

O alinhamento entre unidades prosódicas não terminais e frases gestuais

Implicações da Teoria da Língua em Ato

Camila Barros

Faculdade de Letras
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil
camila-ab@ufmg.br

Heliana Mello

Faculdade de Letras
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil
hmello@ufmg.br

Abstract: A relação entre gestos e prosódia pode ser estudada por meio de uma série de perspectivas teóricas e metodológicas que buscam maneiras de ancorá-la temporal e funcionalmente. O objetivo deste trabalho é apresentar um estudo de corpus que analisou qualitativa e quantitativamente unidades prosódicas não terminais, dentro do enquadre teórico da Teoria da Língua em Ato (CRESTI, 2000; MONEGLIA, RASO, 2014), e frases gestuais (KENDON, 2004). A análise qualitativa ateu-se ao estudo da unidade informacional de Parentético, a qual insere um comentário sobre o conteúdo veiculado no restante do enunciado (TUCCI, 2009). A análise quantitativa enfocou o alinhamento das fronteiras iniciais e finais das unidades prosódicas não terminais e fronteiras gestuais a fim de elucidar qual a relação que elas estabelecem entre si. Os resultados apontam para uma paridade entre uma ruptura do padrão prosódico e gestual durante a realização da unidade de Parentético e para um alto acordo entre unidades não terminais e frases gestuais, que distam em torno de 500 ms para fronteiras iniciais e finais, variando de acordo com a unidade informacional que veiculam.

Gestos; Prosódia; Alinhamento entre gestos e prosódia; Parentético;

I. INTRODUÇÃO

O entendimento de gestos como uma parte constitutiva da comunicação linguística, que reflete processos cognitivos, é possivelmente o ponto de concordância mais fundamental e pacífico sobre gestos concomitantes com a fala (*co-speech gestures*). No entanto, a maneira com que a produção da fala e a produção de gestos estão ligadas entre si ainda é controversa e permanece sob constante escrutínio [1, 2]. O ponto específico que este texto pretende abordar se refere ao alinhamento entre unidades prosódicas não terminais e frases gestuais, buscando entender quais os limites da sincronia entre fala e gestos em termos temporais e funcionais. A paridade entre essas unidades é necessária para que se possa entender como unidades informacionais se relacionam com gestos.

Muito da literatura da área se debruça sobre quais são os pontos de ancoragem de fala e gestos, com base nos picos prosódicos e picos de esforço dos movimentos manuais [3–8], a fim de entender como se dá o processamento e produção de ambos. No entanto, é imprescindível também entender o que ocorre em um nível intermediário de análise, isto é, em um nível que não tenha uma granularidade tão fina quanto os picos de esforço e nem uma granularidade tão grande como os turnos de fala. O presente trabalho adota parâmetros teóricos-metodológicos que permitem elucidar como unidades não terminais de fala e gesto estão funcional e temporalmente alinhadas e, com isso, constroem conjuntamente a mensagem, veiculada na fala como unidades informacionais [9] e, nos gestos, como determinados padrões representacionais [10].

Esse trabalho foi elaborado com apoio e financiamento da Fapemig.

Para isso, foi compilado o BGEST corpus [11], um corpus multimodal que integra os corpora do projeto C-ORAL-BRASIL; e que segue os seus mesmos princípios metodológicos de forma a que fosse comparável aos demais corpora de fala da família C-ORAL [12, 13]. Com isso, foi possível analisar tanto as unidades prosódicas não terminais em sua relação com frases gestuais, quanto as unidades informacionais por elas veiculadas.

II. REVISÃO DE LITERATURA

Grosso modo, gestos são movimentos de mãos e braços que coocorrem com a fala [14]. Esses movimentos não são aleatórios ou desconexos, mas sim compõem com a fala a mensagem, trazendo principalmente nuances imagéticas que não são codificadas verbalmente [3, 8, 10]. Outro ponto crucial do entendimento de gestos concomitantes com a fala é que eles são produto de um processo cognitivo que está fortemente ligado com a produção linguística. Assim, muitos estudos sobre a gestualidade se debruçam sobre a investigação de quais processos engatilham, ancoram e executam os gestos.¹ Este trabalho se enquadra entre aqueles que buscam a ancoragem da gestualidade. A ancoragem dos gestos na fala se apoia na segmentação, que se divide em três níveis: unidade, frase e fases gestuais [4, 15]. O nível de análise mais baixo é o das fases gestuais, que compreende o pico de esforço que carrega a expressão do gesto (*stroke* ou ataque)², sua preparação e repouso [14]. Cada ataque realizado constitui a fase necessária e suficiente para a realização do gesto, sendo entendida como completa. Essa fase pode ser acompanhada pela preparação e repouso, constituindo as frases gestuais, que estão em um nível intermediário de análise. São as frases e frases gestuais os construtos que tradicionalmente são estudados em conjunto com a fala, pareados com unidades prosódicas não terminadas [4, 6, 16]. O nível analítico mais alto é o das unidades gestuais, compreendendo todos os gestos que são realizados entre uma posição de repouso, em que o falante descansa as mãos (normalmente sobre as pernas, caso esteja assentado).

A título de exemplificação, a Figura (1) abaixo indica uma unidade gestual composta por uma única frase gestual; cada um dos frames destacados indica as fases de preparação, ataque e repouso, respectivamente.

Figura 1. Curso do gesto

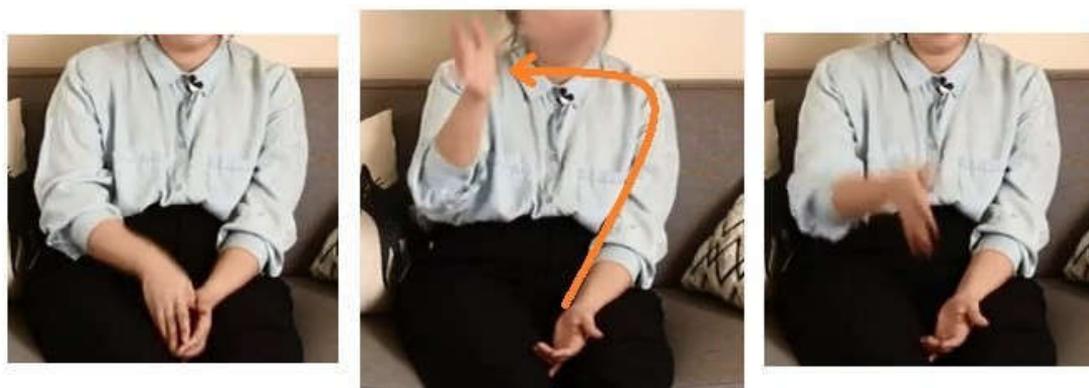


Figura retirada do corpus BGEST [11].

¹ Para uma discussão sobre qual o estado da arte das diferentes abordagens que tentam abordar essas relações, cf. [1].

² Todas os termos técnicos que foram considerados pertinentes por terem diferentes traduções em português brasileiro foram dados com o nome original e a tradução sugerida em [11].

Para as unidades informacionais, adota-se a perspectiva da Teoria da Língua em Ato [9, 17] (doravante, L-AcT). A L-AcT aponta que o fluxo da fala pode ser estudado informacionalmente, adotando-se como unidade básica de referência, as unidades prosódica e pragmaticamente autônomas, os enunciados. Essas unidades são delimitadas por uma fronteira prosódica terminal e veiculam um ato de fala [18]. Sequências terminadas podem ser compostas por um único ato de fala (enunciado) ou, menos comumente, por vários atos de fala justapostos e separados por fronteiras não terminais (estrofe). Por razões de espaço, só será discutido em detalhe aqui o enunciado. Ele pode ser simples, quando não é segmentado internamente por fronteiras não terminais, ou complexo, quando é internamente segmentado por fronteiras não terminais. Essas unidades não terminais, que constituem o enunciado complexo, emolduram a locução veiculada pela unidade de Comentário, tornando-a pragmaticamente apropriada e a compõem textualmente. A L-AcT prediz que existe um isomorfismo entre a unidade informacional e a forma prosódica, o que significa que a realização prosódica e a distribuição dentro do enunciado vão atribuir diferentes valores informacionais à unidade não terminada. No exemplo abaixo, temos três exemplos de enunciados. O exemplo (1) é um enunciado simples, constituído apenas pelo Comentário. Os demais são enunciados complexos. O exemplo (2) é uma estrutura de Tópico-Comentário, em que a primeira unidade estabelece o domínio de atuação da força ilocucionária. E o exemplo (3), além de outras unidades, contém um Parentético, uma unidade que insere uma informação extra sobre o enunciado hospedeiro.³

Exemplo 1. *Enunciado simples, bfamcv01 [12]:*

*GIL [31]: esse que é o ponto //COM=

Exemplo 2. *Enunciado complexo com padrão Tópico-Comentário, bfamd03 [12]:*

*LAU [148]: departamento /=TOP= Artes Plásticas //COM=

Exemplo 3. *Enunciado complexo com Parentético, bfamcv04 [12]:*

*BRU [183]: se /=DCT= for uma mesa /=TOP= por exemplo /=PAR= cê não pode fazer assim //COM=

Cantalini [16] explorou a relação entre estrutura informacional e gestos, apontando uma sobreposição de quase todas as unidades tonais e unidades gestuais nos dados por ela analisados. A autora concluiu também que existe uma correspondência funcional em diferentes níveis, isto é, fala e gestos codificam a mesma informação: no nível da palavra, unidade informacional, função informacional e valor ilocucionário. Além disso, exemplos na literatura [3, 6] indicam que gestos mudam de padrão em uma mesma unidade para acompanhar a mensagem veiculada verbalmente. O exemplo normalmente mencionado é o de que certos Parentéticos, unidades que inserem um comentário no enunciado que os hospeda [21], são acompanhados⁴ por gestos que quebram o padrão estabelecido para o restante do enunciado em que estão inseridos. Isso indica que existe um nível de análise que não está ligado às proeminências que ancoram gestos e fala, mas sim às unidades intermediárias que os conectam. O estudo dessas instâncias intermediárias permite entender como nuances da fala podem ser veiculadas e entendidas em um nível microanalítico.

³ Esses exemplos foram retirados do C-ORAL-BRASIL I [12]. O acrônimo do arquivo indicado na legenda do exemplo indica qual o texto de origem do enunciado: “b” indica que é em português brasileiro, “fam” indica que o texto é de um contexto familiar e “cv” ou “dl” indica se o texto é uma conversação ou dialógico, respectivamente. Para mais detalhes, cf. [19]. As normas de transcrição indicadas nos textos combinam critérios ortográficos e não ortográficos que são detalhados em [20]. As convenções mais pertinentes para a legibilidade dos exemplos é que quebras não terminais são marcadas por barras simples “/” e as terminais por barras duplas “//” e que o pronome “você” é reproduzido como é falado na transcrição, estando em sua forma reduzida “cê”.

⁴ *Acompanhado*, aqui e por todo o texto, significa “realizado concomitantemente com”.

III. MÉTODOS

A. O corpus BGEST

O corpus BGEST [11] foi elaborado para ser um projeto piloto dentro do projeto C-ORAL-BRASIL. Sendo assim, ele tem uma arquitetura similar à seção de monólogos do C-ORAL-BRASIL I, em que um único falante ocupa a maior parte dos turnos de fala. O BGEST teve como objetivo prover uma metodologia coerente para segmentar gestos e fala e, assim, propiciar o estudo concomitante de unidades informacionais e gestos. Essa proposta se inspirou no trabalho de Cantalini

[16] para o italiano, realizado analisando-se dados de fala espontânea e recitada. O corpus BGEST é composto de dez excertos de até três minutos de monólogos de fala espontânea em português brasileiro (sem nenhum estímulo prévio). No total, o corpus tem aproximadamente 24 minutos de duração, 4.000 palavras e 450 gestos.⁵

TABELA I. CARACTERÍSTICAS PRINCIPAIS DO CORPUS BGEST

Arquivo	Duração (mm:ss)	# Palavras	# Enunciados
bgest_001	02:24	330	18
bgest_002	02:24	386	33
bgest_003	02:40	417	22
bgest_004	02:11	382	19
bgest_005	02:04	341	19
bgest_006	02:40	375	36
bgest_007	02:39	514	39
bgest_008	02:04	276	18
bgest_009	02:39	545	35
bgest_010	02:43	418	36
Total	24:28	3984	275

Adaptado de Barros [11].

A transcrição dos textos foi feita utilizando-se os critérios adotados no C-ORAL-BRASIL I, descritas em [19, 20]. Esses critérios, de base ortográfica, visam capturar fenômenos de gramaticalização e lexicalização em curso na língua, mantendo sempre que possível a legibilidade textual. A anotação foi conduzida em duas etapas, separadas entre fala e gestos. A parte de fala foi realizada utilizando-se o software Praat [23] apenas com acesso às oitivas e procedendo-se à segmentação em unidades terminais (enunciados), unidades não terminais (unidades tonais) e anotação de etiquetas informacionais [9]. A parte gestual foi segmentada e anotada usando-se o software ELAN [24] para os seguintes parâmetros: unidade gestual, frase gestual, fase gestual, tipo e direção de movimento, forma de mão, posição. Os gestos foram anotados com acesso à oitava porque entendemos que se estamos analisando gestos concomitantes com a fala, seu sentido só emerge na presença da fala [6, 25]. Assim, segmentar e anotar os gestos sem acesso às oitivas seria perder dados relevantes. Abaixo, a Figura 2 exemplifica a transcrição e anotação completa no ELAN, retirada de [11].

⁵ Para uma discussão sobre os critérios de seleção dos textos, adaptação dos critérios do C-ORAL para a corpus BGEST, cf. [11, 22].

Figura 2. Exemplo de transcrição e anotação usando ELAN

The screenshot displays the ELAN 6.2 software interface. At the top left, there are two video windows showing a person sitting on a couch. Below the videos are playback controls. The main area is divided into several tracks:

- Transcription:** A table with columns for 'Grade', 'Texto', 'Legenda', 'Lexicon', 'Comments', 'Reconhecedores', 'Metadados', and 'Controles'. It contains several lines of text with corresponding time stamps.
- Timeline:** A horizontal axis at the bottom showing time intervals from 00:00:00.000 to 00:01:06.000.
- Annotation Tracks:** Multiple tracks below the transcription, including 'Transcription', 'InfoStructure', 'GE-Units', 'GE-Phrase', 'right-GE-Position', 'right-GE-Hand-shape', 'right-GE-Orientation', 'right-GE-Movement', 'left-GE-Position', 'left-GE-Hand-shape', 'left-GE-Orientation', and 'left-GE-Movement'. These tracks contain detailed gesture annotations such as 'retraction', 'stroke', 'periphery', 'flat-hand', 'away center', and 'arcad'.

Figura retirada de BGEST, texto 6 [11]

No canto superior esquerdo estão os vídeos que podem ser consultados durante e após a anotação, no canto superior esquerdo estão os controles do software que permitem acessar a transcrição ou a anotação, além de controlar o som e vídeo. Na parte inferior está a transcrição (primeiras duas linhas) e a anotação gestual, dividida em mão esquerda e direita.

Igualmente relevante foi revisar a anotação informacional com acesso aos gestos, o que permitiu uma maior riqueza de detalhes e, com isso, uma anotação mais precisa. A escolha para o uso dos softwares Praat [23] e ELAN [24] deu-se por serem eles livres, gratuitos, amplamente utilizados pela comunidade científica e possuírem interoperabilidade. Um script para a análise de unidades prosódicas e gestuais também foi elaborado e se encontra atualmente em versão beta [25]. Esse script é uma ferramenta computacional que permite analisar a sobreposição de unidades prosódicas e gestuais, tomando-se como parâmetro inicial o ataque, fornecendo a porcentagem e o tempo (em milissegundos) da sobreposição de unidades prosódicas não terminais, frases gestuais e ataque, além de ser possível fazer-se a análise filtrando-se por unidade informacional.

B. O acordo entre unidades prosódicas e gestuais

Existem muitos caminhos para se avaliar o acordo entre unidades prosódicas e gestuais, uma vez que essa denominação é muito geral. A primeira grande divisão que se encontra entre as opções analíticas disponíveis é entre a coordenação de unidades e a coordenação de proeminências (ou picos). Vamos nos deter na primeira subdivisão (coordenação de unidades), que é abordada nesse trabalho. Ela se divide em duas outras possibilidades: sobreposição e alinhamento. *Sobreposição* denota a sincronia de unidades prosódicas e gestuais, independentemente de tipo e da relação funcional entre elas. Já *alinhamento* designa uma correspondência entre unidades prosódicas e gestuais que são funcionalmente compatíveis e que apresentam uma distância de proeminências suficientemente próximas para que possam ser

compreendidas em conjunto. Nesse trabalho explorou-se como o alinhamento entre unidades tonais não terminais e frases gestuais relacionava-se, a fim de se observar o reflexo deste alinhamento nas unidades informacionais.

Assim, a análise se baseou na busca por unidades tonais não terminais que englobariam um ataque gestual, para depois se encontrarem as diferenças entre as fronteiras iniciais e finais de unidades não terminais e frases gestuais. A partir desses valores pode-se chegar à proporção de alinhamento, bem como pode-se calcular qual a distância entre essas fronteiras. O protocolo de análise tentou conciliar o rigor metodológico de Loehr [6],⁶ principalmente no que se refere à replicabilidade, transparência e cuidado com as fronteiras iniciais e finais, e a abordagem de Cantalini [16] na escolha de quais unidades seriam pertinentes para o estudo. Cantalini analisou principalmente a sobreposição de unidades prosódicas terminais com unidades e frases gestuais, buscando principalmente entender se havia alguma unidade informacional que era concomitante com gestos. O critério adotado foi que quaisquer unidades seriam consideradas sobrepostas se estivessem até 200 ms umas das outras. Isso é uma metodologia exploratória, da qual se adotou a relação com as unidades informacionais, mas não o limite temporal. Assim, os valores encontrados puderam ser analisados à luz dos achados dos autores anteriores, bem como ofereceram indicações de possibilidades para estudos quantitativos futuros, além de serem pertinentes para estudos posteriores. A análise qualitativa contemplou todos os Parentéticos do corpus (unidades que geraram dúvida foram excluídas da amostra) para entender quais as manifestações que poderiam ser encontradas nos gestos.

IV. RESULTADOS E DISCUSSÃO

A análise quantitativa revelou que as unidades prosódicas estavam contidas nas frases gestuais e incluíam os ataques gestuais. Com isso as frases gestuais são as unidades intermediárias de maior extensão temporal (~747 ms) e que se iniciam antes das demais. Elas são seguidas por unidades prosódicas não terminais, que se iniciam por volta de 200 ms depois. Dentro de uma dada frase gestual e de uma unidade não terminal, o ataque é realizado. A maior contribuição desse estudo é mostrar que o intervalo de tempo entre o início da fronteira prosódica e o início do ataque varia de acordo com a unidade informacional à qual está associada, sendo este intervalo de, em média, 118 ms quando consideradas quaisquer unidades informacionais, cerca de 76 ms para Comentários e 87 ms para Parentéticos. Os resultados estão sumarizados na tabela (2), abaixo:

TABELA II. PRINCIPAIS MEDIDAS CAPTURADAS NOS DADOS

Unidade informacional	Duração do Ataque	Início de frase gestual e unidade tonal	Final de frase gestual e unidade tonal	Duração da frase gestual	Início de ataque e unidade tonal	Final de ataque e unidade tonal	Duração de unidade tonal	Taxa de alinhamento de frase gestual e unidade tonal	Taxa de alinhamento de ataque e unidade tonal
Todas as unidades informacionais	747	-260	73	1801	118	-356	1334	0.35	0.99
Comentário	733	-286	160	1815	76	-390	1270	0.35	0.94
Parentético	704	-290	-4	1768	87	427	1255	0.30	0.91

Dados retirados e traduzidos de [26].

Qualitativamente, observou-se que Parentéticos são gestualmente marcados por rupturas no padrão que se estabelece dentro do restante do enunciado. São possíveis duas estratégias de ruptura gestual: a suspensão momentânea ou a inserção

⁶ Cabe apontar que, apesar de comparáveis, as metodologias aplicadas aqui e por Loehr [6] se diferem principalmente porque o autor pré-estabeleceu uma medida que seria considerada como acordo com base na dispersão, enquanto aqui foi feita uma análise do alinhamento com base nas unidades que incluíam um ataque gestual. Essa diferença está presente em uma discussão mais extensa da análise quantitativa, cf. [25]

do gesto entre repousos. Essa estratégia se adequa à unidade gestual em que se insere mas, principalmente, coordena um padrão com o enunciado, adequando o gesto simultâneo ao Parentético para necessariamente marcar uma ruptura.

V. CONCLUSÃO

O trabalho apresentado debruçou-se sobre como estudar prosódia e gestos a partir de unidades intermediárias, demonstrando que uma abordagem a partir da L-AcT, permite adotar uma metodologia coerente com a análise almejada. Esse ponto é importante porque assegura que os dados estão foram tratados, classificados e analisados tomando-se os mesmos critérios, corroborando propostas explicativas aventadas pela teoria. Os resultados quantitativos apontaram para um alinhamento escalonado, que se inicia pela frase gestual, é seguida pela unidade não terminal e pelo ataque. Os resultados qualitativos indicaram que os gestos seguem padrões usados na fala, tais como adequação contextual às unidades adjacentes, para reforçar o conteúdo que é veiculado verbalmente. Com isso, temos resultados promissores sobre a importância do estudo de unidades intermediárias, principalmente para se entender como a segmentação se organiza vinculando prosódia e gestos.

ACKNOWLEDGMENT

Avoid the stilted expression, “One of us (R. B. G.) thanks . . .” Instead, try “R. B. G. thanks”. Put sponsor acknowledgments in the unnum-bered footnote on the first page.

REFERENCES

- [1] Pouw W, de Nooijer JA, van Gog T, et al. Toward a more embedded/extended perspective on the cognitive function of gestures. *Front Psychol*; 5. Epub ahead of print 24 April 2014. DOI: 10.3389/fpsyg.2014.00359.
- [2] Müller C. Gesture and Sign: Cataclysmic Break or Dynamic Relations? *Front Psychol* 2018; 9: 1651.
- [3] McNeill D. *Hand and mind: What gestures reveal about thought*. Chicago, IL, US: University of Chicago Press, 1992.
- [4] Kendon A. Relationships between body motion and speech. 1972, pp. 177-210.
- [5] McClave E. Gestural beats: The rhythm hypothesis. *J Psycholinguist Res* 1994; 23: 45–66.
- [6] Loehr D. *Intonation and Gesture*. Doctoral Dissertation, University of Georgetown, 2004.
- [7] Pouw W, Harrison SJ, Dixon JA. Gesture–speech physics: The biomechanical basis for the emergence of gesture–speech synchrony. *Journal of Experimental Psychology: General* 2020; 149: 391–404.
- [8] Kita S. How representational gestures help speaking. In: McNeill D (ed) *Language and Gesture*. Cambridge University Press, pp. 162–185.
- [9] Moneglia M, Raso T. Appendix: Notes on the Language into Act Theory. In: Raso T, Mello H (eds) *Studies in Corpus Linguistics*. Amsterdam: John Benjamins Publishing Company, pp. 468–495.
- [10] Kita S, Özyürek A. What does cross-linguistic variation in semantic variation. *Journal of Memory and Language* 2003; 48: 16–32.
- [11] Barros C. *A relação entre unidades gestuais e quebras prosódicas: o caso da unidade informacional Parentético*. Mestrado, Universidade Federal de Minas Gerais, 2021.
- [12] Raso T, Mello H (eds). *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora UFMG, 2012.
- [13] Cresti E, Moneglia M (eds). *C-ORAL-ROM: integrated reference corpora for spoken Romance languages*. Amsterdam; Philadelphia, PA: J. Benjamins, 2005.

- [14] Kendon A. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press. Epub ahead of print 2004. DOI: 10.1017/CBO9780511807572.
- [15] Ladewig S, Bressem J. 69. A linguistic perspective on the notation of gesture phases. In: Müller C, Cienki A, Fricke E, et al. (eds) *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/1*. Berlin, Boston: DE GRUYTER. Epub ahead of print 16 January 2013. DOI: 10.1515/9783110261318.1060.
- [16] Cantalini G. *La gestualità co-verbale nel parlato spontaneo e nel recitato*. Università degli studi Roma Tre, 2018.
- [17] Cresti E. *Corpus del italiano parlato*. Firenze: Accademia della Crusca, 2000.
- [18] Austin J. *How to do things with words*. Oxford: Oxford University Press, 1988.
- [19] Mello H. Methodological issues for spontaneous speech corpora compilation: The case of C-ORAL-BRASIL. In: Raso T, Mello H (eds) *Studies in Corpus Linguistics*. Amsterdam: John Benjamins Publishing Company, pp. 27–68.
- [20] Mello H, Raso T, Mittmann MM, et al. Transcrição e segmentação prosódica do corpus C-ORAL-BRASIL: critérios de implementação e validação. In: *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora UFMG, 2012, pp. 125–176.
- [21] Tucci I. Obiter dictum: La funzione informativa delle unità parentetiche. In: Pettorino M, Giannini A, Dovetto F (eds). Napoli: Università l’Orientale Press, 2010, pp. 365–654.
- [22] Barros C, Mello H. The C-ORAL-BRASIL proposal for the treatment of multimodal corpora data: the BGEST corpus pilot project. In: *Language Technologies and Digital Humanities*. Chile: UFRO, in press.
- [23] Boersma P, Weenink D. *Praat: doing phonetics by computer*, <http://www.praat.org/> (2020, accessed 7 March 2020).
- [24] Wittenburg P, Brugman H, Russel A, et al. ELAN: a Professional Framework for Multimodality Research. In: *Proceedings of LREC 2006*. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive, pp. 1556–1559.
- [25] Cantalini G, Moneglia M. The annotation of gesture and gesture/prosody synchronization in multimodal speech corpora. *Journal of Speech Sciences* 2020; 7–30.
- [26] Barros C, Santos SM. A methodology and a resource for comparing gesture and prosodic boundaries in multimodal corpora. In: *Proceedings of the PROPOR 2022*. in press.