa;

Fase 2 – curso: braço direito semifletido com mão segurando o garfo e tocando nas rodela de banana, braço esquerdo tocando o tronco acima do peito, olhar direcionado para a rodela de banana que o garfo toca, face tensa;

Fase 3 – meio do curso: braço direito suspenso à frente do corpo, semifletido, com mão produzindo o gesto de apontar com o dedo indicador apontado para baixo em direção à uma rodela de banana específica, gesto dêitico, olhar direcionado para a rodela de banana, olhar direcionado para a rodela de banana apontada, face tensa, enquanto produz o bloco de enunciado “**aqui/ quer esse não**”.

Fase 4 – retração: braço direito em movimento descendente, voltando a lateral do corpo na altura da cintura.

A sincronia gesto-fala ocorre na fase 3, no momento de produção do bloco de enunciado “**aqui/ quer esse não**”, aliado ao gesto dêitico descrito nessa fase.

## Considerações Finais

Neste artigo apresentamos uma breve introdução aos estudos de multimodalidade tendo como referência a matriz gesto-fala tal como propõem Kendon (1982) e McNeill (1985, 1992, 2000), quando defendem o caráter concomitante e sincrônico dessas instâncias multimodais. A partir dessa noção basilar, construímos uma proposta de análise de dados em aquisição da linguagem, em contextos interativos entre mãe e criança, apresentando um modelo metodológico de transcrição e análise que se compõem de trilhas de componentes/modalidades – gesto, olhar, produção vocal, prosódia – que vão se estruturando como um envelope multimodal e dando ritmo e cadência às interações entre o adulto e a criança de modo a possibilitar um olhar mais acurado acerca do fenômeno da aquisição da linguagem.

Buscamos mostrar como a instância multimodal (gesto e fala) funciona em crianças em fase de aquisição de linguagem. Para que isso fosse possível, lançamos mão de uma tipologia gestual proposta por McNeill (1992) e a adaptamos para as crianças, visto que as pesquisas deles se voltam para os adultos.

Baseamo-nos também em conceitos prosódicos para que análises acústicas da fala pudessem ser feitas e utilizamos a proposta de trabalhos anteriores (BARROS, 2012; FONTE, R.; BARROS, A. et al., 2014) de estágios de desenvolvimento entonacional, a fim de observar se mesmo em estágios iniciais a criança já trabalha com a matriz gesto-fala. Nossos dados mostraram, portanto, que, mesmo ainda pequena, a criança faz uso de instâncias multimodais para fins comunicativos, corroborando então com os dados apresentados por Butcher e Goldin-Meadow (2000) em que crianças apresentaram gesto e fala integrados numa mesma matriz de significação.

Como vimos, a sincronia faz parte do processo aquisicional e é composta de uma matriz multimodal colaborativa gestuo-vocal que constituirá a matriz linguística da fala infantil. Compreender a sincronia, nessa perspectiva, permite entre outras coisas considerar o gesto como co-partícipe da produção vocal, que aliados promovem aquilo

que denominamos discurso fluente. Nesse sentido, quando vislumbramos as terapias de linguagem que têm como foco a disfluência infantil, devemos levar em consideração essa matriz e a precocidade de sua presença nas produções infantis, sustentadas nas interações com o adulto.

## REFERÊNCIAS

ÁVILA-NÓBREGA, P. V. *Dialogia mãe-bebê: a emergência do envelope multimodal em contextos de atenção conjunta*. 2010. 190 f. Dissertação (Mestrado em Linguística) – Universidade Federal da Paraíba, João Pessoa.

ÁVILA-NÓBREGA, P. V.; CAVALCANTE, M. C. B. Aquisição de linguagem em contextos de atenção conjunta: o envelope multimodal em foco. *Revista Signótica*. v.24, n.2, p. 469-491, jul./dez. 2012a.

\_\_\_\_\_. Aquisição de linguagem e dialogia mãe-bebê: o envelope multimodal em foco em contexto de atenção conjunta. *Revista Investigações*, v.25, n.2, p. 157-183. jul. 2012b.

BARROS, A. T. M. de C. *Fala inicial e prosódia: do balbucio aos blocos de enunciados*. 2012. 180 f. Dissertação (Mestrado em Linguística) – Universidade Federal da Paraíba, João Pessoa.

BRANDÃO, L. W. P. *A fala materna dirigida ao bebê surdo implantado: entre o “ouvinte suposto” e o “aprendiz de ouvinte”*. 2010. 280 f. Tese (Doutorado em Linguística) – Universidade Federal da Paraíba, João Pessoa.

CAVALCANTE, M. C. B. Rotinas interativas mãe-bebê: constituindo gêneros do discurso. *Investigações*, Recife, p. 153-170, 2009.

CAVALCANTE, M. C. B. O gesto de apontar como processo de co-construção nas interações mãe-criança. 1994. 294 f. Dissertação (Mestrado em Linguística) – Universidade Federal de Pernambuco, Recife.

CAVALCANTE, M. C. B.; BARROS, A. T. M. de C.; ÁVILA NÓBREGA, P. V.; SILVA, P. M. S. da. Gestualidade como uma pista importante da fluência infantil. *Revista ProLíngua*, João Pessoa, v.10, n. esp., p. 43-50, Editora da UFPB, 2015.

COSTA FILHO, J. M. S. da. *“Olá, Pocoyo!”: a constituição da atenção conjunta infantil com o desenho animado*. 2011. 200 f. Dissertação (Mestrado em Linguística) – Universidade Federal da Paraíba, João Pessoa.

DROMI, E. *Babbling and early words*. In: SALKIND, N.J. (Ed.). *Child development*. Macmillan psychology reference series. Mcmillan, 2002.

FONTE, R.; BARROS, A.; CAVALCANTE, M.; SOARES, P. A matriz gesto-fala na aquisição da linguagem: algumas reflexões. In: \_\_\_\_\_. *Aquisição, desvios e práticas de linguagem*. Editora CRV, 2014.

FONTE, R. F. L. da. *O funcionamento da atenção conjunta na interação mãe-criança cega*. 2010. 311 f. Tese (Doutorado em Linguística) – Universidade Federal da Paraíba, João Pessoa.

GOLDIN-MEADOW, S.; MORFORD, M. Comprehension and production of gestures in combination with speech in one-word speakers\*. *J. Child lang.* 19, p. 559-580, 1992.

KENDON, A. The study of gesture: some observations on its history. *Recherches Semiotique/Semiotic Inquiry*, 2 (1), p. 25-62, 1982.

\_\_\_\_\_. Language and gesture: unity or duality? In: MCNEILL, D. (Ed.). *Language and Gesture*. Cambridge: Cambridge University Press, 2000. p. 47-63.

KNAPP, M. L., HALL, J. A. *Comunicação não-verbal na interação humana*. São Paulo: JSN, 1999. 948 p.

LOCKE, J. L. Desenvolvimento da capacidade para a linguagem falada. In: FLETCHER, P.; MACWHINNEY, B. (Eds.) *Compêndio da Linguagem da Criança*. Tradução de M. A. G. Domingues. Artes Médicas. Porto Alegre, 1995. 632 p.

LOPES, J. C. M. *Dinâmicas dialógicas singulares: a multimodalidade na criança com autismo*. 2011. 281 f. Tese (Doutorado em Linguística) – Universidade Federal da Paraíba, João Pessoa.

MCNEILL, D. So you think gestures are nonverbal? *Psychological Review*, v.92 (3), p. 350-371, 1985.

\_\_\_\_\_. *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press, 1992. 423 p.

\_\_\_\_\_. Introduction. In: MCNEILL, D. (Ed.). *Language and gesture*. Cambridge: Cambridge University Press, 2000. p. 1-10.

MCNEILL, D.; QUEK, F.; BRYLL, R.; DUNCAN, S. Multimodal human discourse: gesture and speech. *ACM Transactions on Computer-Human Interaction*, v.9, n.3, p. 171-193, sept. 2002.

MELO, E. S. de. *Gestos pantomímicos e produção verbal na Aquisição da Linguagem*. 2014. 45 f. Monografia (Graduação em Letras). – Universidade Federal da Paraíba, João Pessoa.

QUEK, F.; MCNEILL, D.; BRYLL, R. et al. Multimodal human discourse: gesture and speech. *ACM transactions on computer-human interactions*, v. 9, n. 3, sept. p. 171-193, 2002.

SCARPA, E. M. A Aquisição da prosódia: dupla face, dupla vocação. In: AGUIAR, M. A. M.; MADEIRO, F. (Orgs). *Em-Tom-Ação: a prosódia em perspectiva*. Recife: Editora Universitária da UFPE, 2007. 251 p.

\_\_\_\_\_. Desenvolvimento da Intonação e A Organização da Fala Inicial. *Cadernos de Estudos Linguísticos*, UNICAMP, n.14, p. 65-84, 1988.

SCLIAR-CABRAL, L. Evolução das pesquisas em aquisição da linguagem oral monolíngue no Brasil. In: FINGER, I.; QUADROS, R. M. (Org.). *Teorias de Aquisição de Linguagem*. Santa Catarina: Editora da UFSC, 2008, 304 p.

SILVA, P. M. S. *Gestos e produção verbal: a fluência multimodal em aquisição da linguagem*. 2014. 172 f. Dissertação (Mestrado em Linguística) – Universidade Federal da Paraíba, João Pessoa.

TOMASELLO, M. *Origens culturais da aquisição do conhecimento humano*. São Paulo: Martins Fontes, 2003. 342 p.

**Recebido em:** 06/03/2016

**Aprovado em:** 26/04/2016