



Artigos

O trabalho técnico-metodológico e analítico com dados interacionais audiovisuais: a disponibilidade de recursos multimodais nas interações

The technical, methodological and analytical work with audiovisual interactional data: the availability of multimodal resources in interactions

Fernanda Miranda da Cruz¹

Ana Cristina Ostermann²

Daniela Negraes Pinheiro Andrade³

Minéia Frezza⁴

RESUMO

A partir de uma coleção de dados audiovisuais de situações de complexa ecologia interacional e material, abordamos questões de caráter técnico-metodológicas, teóricas e epistemológicas oriundas do trabalho

1. Universidade Federal de São Paulo (Unifesp). Guarulhos, São Paulo - Brasil. <http://orcid.org/0000-0002-5051-7759>. E-mail: fernanda.miranda.cruz@gmail.com

2. Universidade do Vale do Rio do Sinos (Unisinos). São Leopoldo, Rio Grande do Sul - Brasil. <https://orcid.org/0000-0001-9017-6314>. E-mail: aco@unisinos.br

3. Interagis Serviços de Linguística Aplicada. São Leopoldo, Rio Grande do Sul - Brasil. <https://orcid.org/0000-0003-4580-5307>. E-mail: danielanegraes23@gmail.com

4. Instituto Federal do Rio Grande do Sul, Campus Farroupilha, Rio Grande do Sul - Brasil. <https://orcid.org/0000-0002-4480-898X>. E-mail: mineia.frezza@farroupilha.ifrs.edu.br



This content is licensed under a Creative Commons Attribution License, which permits unrestricted use and distribution, provided the original author and source are credited.

com vídeo destinado a análises linguístico-interacionais de orientação multimodal. Dentre elas destacamos: formas de registro; uso de softwares; práticas de transcrição e representação de dados audiovisuais; distinção entre descrição e transcrição de movimentos corporais. Inspiradas nas reflexões produzidas no campo da videoanálise, apresentamos como resultado alguns procedimentos de pesquisa mobilizados para dar conta da transcrição, representação e publicização dos recursos multimodais relevantes para os/as participantes nas interações.

Palavras-chave: *Videoanálise; Multimodalidade; Princípio de disponibilidade; Transcrição multimodal.*

ABSTRACT

Based on a collection of audiovisual data from diverse situations involving complex interactional and material ecology, this paper discusses the technical, methodological, theoretical, and epistemological questions that emerge when working with video for the purposes of linguistic-interactional and multimodal analyses, among which: forms of documentation; software use; audiovisual data transcription and representation practices; distinction between description and transcription of body movements. Fundamented in knowledge produced in the field of video analysis, the paper presents the results of the decision-making on research procedures that has enabled us to attain matters of transcription, representation, and publicization of multimodal resources demonstrated to be relevant for the participants of the interactions analyzed.

Keywords: *Video analysis; Multimodality; Availability Principle; Multimodal transcription.*

Introdução

Neste artigo, abordamos desafios de ordem técnica e metodológica presentes em práticas de pesquisas que subsidiam e viabilizam uma análise multimodal de interações cotidianas registradas em vídeo em contextos naturais de ocorrência. Estudos sobre multimodalidade da interação têm sido aprofundados a partir da chamada perspectiva corporificada da interação (Streeck, Goodwin e LeBaron 2011a). A partir dessa perspectiva, considera-se que construímos os espaços interacio-

nais multimodalmente; ou seja, que nossas ações sociais são construídas graças a uma ecologia de sistemas semióticos, estruturalmente distintos entre si, mas intrinsecamente relacionados (Goodwin 2010). Esses recursos são de diferentes modalidades, verbais e não-verbais e, portanto, compreendem recursos: linguísticos (aspectos gramaticais, prosódicos, sintáticos, entoacionais e lexicais); corporais (posturas, gestos, direcionamentos do olhar) e materiais (referentes às relações múltiplas que temos de manuseio, referência e percepção sensorial de objetos e de elementos do espaço físico).

Os procedimentos para o desenvolvimento de uma investigação multimodal da interação social estão intimamente relacionados às possibilidades de se capturar por vídeo, e de se reconstituir, metodológica e analiticamente, a complexa coordenação temporal e espacial entre recursos de diferentes modalidades mobilizados pelos/as participantes durante a construção de suas ações. Uma investigação multimodal requer, assim, a adoção de práticas de transcrever e de representar tais interações de forma a dar conta da integração articulada dessas várias modalidades.

Mondada (2003a, 2008a, 2008b) é uma das analistas da interação que tem mostrado como as práticas de registro, transcrição e representação de dados interacionais são regidos por um princípio de disponibilidade, qual seja, se os recursos multimodais são relevantes antes de tudo para os participantes (perspectiva êmica). Então, para a análise desses dados é importante que tais recursos sejam também disponibilizados por meio dos próprios registros (escolhas de equipamentos, ângulos, edições ou aproximações das lentes da câmera, por exemplo) e por meio de como esses dados são transcritos, representados e apresentados impressa ou oro-audiovisualmente.

Preocupações dessa natureza encontram ressonância nas reflexões sistematizadas em torno de questões relativas às metodologias para análise de registros feitos com vídeo, abordadas pelo campo da videoanálise (Knoblauch *et al.* 2006). Knoblauch *et al.* (2006) propõem que o trabalho com vídeo seja *em si* um tópico de reflexão teórico-analítica. Refletir sobre os variados procedimentos realizados sobre os dados primários (os registros realizados em vídeo de determinada situação de interação naturalística) tem implicações diretas para a

forma como concebemos, produzimos e exploramos analiticamente as ações construídas *na e pela* fala-em-interação. Nesse aspecto, importa pensar em uma abordagem praxiológica das práticas de vídeo no campo dos estudos linguístico-interacionais (Mondada 2006, 2008b) de modo a se discutir *se e como* o uso do vídeo tem consequências não apenas metodológicas, mas também epistemológicas para os estudos interacionais. Isso porque o vídeo pode ser, derradeiramente, um lugar de preservação de recursos multimodais importantes da organização da interação, ao mesmo tempo em que configura reflexivamente os fenômenos de análise.

Afinado com essa agenda, nosso interesse consiste em tomar nossas próprias práticas de pesquisa baseadas em registros audiovisuais como ponto de partida para discutirmos a disponibilidade, do ponto de vista dos/as participantes, e o tratamento e a disponibilização, do ponto de vista do/a pesquisador/a analista, dos recursos multimodais relevantes para a construção das interações. Frente a isso, propomos aqui explicitar, a partir de uma coleção de dados de três situações interacionais, um conjunto de questões de caráter metodológico e técnico oriundo do trabalho com vídeo. Levamos em consideração o duplo e reflexivo papel do vídeo nos estudos linguístico-interacionais: o de preservar os detalhes mas também configurar os elementos de análise. Desse modo, abordaremos neste artigo: (1) formas de registro; (2) decisão sobre o uso (ou não) de softwares; (3) práticas de transcrição e escolhas de notação; (4) a distinção entre descrição e transcrição de movimentos corporais; (4) incorporação ou não dos elementos do ambiente físico na interação e das inter-relações entre mundo material e interação verbal; (5) práticas de representação dos dados e (6) as formas de publicização do material empírico para uma análise da interação social multimodalmente orientada.

Como dar conta dos recursos multimodais para análise das interações?

Organizamos a discussão introduzida a partir da análise de quatro excertos e nos perguntamos, centralmente, como dar conta do tratamento, da transcrição, da representação e da publicização dos

O trabalho técnico-metodológico e analítico com dados interacionais audiovisuais

detalhes e recursos multimodais relevantes para os/as participantes das interações?

O Excerto 1 refere-se a uma interação entre uma médica e uma paciente no momento de realização de um exame de ultrassom obstétrico. Em interações nesse contexto, a imagem em movimento do feto ou de certas posições e partes do feto projetadas no monitor são destacadas pelos/as participantes como relevantes para a atividade de construção conjunta de referentes durante a interação. Analisamos os recursos multimodais relacionados à manipulação de materiais mobilizados pelas participantes para ver e fazer a outra ver o que está na tela.

Os Excertos 2 e 3 referem-se a uma mesma interação representada por dois sistemas distintos. Temos uma interação envolvendo uma paciente e um médico no momento em que ele anuncia, como prescrição, a necessidade da paciente parar de fumar. A resposta à primeira parte do par adjacente *entrega da notícia por parte do médico/recepção da notícia por parte da paciente* é realizada multimodalmente através de movimentos corporais e aspectos verbais sequencial e temporalmente estruturantes da ação em questão. Discutiremos as implicações, para a análise, de duas práticas distintas: a de descrição de gestos corporais inscritas na transcrição verbal por meio de parênteses duplos e a de transcrição das ações corporais coordenadas temporal e sequencialmente ao curso da fala.

O Excerto 4 apresenta uma interação em um ambiente domiciliar entre uma criança diagnosticada dentro do espectro do transtorno autista (TEA) e sua irmã. Nessa interação, a análise e a inteligibilidade da produção verbal da criança autista dependem necessariamente de sua inscrição corporificada na interação, o que significa levar em consideração uma coordenação fina de temporalidades complexas envolvendo recursos materiais, corporais e linguísticos mobilizados pelas participantes. O tratamento multimodal desse excerto nos conduz a uma reflexão sobre a implicação para a análise do uso ou não de softwares que permitem um alinhamento temporal e sequencial mais preciso de modalidades verbais e não verbais.

Perceber, ver e fazer ver: um problema para participantes e para analistas de interações

De forma geral, as escolhas de registro audiovisual de uma interação incluem decisões sobre: (a) local de instalação dos equipamentos; (b) formas de operação dos equipamentos durante o registro; (c) quantidade e tipos de equipamentos; (d) tipos de *zoom* ou de movimentos de captura específicos realizados durante os registros, entre outras. Tais escolhas não são feitas sem implicações importantes para o que temos como material empírico primário – e que será submetido à transcrição, a representações e à análise. Como alerta Mondada (2008c), o registro em vídeo pode gerar dados primários extremamente complexos, sujeitos a várias transformações que os tornem aptos para análises. A autora sugere, inclusive, que cada vez mais os/as analistas se atentem para a distinção entre as ações de ver *através* da câmera (Büscher 2005) e ver *com* a câmera (Mondada 2012a). A ação de ver *com* a câmera reconhece e problematiza o vídeo “como configuração e organização de detalhes relevantes” (Mondada 2006:56⁵), reconstituído (ou mesmo coconstruído) por ações e decisões de quem está fazendo o registro, como ajustes na câmera de vídeo, seleção de cenas e/ou enfoque momentâneo sobre um/a participante específico/a, contingências do próprio contexto e situação de pesquisa.

O Excerto 1 nos oferece a oportunidade de tratarmos reflexivamente essa discussão. Ele foi extraído de um *corpus* de registro de interações entre médicas/os e gestantes ao longo de exames de ultrassom obstétrico. O *corpus* foi gerado entre 2013 e 2015⁶, totalizando 138 ultrassonografias obstétricas de médio e alto risco que passaram por uma escolha de configuração mista de registro que leva em consideração uma equação entre aspectos éticos referentes ao registro de consultas de potencial vulnerabilidade e aspectos metodológicos relacionados ao registro (captura visual). Como resultado dessa equação, em termos de dados primários, para esse excerto, temos a gravação em áudio da consulta, mais especificamente da conversa (oral) que se deu entre uma

5. “as configuring and assembling relevant details”.

6. O contexto do estudo está inserido em um projeto de pesquisa maior coordenado por Ana Cristina Ostermann e intitulado “Uma mulher, um feto, e uma má notícia: a entrega de diagnósticos de síndromes e de malformações fetais – em busca de uma melhor compreensão do que está por vir e do que pode ser feito” (2013).

médica e uma gestante, e o registro em vídeo da tela do computador em que é realizado o exame de ultrassom. Contudo, faz-se relevante registrar aqui que o ângulo da câmera captura não apenas um elemento físico aleatório do consultório médico, mas “faz ver” (no sentido de tornar visível) um tipo de problema que é manifestado, antes de tudo, como relevante para as participantes dessa interação.

Para que profissionais e participantes leigos/as construam a intersubjetividade em exames de ultrassonografia fetal, é preciso que ambos/as tenham acesso simultâneo aos referentes que giram em torno do feto (objeto de escrutínio visual) na interação e que passam a ter visibilidade por meio da tela do computador que registra o ultrassom. Contudo, a ecologia do contexto de exames de ultrassom fetal apresenta peculiaridades que geram determinadas contingências com as quais os/as participantes precisam lidar para se entenderem, sendo uma das mais desafiadoras a efemeridade dos referentes relacionados às imagens produzidas ao longo do exame e projetadas na tela de um computador – e, assim, disponibilizadas para visualização das/dos participantes. Diferentemente de outros exames imagéticos (*e.g.*, radiografia e tomografia), a ultrassonografia permite que os/as participantes interajam sobre e com as imagens ao longo da própria realização do exame. Porém, como o transdutor que capta as imagens no abdômen da gestante está constantemente em movimento (bem como frequentemente está o feto examinado), um dos desafios para esses/as participantes torna-se justamente oportunizar/fazer coincidir a simultaneidade entre aquilo que é falado e aquilo *sobre* o que é falado – *i.e.* imagens projetadas na tela. Para lidar com esse desafio, os/as interagentes, frequentemente, utilizam recursos multimodais de naturezas distintas para a construção de atenção conjunta, como o congelamento e/ou a ampliação da imagem, o gesto de apontar com a mão ou com o cursor do aparelho e símiles (Ostermann e Frezza 2018a; 2018b; 2018c).

Além das restrições físico-espacial-temporais de tornar as imagens disponíveis durante o exame – já que o/a profissional, que é quem opera o aparelho, precisa parar o exame para acionar algum desses mecanismos de construção de atenção conjunta, há de se considerar que os diferentes *status* epistêmicos dos/as participantes envolvidos/as nessas interações podem ser outro fator que gera problemas de intersubjetividade. O/A profissional tem o conhecimento técnico para

operar o equipamento e interpretar as imagens fetais, enquanto a gestante (e familiares) não.

No Excerto 1, as participantes (indicadas por MED, médica, e GES, gestante) se orientam para a imagem projetada na tela do computador do exame como foco central de atenção. A imagem, resultante das operacionalizações que a médica faz com o aparelho, acaba se tornando fonte de problema de intersubjetividade, atingindo de forma fulcral a progressividade interacional.

Além de analisar o problema de intersubjetividade (problema que deflagrou a discussão desse excerto em particular), nos cabe, como pesquisadoras, disponibilizar para nossos pares, em artigos impressos e/ou apresentações audiovisuais, aquilo que se torna relevante para os/as próprios/as participantes das interações investigadas. Em apresentações audiovisuais, essa tarefa é consideravelmente facilitada se tivermos à disposição (e aval ético) a opção de projetar vídeos. Contudo, no que tange a publicações impressas, ainda são raras as oportunidades de divulgação de *links* com dados em vídeo.

Vejamos, então, o resultado de nossa reflexão e trabalho para disponibilizar para nossos/as pares o problema de intersubjetividade gerado e resolvido pela ecologia do ambiente interacional do Excerto 1.

Vale mencionar que a imagem projetada no monitor do ultrassom é consequência direta dos movimentos que a médica faz com o transdutor sobre abdômen da gestante. Assim, a imagem projetada em tela está em constante movimento durante toda interação. Interessantemente, a possibilidade de congelamento da imagem ultrassonográfica (para fins de análise para nós, pesquisadoras) não está limitada à pesquisa. Pelo contrário, os/as profissionais recorrentemente utilizam desse recurso tanto para a realização do exame em si (*e.g.*, registro de medidas do feto), quanto para a explicação das imagens para as gestantes.

Desse modo, as imagens 1 M a 6 M⁷ foram realizadas através de capturas da tela do vídeo do exame de ultrassom pausado. Além de serem “imagens da imagem” (“picture-in-picture”) (Mondada 2003b, 2006, 2009, 2014c), tratam-se de imagens de uma imagem *em movimento*, o que explica a inserção da letra “M” na descrição de cada captura da tela (*i.e.*, “imagem em movimento”). Para registrar essa complexidade da ecologia local e de forma a oportunizar que o/a leitor/a também “veja” e compreenda essa complexidade, as imagens foram inseridas conforme as convenções propostas por Mondada (2014d) ao final do excerto.

No caso dessa interação, o movimento da imagem projetada na tela é justamente o que dificulta, e, em certa instância, impossibilita, a simultaneidade entre referente (na tela) e referido. Quando a gestante inicia seu turno de fala “aquela parte que tá aí >parece um tatu bola< né xx” (l. 427-428), a Imagem 1 M estava visualmente disponível na tela. Contudo, como a médica continua o exame sem pausar a imagem, ao final da produção do turno da gestante, a Imagem 1 M já não está mais disponível. Para dar conta de demonstrar a efemeridade do referente no monitor do ultrassom, disponibilizamos capturas da tela de três momentos relacionados à produção do turno das linhas 427-428:

Imagem 1 M: capturada um segundo antes do turno de modo a representar a imagem percebida pela gestante como um “tatu bola”;

7. Reservamos o termo imagem aos conteúdos do registro audiovisual extraídos diretamente do vídeo e utilizamos o termo figura como elemento gráfico-textual referente a tabelas e figuras construídas, como capturas de tela do ELAN ou *Praat*. Para uma discussão mais ampliada sobre essas escolhas, vide Cruz (2018).

O trabalho técnico-metodológico e analítico com dados interacionais audiovisuais

Imagem 2 M: capturada no meio do turno da gestante para representar uma pequena alteração em relação à Imagem 1 M, quando do início de seu turno;

Imagem 3 M: capturada no local relevante para transição de turno, *i.e.*, na micropausa da linha 429, representando a disparidade entre a imagem anterior ao turno e a imagem posterior a ele.

Em resposta à gestante, a médica inicia um reparo (l. 430), tratado primeiramente como um indicativo de problema de audição pela gestante, que repete: “parece um °tatu {{rindo} bola°)” (l. 431). Em seguida, contudo, a falta de simultaneidade entre fala e imagem é evidenciada como a fonte do problema de intersubjetividade quando a gestante anuncia que mostrará novamente a imagem à profissional, assim que ela (re)aparecer na tela (l. 434).

Na linha 441, a imagem a que a gestante se referia no início do Excerto 1 volta a aparecer e, dessa vez, ela aponta para o monitor circulando duas vezes com o dedo indicador na parte que seria o “tatu bola” (Imagens 4 M a 6 M), sustentando esse gesto dêitico até o momento em que a profissional oferece um marcador de “mudança de estado” “a : : :” (l. 442) (Heritage 1984) que, nesse caso, sinaliza o entendimento da fonte do problema (*i.e.*, a visualização da imagem do que seria o “tatu bola”). Observe-se que é apenas após essa manifestação de “mudança de estado” – *i.e.* de passar a entender algo de forma diferente –, que a gestante inicia o movimento de retração do gesto dêitico, indicando que a ação de construir a atenção conjunta para o objeto da atividade de referência (indicialidade, indexicalidade) está concluída. Logo, a trajetória do gesto dêitico de apontar para o referente se torna relevante nessa interação, uma vez que ela representa a coordenação temporal e sequencial entre fala e gesto em interação, indicando a natureza da ação aqui desempenhada: construir a atenção conjunta para identificar o referente visual de tatu bola, *i.e.*, fazer ver junto, ver o mesmo referente.

Por conta do refinamento da transcrição, é possível perceber que ausência de fala de sete décimos de segundo da linha 441 integra o turno de fala, já que durante tal ausência de fala temos o gesto de apontar, que de forma sincrônica e coordenada aos recursos verbais “essa parte aí:” (l. 441), constitui multimodalmente essa “unidade de

construção de turno” (UCT). Hindmarsh e Heath (2000) argumentam que expressões dêiticas isoladas não funcionam como um *pointer*, nem especificam o referente. A gestante se orienta para a incompletude de referência das expressões dêiticas acionadas em seu turno (“essa parte aí:”) complementando-a com o gesto dêitico de apontar. Essa orientação, por sua vez, resolve a fonte do problema das linhas 427-428.

Perceber, ver e fazer ver: da descrição subjetiva à complexidade das temporalidades

Nesta seção, propomos uma reflexão sobre como as ações de perceber, ver e fazer ver, no que tange à atividade de pesquisa de AC a partir de dados multimodais, estão condicionadas à metodologia de transcrição adotada. Dentre as metodologias frequentemente utilizadas para transcrever condutas corporificadas, Hepburn e Bolden (2017) discutem o uso de: (a) comentários em parênteses duplos; (b) sistemas de segmentação e notação e transcrição multimodal e (c) fotografias de imagens extraídas de vídeos ou desenhos representativos das condutas corporificadas de interesse. Ancoradas em estudos de referência na tradição analítica interacional multimodal (*e.g.*, Goodwin 1981; Kendon 2004; Mondada 2001; 2016; Rossano 2012; Streeck *et al.* 2011b), as autoras discorrem sobre as motivações analíticas que subjazem cada uma dessas metodologias, sendo o modo de representação visual da conduta corporificada, alinhada ou não à fala, a variante determinante para a opção entre uma ou outra metodologia.

Enquanto comentários entre parênteses duplos mostram-se adequados quando a temporalidade da conduta corporificada não é foco de atenção na análise, o trabalho de segmentação, notação e transcrição multimodal torna-se fundamental para entendermos como as condutas corporificadas alinham-se temporalmente com a fala (Hepburn e Bolden 2017). Nas palavras de Streeck *et al.* (2011b:2), a temporalidade das “estruturas de participação corporificadas” é parte constitutiva da ação e, portanto, deve ser considerada como elemento a ser analisado se quisermos tentar responder “por que isso agora?” Representações visuais a partir de fotografias ou desenhos, por sua vez, são alternativas frequentemente associadas à transcrição, pois “têm a vantagem de

serem facilmente interpretáveis e mais holísticas em termos de representação” (Hepburn e Bolden 2017:122). De uma forma ou de outra, a maior de todas as pretensões de qualquer metodologia de transcrição é capturar a “essência” do momento, *i.e.*, os recursos multimodais mobilizados pelos/as participantes e suas temporalidades específicas, a fim de disponibilizá-los para que o/a leitor/a possa analisar os dados de maneira independente (Hepburn e Bolden 2017; Mondada 2001). De maneira, então, a refletir sobre diferentes práticas de percepção e representação envolvidas na manipulação de dados audiovisuais, nesta seção, propomos discutir alguns dos impactos a que estão sujeitas as análises propostas a partir de duas convenções de representações (comumente tratadas como “convenções de transcrição”) de interações envolvendo ações corporificadas, mais especificamente, representações por meio de comentários/descrições entre parênteses duplos e sistemas de segmentação e notação e transcrição multimodal.

Os Excertos 2 e 3 referem-se a uma mesma interação (representada por convenções distintas) e fazem parte de uma coleção, registrada em 2013, de 25 encontros face a face entre profissionais de saúde e pacientes cardiopatas. Os encontros, que foram registrados na ala de cardiologia de um hospital da rede de saúde privada localizado no sul do Brasil, tinham como objetivo oportunizar a pacientes hospitalizados acesso a orientações sobre autocuidados de saúde para fins de controle de fatores de risco que incidem sobre doenças cardiovasculares (DCV)⁸.

O evento interacional em análise ocorreu entre um médico cardiologista e uma paciente cardiopata no dia da consulta de alta da paciente e inicia no ponto em que o médico recomenda à paciente que ela abandone o cigarro. Para a reflexão a que este artigo se propõe, concentramos o foco de atenção aqui nas condutas corporificadas mobilizadas pela paciente em resposta à recomendação do médico. Para construir o argumento central dessa discussão, em um primeiro momento, fazemos uso das convenções jeffersonianas para transcrever o excerto e, desse modo, recorreremos ao recurso do comentário/descrição entre parênteses duplos para representar as condutas corporificadas dos interagentes. Em

8. Maior detalhamento sobre a geração dos referidos dados está disponível em Andrade (2016).

um segundo momento, fundamentadas no trabalho de segmentação do dado por meio do software ELAN (Wittenburg *et al.* 2006) transcrevemos o excerto multimodalmente a partir das convenções propostas por Mondada (2014d), em que a autora sugere que os comentários devem restringir-se a condutas verbais não convencionadas (*e.g.*, tosse) ou opcionalmente não transcritas (*e.g.*, aponta).

3.1. Parênteses duplos, segmentação e transcrição multimodal:

Excerto 2: HNS_Lauriano_Silvia_22_08_13_abandoná o cigarro

```
58 LAU:      então a senhora tem que abandoná o cigarro.
59 SIL:      ((respira fundo e expressa insatisfação/desconforto))
60 LAU:      vamo começá pra:: (.) a partir daí.
61           (0.7)
62 LAU:      tá:?
63 SIL:      >°pois é° mas não é fácil.<=
64 LAU:      =porque o cigarro é um problema que::
65           atrapalha bastante a questão dos vasos.
66           (1.8)
67 LAU:      tá:?
68           (0.8)
69 SIL:      abandoná o ciga- na- é fácil dizê. a|bandona o cigarro.
70 LAU:      =pra mim é muito fácil.
71 SIL:      claro. o senhor não fuma né?
```

!

O modelo de transcrição adotado no Excerto 2 privilegia a análise com base na organização de cursos de ação estabelecida por turnos de fala relacionados adjacientemente (Schegloff 2007). A relação de adjacência faz emergir sequências sucessivas formadas por primeiras e segundas partes de pares adjacentes e que, de acordo com seu formato e posição, implementam ações que mantêm relação de coerência e ordem umas com as outras. Em outras palavras, esse modelo de transcrição favorece a análise a partir de sequências mínimas formadas por pares adjacentes de dois turnos (Schegloff 2007). Nesse modelo de transcrição, condutas corporificadas, ou outros eventos extralinguísticos potencialmente relevantes para a interação, tendem a ser descritos de modo a privilegiar: (a) apenas uma das modalidades (*e.g.*, direcionamento de olhar) e (b) a temporalidade e a sequencialidade sem uma

coordenação fina com a temporalidade e a sequencialidade da fala. Além disso – e sobremaneira importante para a nossa discussão –, no modelo apresentado, a depender dos eventos interacionais de interesse para a análise, as descrições são realizadas a partir de interpretações de natureza subjetiva e dependentes da/o analista.

Levando-se em conta essas considerações, no Excerto 2, ao buscaremos dar conta de descrever a conduta corporificada de Sílvia, uma das partes dos comentários entre parênteses duplos remete à ação física socialmente disponibilizada pela participante – “respira fundo” –, enquanto a outra recai sobre seu estado emocional – “expressa insatisfação/desconforto”. Em termos de representação gráfica, a descrição da conduta corporificada, em conjunto com a interpretação subjetiva acerca do estado emocional da paciente, ocupa uma linha da transcrição (l. 59), sendo apresentada como uma UCT. A partir da perspectiva analítica sequencial, a UCT preenche a segunda posição do par adjacente da sequência recomendação – aquiescência/resistência (Stevanovic e Peräkylä 2012) iniciada por Lauriano.

Desse modo, a partir dessa decisão de transcrição, observa-se que o/a analista percebe e faz ver a conduta corporificada de Sílvia como tão somente respondendo à primeira parte do par adjacente (PPP) da sequência iniciada pela recomendação. A conduta descrita transparece como consequência direta para a ação subsequente de Lauriano, a saber, uma proposta para que Sílvia cesse o hábito de fumar (l. 60), ao que a paciente reage com ausência de fala (l. 61). “Dado [que, devido a]o poder de projeção da adjacência dos pares, as ações feitas em posições de segundas partes são altamente condicionadas (Levinson 2013:103), a ausência de fala após uma recomendação sinaliza desalinhamento ao curso de ação iniciado por Lauriano, o que pode ser interpretado como constituindo a ação despreferida de resistir.

A leitura que se pode fazer a partir da transcrição apresentada no Excerto 2 é a de que a resistência de Sílvia gera consequências para a próxima ação de Lauriano. Ao se deparar com (duas) respostas antagônicas independentes e sucessivas da paciente, *i.e.*, a conduta física visível e a ausência de fala, Lauriano produz uma ação que pode ser interpretada como busca de alinhamento/afiliação (Heritage e Lindström 2012). Sílvia, no entanto, persevera em demonstrar resistência

à proposta de cessação do tabagismo, dessa vez, por meio de conduta verbal (l. 63).

Acompanhando a transcrição, após o momento em que Sílvia verbaliza pela primeira vez seu posicionamento sobre o que está em jogo nesse instante interacional, *i.e.*, a negociação que envolve seu hábito de fumar, tem-se a impressão de que Lauriano não se orienta para a dificuldade que Sílvia enfrenta. Esse argumento leva em consideração o fato de que a justificativa que preenche o próximo turno de Lauriano (l. 64) parece operar como complemento de uma linha de raciocínio que inicia com a proposta lançada no turno da linha 60 e que precisa ser perseguida frente ao silêncio produzido por Sílvia, na linha 61. Após a justificativa, Lauriano depara-se com nova ausência de fala por parte de Sílvia (l. 66), o que gera nova busca dele por alinhamento/afiliação (l. 67), mas sem sucesso algum.

Com base nessa decisão de transcrição, e conseqüente lógica analítica, na maior parte do Excerto 2, a formação das ações de Lauriano mostra-se sob influência da ação imediatamente anterior de Sílvia (com exceção, talvez, da ação produzida pelo turno representado na l. 64). Não há, na transcrição, indicação de que a temporalidade da conduta corporificada de Sílvia, que perdura, como veremos a seguir, para além do turno imediatamente responsivo à recomendação. Tampouco há sinal de que Lauriano está orientado para a conduta corporificada de Sílvia (l. 60).

Ao submetermos o mesmo dado ao sistema de segmentação disponibilizado pelo ELAN, podemos perceber que a conduta corporificada de Sílvia transpõe o limite de uma UCT. Sílvia, de fato, inicia a conduta na linha 59, mas a sua realização perdura até o momento representado pela linha 68 na transcrição apresentada no Excerto 2. Além disso, afora a ação física de inspirar e expirar o ar de maneira visível, a paciente realiza outras condutas: movimentos de cabeça, fricção da mão esquerda sobre o colo e mudança no direcionamento de olhar. As “múltiplas dimensões multimodais sincrônicas e coordenadas” (Mondada 2007b:813), e socialmente disponibilizadas por Sílvia, que, no Excerto 2, respondem pela descrição subjetiva interpretativista sobre seu estado mental, tornam-se perceptíveis por meio de transcrição multimodal detalhada, como mostra o Excerto 3:

O trabalho técnico-metodológico e analítico com dados interacionais audiovisuais

Excerto 3: HNS_Lauriano_Sílvia_22_08_13_abandoná o cigarro_multimodal

```

01 LAU:  @*então a senhora tem que abandoná
02      @>> olha para sil----->>
03 sil   Δ>> olha para lau -----> (1.06)
04 sil   *..--move ME da testa até o colo----->
05      o ciga:rro.*Δ
06 sil   ----->.*Δ
07 SIL:  #†Δ.h:::h:::=-
08      #ima. 1
09      †..retrai cabeça--> (1. 11)
10      Δ..--- olha p E--> (1. 16)
11 LAU:  #*=vamo começá†# pra::#(.) a partir daí.#
12 sil   ----->†
13      #ima. 2
14 sil   *...mov. vai e vem ME no peito -----> (1. 21)
15      #ima. 3          #ima. 4          #ima.5
16 SIL:  .h:::hh
17 LAU:  #†tá::?=
18 sil   †..balança cabeça neg. --> (1. 24)
19      #ima.6
20 SIL:  =# '>pois é<' mas não* éΔ fáΔ [cil.
21      ----->Δ...Δ...olha p lau---> (1.26)
22      ----->*
23 LAU:  [porque
24      o cigarro† é um problema que:: atrapalha
25 sil   ----->†
26      Δ>bastante< a questão dos va:sos.
27      Δ...Δ----- olha p E.. ----->(1. 31)
28 SIL:  Tsk
29      (0.7)
30 LAU:  'tá:?'
31 sil   (0.8)Δ
32      ----->Δ
33 SIL:  Δ a†bandoná o ciga- na- é †fácil dizê.
34      Δ...-olha para lau ----->>
35      abanidona o cigarro.=
36 LAUR:  ¬ =pra mim é muito¬ fácil.=
37      ¬.....¬ aponta polegar p o peito
38 SIL:  †=claro. o senhor não fuma né?†
39      †....balança cabeça neg.----- > †

```

Legenda:

† movimento de cabeça
Δ direcionamento de olhar
* movimento da mão esquerda (ME)
¬ movimento do polegar

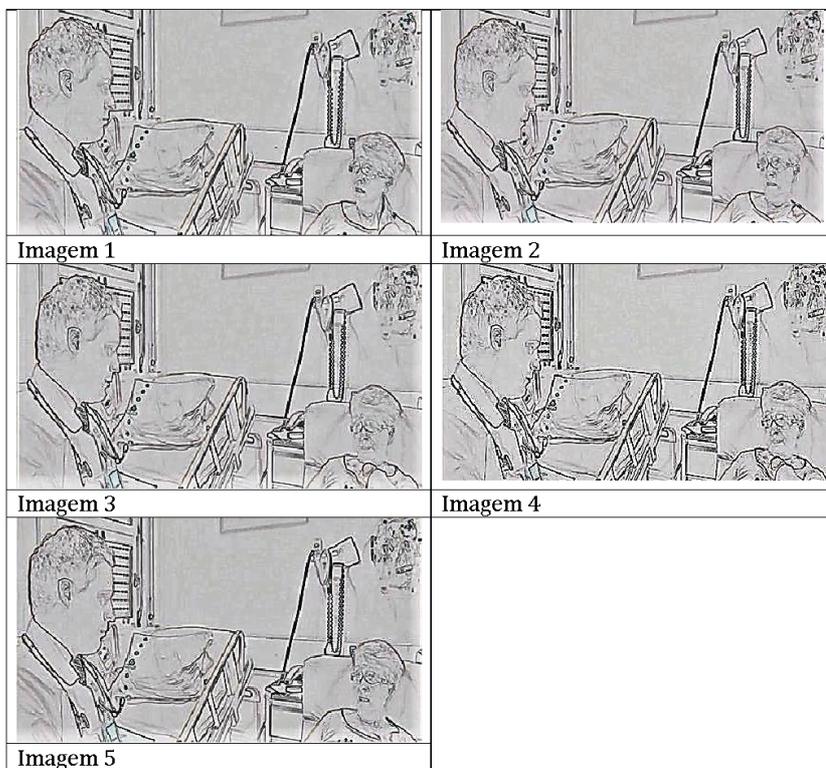


Figura 2 – Capturas de tela de vídeo do Excerto 3: imagens 1 a 5 sincronizadas aos turnos de fala transcritos (#).

Conforme evidenciado no Excerto 3, as condutas corporificadas de Sílvia, anteriormente descritas de forma interpretativa (“respira fundo” e “expressa insatisfação/desconforto”), são materializadas de modo a contemplar o alinhamento temporal de cada dimensão multimodal acionada pela paciente ao desdobramento da sequencialidade das falas (Mondada 2007b).

Nessa distinta representação, é possível perceber que durante o turno em que Lauriano faz a recomendação (“então a senhora tem que abandoná o ciga:rrro”), Sílvia está com o olhar direcionado para ele. Contudo, no lugar relevante para tomada de turno, a paciente desvia o olhar para a esquerda, movimento que inicia em concomitância com os movimentos de inspiração e expiração audíveis e com

a retração da cabeça, esse último movimento atingindo sua completude apenas após Lauriano proferir a palavra “começá” (l. 11). As condutas corporificadas das dimensões do olhar, da respiração e da cabeça coocorrem com o movimento de deslocar a mão esquerda da testa para o colo (l. 4). Uma vez com a mão esquerda repousada no colo, Sílvia a esfrega em movimento de vai e vem, atividade física que perdura até o instante em que ela produz a palavra “não” (l. 20). Vê-se também que Sílvia balança a cabeça negativamente em resposta à solicitação explícita de concordância feita por Lauriano (“tá: :?”, l.17), a de que Sílvia comece os autocuidados da saúde pelo abandono do tabagismo (“vamo começá pra:(.) a partir daí.”). O balanço negativo de cabeça de Sílvia, que inicia na linha 18, cessa no mesmo instante em que ela redireciona o olhar para Lauriano, i.e., após a produção da palavra não (l. 19). Assim, Sílvia indexicaliza posicionamento antagônico em relação ao abandono do tabagismo (“>pois é< mas não* é fá[cil.”), posicionamento esse que, em certa medida, se exacerba, no desenrolar da interação, conforme evidente na parte restante do fragmento.

No que se segue, as ações corporificadas de Sílvia – desvio do direcionamento o olhar, ausência de contato visual com o médico (l. 27-32) e balanço lateral de cabeça (l. 39) – juntamente com as ações verbais, constituem uma manifestação interacional que problematiza, de forma bastante profunda, as dificuldades de cessar o tabagismo. Nesse aspecto, Sílvia propõe que o nível de dificuldade altera de acordo com a ótica de quem fuma e de quem não fuma, ou seja, fumantes e não-fumantes.

O médico, que antes não havia se orientado para os recursos corporais de Sílvia, ao final da interação, alinha-se então à problematização apontada por ela sobre a diferença entre recomendação de parar de fumar direcionada a um fumante ideal e a situação real. O alinhamento do médico à Sílvia se confirma com o uso de primeira pessoa: “=pra mim é muito- fácil.” Esse alinhamento produz como efeito a ação de sair da posição médico que faz uma recomendação geral e situar-se relativamente à posição de sua interlocutora, como um não-fumante.

Com efeito, em termos metodológicos, a segmentação dos dados e a “representação multifacetada dos fluxos simultâneos e coordenados das ações” (Mondada 2007b: 813) geram consequências para a maneira

de perceber e dar visibilidade como se dá a formação das ações dos/as participantes da interação em análise. Nesse sentido, concordamos com Streeck *et al.* (2011b: 5) que, com base em Goodwin (1981) e M. H. Goodwin (1980), afirmam que:

[...] em situações inteiramente corporificadas, enunciados não são constituídos exclusivamente dentro do fluxo de fala pelas ações do/a falante. Ao contrário, as ações visuais dos/as ouvintes, incluindo tanto orientações para o/a falante quanto ações em resposta às especificidades da fala ao longo de sua produção, podem sistematicamente levar o/a falante a modificar a estrutura do enunciado em progresso.

Considerando, então, que os cursos de ação vão se moldando às (re)ações do/a ouvinte numa relação de interdependência com mais diversas modalidades acionadas e disponibilizadas pelos/as participantes de maneira sucessiva e coordenada no fluxo da interação, a transcrição multimodal favorece o alargamento da compreensão sobre “por que isso agora?” Nesse aspecto, é possível observar, por exemplo, como as condutas corporificadas acionadas por Sílvia estão estrutural, temporal e sequencialmente articuladas à produção dos cursos de ação de Lauriano (*e.g.*, a busca pela confirmação de Sílvia, l. 17).

Em comparação com o modelo de transcrição adotado no Excerto 2, o modelo de transcrição multimodal da conduta corporificada observado no Excerto 3 reflete mais adequadamente a articulação fina entre os recursos verbais e não verbais organizados temporal e sequentemente para construir as ações dessa interação face a face (Streeck *et al.* 2011b) em particular. Explicando de outra maneira, ao deixar de alinhar a temporalidade das condutas corporificadas de Sílvia, a transcrição do Excerto 2 não evidencia de maneira granular como suas trajetórias geram consequências para as ações interacionais subsequentes de Lauriano, para além do turno produzido imediatamente posterior à recomendação pela cessação do fumo.

Vale esclarecer que concordamos com Streeck *et al.* (2011b), Mondada (2001) e Hepburn e Bolden (2017), no sentido de não se tratar de defender a transcrição multimodal sobre a transcrição descritiva, mas de refletir sobre em que medida um ou outro modelo permite ao/à analista, antes de qualquer coisa, perceber de forma mais ou menos granular

o que está acontecendo no segmento interacional sob investigação. Mondada (2007b:811), ao problematizar as práticas de transcrição e a produção de excertos, argumenta que:

[...] mesmo que as possibilidades tecnológicas aprimorem a complexidade dos excertos e das práticas de transcrever, a prática de transcrever continua sendo uma atividade profundamente interpretativa: os interesses analíticos guiam a seleção e a relevância dos detalhes preservados nos excertos que permeiam cada passo da produção do dado.

Mondada (2007b:812) considera ainda que a “variabilidade [das práticas de transcrição] está relacionada à precisão dos excertos, seus níveis de granularidade, sua seletividade e à configuração da representação dos detalhes considerados relevantes. Nosso questionamento recai sobre em que medida a configuração da representação dos detalhes considerados relevantes é uma escolha do/a analista feita de maneira suficientemente informada, uma vez que diferentes modelos de transcrição suscitam distintos entendimentos sobre quais e como as diversas modalidades estão em jogo na formação das ações, não apenas do/a falante corrente, mas também do/a próximo/a.

Nesse ponto, defendemos que o/a analista mantenha-se atento/a às possibilidades de recursos disponíveis para manipulação de dados, que constam as atividades de gravação, notação, representação, textualização, digitalização, compressão, anonimização, animação, armazenamento etc. Em suma, entendemos que perceber, ver e fazer ver o óbvio são atividades de pesquisa de analistas da conversa. Todavia, em casos de dados audiovisuais, essas atividades requerem aprimoramento tecnológico para que o óbvio se manifeste de maneira precisa para o/a analista.

As várias camadas de visualização incorporadas em uma transcrição multimodal

Nesta seção, a disponibilidade de detalhes multimodais mobilizados pelos/as participantes será discutida a partir do recurso a *softwares* que auxiliam na exploração do vídeo, no alinhamento temporal das modalidades gestuais e verbais e das múltiplas ações realizadas e no

trabalho posterior de transcrição multimodal. A pergunta central explorada será: dentro da mentalidade analítica dos estudos interacionais de perspectiva multimodal, como visualizar um material em vídeo? Como desdobramento dessa pergunta, também buscamos responder a um segundo questionamento: como reconstruir os movimentos que vemos no vídeo e representá-los a partir de uma transcrição multimodal em formato textual?

No Excerto 4 temos uma interação entre duas participantes, Luiza, uma criança de 10 anos à época dos registros, e sua irmã adulta, Clara. Luiza tem o diagnóstico de TEA, transtorno do espectro autista⁹. Apesar de limitações linguístico-interacionais, Luiza realiza complexas articulações entre recursos multimodais distintos, verbais, gestuais e materiais presentes no ambiente interacional, organizados sequencial e temporalmente nas atividades de fala-em-interação das quais participa.

Na interação aqui analisada, Luiza introduz um turno de fala direcionado à sua irmã Clara com a estrutura verbal [tsau pam\]. As produções verbais de Luiza referem-se a um episódio ocorrido segundos antes, e ilustrado na Figura 3: Clara, ao sentar-se em uma poltrona, derruba, com o movimento de uma coberta, uma pipoca que estava no canto esquerdo da outra poltrona onde Luiza estava sentada (imagem 1).

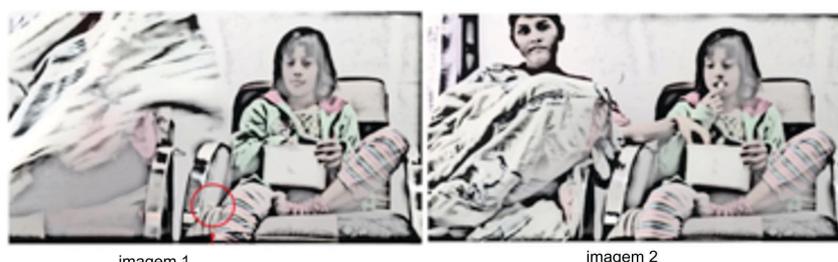


Figura 3 – Sequência: posições corporais de Clara e Luiza capturadas via câmera filmadora frontal.

9. As interações das quais Luiza participa estão documentadas no *corpus* CELA (Corpus para Estudo da Linguagem no Autismo), gerado pela pesquisadora Caroline Paola Cots. Para uma descrição mais detalhada do corpus vide Cruz (2017), Cruz e Cots (2017), Cots (2018), autorização do Comitê de Ética (CEP UNIFESP 59128416.3.00005505). A geração de *corporapara* investigação das interações envolvendo crianças autistas inscreve-se no escopo do projeto *Ao mínimo gesto* (Fernanda Miranda da Cruz, financiamento FAPESP-Fundação de Amparo à Pesquisa no Estado de São Paulo, processo 2018/07-565-7).

Ao sentar-se, o movimento de Clara com a coberta derruba, sem que ela perceba, uma pipoca que estava no canto do sofá em que Luiza está sentada (indicado na imagem 1 por um círculo). A informação de que havia uma pipoca no canto do sofá é central para nós, analistas, precisamente por ser relevante, antes de mais nada, para os próprios participantes da interação. Na imagem 2, Luiza percebe (visualmente) a ausência da pipoca. Clara, por sua vez, que está corporalmente em posição lateral à Luiza, não se orienta visualmente para o fato de que fora seu movimento que derrubara a pipoca; Clara também não se orienta para os movimentos de Luiza que indicam, por meio do direcionamento de seu olhar, que ela percebera a ausência da pipoca. Segundos depois, Luiza inicia uma conversa com Clara fazendo referência, construída corporificadamente, a esses dois fatos: o movimento de Clara com a coberta e a queda da pipoca.

Excerto 4: CELA004_2015.eaf

```

01   LUI   *   tsa::u\#           *
02   lui   *-balança a mão -> *
03   im    #ima. 3
04       **(.)
05       **-olha p/ Cla sorrindo-->(1. 12)
06       * #pa[m\ *
07       *bate a mão*
08   im    #ima. 4
09       (1.0)
10   CLA   * <((rindo))você derrubou no Δchão> // *
11       *----->olha p/ o chão----->*
12   lui           Δolha p/ o chão
13       (.)
14   lui   hihihhi hihihihihhi
15       (0.8)
16   LUI   <((rindo)).hΔhhhh>
17           Δ olha p/ o chão
18   CLA   não fui eu *Δhein//
19   cla           *Δ olha p/ Lui
20   lui           Δ olha p/ o chão
21   LUI   Δhihhihi hihihihhiΔ
22       Δ..olha p/Lui-->Δ olha p/ o chão
23       (.)
24   CLA   nã::o fui eu hein//
25       <(4.6)((abaixa-se, pega a pipoca no chão e leva à
26       boca))>
27   CLA   * dá aqui \ Nã:0\* *eu jogo fora\ *
28   cla   * -pega a pipoca--* *coloca no sofá*
29   LUI   ΔPA[M\ você\
30       Δbate a mão

```



Figura 4 – Captura de tela de vídeo do Excerto 4: imagens 3 e 4 sincronizadas aos turnos de fala na transcrição (#).

A transcrição em formato textual apresentada acima traz as seguintes informações: (1) Luiza inicia sua fala [*tsa::u*]; (2) olha para Clara; e (3) continua seu turno [*pa↑m/*]. Para cada uma das duas unidades lexicais *tsau* e *pam*, Luiza realiza um tipo distinto de gesto sincronizado à fala: o gesto de balançar a mão no ar, que acompanha a produção verbal *tsau* e o gesto de bater a mão no ar, que acompanha a produção verbal *pam*. Esse padrão de construção sintático-gestual acontece novamente ao longo da interação, ainda que com algumas modificações

Após a introdução do turno de fala inicial de Luiza, Clara faz uma reformulação [*linha 24: você derrubou no +chão//*]. Tal ação deixa ver que Clara interpreta o turno de Luiza como uma referência à queda da pipoca. Até a linha 24, ambas estão orientadas para o mesmo referente, pipoca. No entanto, o trabalho interacional a partir desse momento da interação está em estabelecerem e alinharem-se sobre *quem/o que* derrubara a pipoca. No decorrer da interação, à medida que recebe a interpretação de Clara sobre quem teria derrubado a pipoca, Luiza realiza um reparo em seu turno de fala e incrementa-o, acrescentando o agente (no caso, *você* referindo-se à Clara) à parte da estrutura produzida anteriormente, na posição de coda, resultando em [*PA↑M/ você (linha 43)*]. Essa estrutura verbal é acompanhada do mesmo gesto de bater, o que nos permite identificar uma correlação semântico-sintática entre os gestos (gesto de bater e postura corporal) e o enunciado verbal produzido por Luiza.

As produções verbais produzidas por Luiza, embora não contemham uma palavra ou unidade morfológica reconhecível do léxico em português, contêm, a partir de uma descrição gramático-interacional (Auer 2009; Mondada 2012; 2014a; 2014b): ordenamento sintático; pausa; fenômenos segmentais reconhecíveis, como alongamento, entonação ascendente ou descendente; acentuação de ênfase tônica sobre a sílaba núcleo. Sincronicamente coordenado às produções verbais ‘tsau’ e ‘pam’, Luiza produz um conjunto de referências corporificadas (Mondada 2012), como o olhar e o corpo direcionados para o local (imagem 1, círculo desenhado) onde estava a pipoca. Somado a isso, Luiza realiza sistematicamente os dois tipos de gestos (balançar e bater a mão) para cada unidade lexical. Em suma, a organização sintático-interacional da fala de Luiza articula-se, em termos de ordenamento, de sequencialidade e de temporalidade aos elementos gestuais e ao elemento material “pipoca” tornados disponíveis. Tratemos mais pormenorizadamente dessa disponibilidade.

No que diz respeito aos ângulos de visão de Clara e do/a analista, a câmera fixada por Clara, que fez os registros das interações em ambiente familiar, oferece ao/à analista um ângulo frontal privilegiado com relação ao ângulo lateral que Clara tem da pipoca. Porém, do ponto de vista do/a analista, o acesso a um detalhe (a presença da pipoca) é obtido não necessariamente por conseguir visualizar um elemento físico-material presente no espaço através de um ângulo privilegiado da câmera. O fato é que durante as primeiras visualizações deste excerto, a pipoca também permaneceu visualmente imperceptível para o/a analista. Sua visibilidade e materialidade foi recuperável, retrospectivamente, durante o próprio trabalho de transcrição, ao acompanharem-se as trajetórias de direcionamentos de olhar de Luiza em seu trabalho de referência corporificada à pipoca e ao se reconstruir milimetricamente a temporalidade em termos de sincronia e coordenação dos movimentos de Luiza e Clara.

De certa forma, a pipoca apenas fora percebida após uma espécie de rebobinagem: um ângulo da câmera deixa ver aqueles movimentos corporais de Luiza que antecederam seu turno de fala; por isso, foi apenas ao submetermos o material registrado em vídeo a um trabalho metodológico de visualização e transcrição específicos, guiado pelos movimentos de direcionamentos do olhar de Luiza para a pipoca, que

tanto esses movimentos como a referida pipoca tornaram-se disponíveis e interacionalmente relevantes também para o/a analista.

Dar conta, analiticamente, dessa articulação temporal fina torna-se um desafio. Ver e tornar visível (ou fazer ver) essa articulação exige que o/a analista dê conta de temporalidades complexas, como argumenta Mondada (2016). Essas temporalidades, que envolvem coordenações, simultaneidades, sequencialidades, acontecem não apenas entre os participantes, entre as ações, mas entre as próprias modalidades (verbais e gestuais).

Ao longo de toda a interação representada no Excerto 4, em nenhum momento, referentes como “pipoca”, “queda da pipoca” ou “coberta” foram produzidos verbalmente, mas são construídos corporal e materialmente. Durante o desenrolar dessa interação, Luiza e Clara se alinham a respeito do que está sendo tratado, *i.e.* a queda da pipoca, mas não chegam a um entendimento mútuo sobre o que realmente aconteceu (e que acabou gerando a queda da pipoca) – fato esse a que Luiza se remete. A análise e a inteligibilidade da fala de Luiza dependem necessariamente da sua inscrição corporificada na interação, das percepções e referências que ela faz à sua ecologia material e de uma correlação estrutural entre língua, corpo e mundo material. É dentro do tópico “dar conta da inteligibilidade da coordenação fina” que o vídeo e o trabalho de reconstrução dos detalhes multimodais realizado pelo/a analista podem ser abordados em termos de camadas de visualização a serem operadas pelos participantes e pelos analistas para que tal coordenação fina possa ser reconstruída, aferida e publicizada.

A pipoca, assim, não é um detalhe, como poderíamos julgar se estivéssemos vendo um vídeo leigamente e sem propósitos analíticos. Ao contrário, ela é a chave de entendimento do que Luiza constrói em seu turno, apesar de suas limitações verbais. Clara e Luiza não encerram a interação com a mesma versão sobre quem/o que derrubou a pipoca. Para o analista, por sua vez, ao visualizar o vídeo, seu olhar não é guiado por ver ou não ver a pipoca que estava no canto do sofá e que fora derrubada. Metodologicamente, o analista visualiza o vídeo atento a *quando* tanto a pipoca quanto quem a derrubou emergem como relevantes para a inteligibilidade do/a analista das produções verbais

tchau, pam, você. Esse *quando* aqui é entendido no sentido de “*quando de um contexto*” dado por Erickson e Shultz (2013[1981]).

Para dar conta com precisão do *quando de um contexto* e das temporalidades complexas das interações, incluindo a temporalidade de recursos multimodais distintos, como vimos nas três interações ao longo deste artigo, o recurso a *softwares* pode ser útil. O trabalho de transcrição multimodal pode variar consideravelmente segundo a escolha em se usar ou não *softwares*. Por isso, tal escolha também gera questões importantes para o campo de pesquisas interacionais baseadas em vídeo, como apontam (Knoblauch *et al.* 2006:16).

A transcrição multimodal dos excertos foi realizada a partir do auxílio da ferramenta ELAN (Wittenburg *et al.* 2006, versão 5.0, vide Figuras 5 a 7). Assim, há um trabalho invisível dos/as videoanalistas envolvido nos estudos multimodais que nem sempre é tornado público, como apontam Knoblauch *et al.* (2006). Para publicizar parte deste trabalho, descreveremos como o ELAN foi utilizado especificamente para exploração dessa situação interativa e do corpus envolvendo interações com crianças autistas do qual o Excerto 4 fora extraído. O ELAN foi utilizado para: (a) visualização do vídeo, ora em velocidade normal, ora em velocidade reduzida, para acompanharmos em câmera lenta as trajetórias de gestos; (b) inserção de anotações relativas aos elementos gestuais e verbais; (c) alinhamento temporal mais preciso entre as modalidades verbais e gestuais (particularmente os gestos das mãos e braços, os direcionamentos de olhar, as ocorrências de riso, os direcionamentos de torso) e a duração dos momentos de ausência de fala; (d) extração dos contornos entonacionais das produções verbais para leitura e análise posterior em outro software, o *Praat* (Boersma e Weenink, 1992, versão 6.0.25). Esse último procedimento é fundamental para analisarmos a correlação entre aspectos prosódico-entonacionais e os gestos em produções verbais que não correspondem a formas convencionalizadas da língua como é o caso de *tsau* e *pam*. Os procedimentos descritos em (a), (b), (c) e (d) estão ilustrados nas figuras abaixo.

Fernanda Miranda da Cruz, Ana C. Ostermann, Daniela N. Pinheiro Andrade, Minéia Frezza

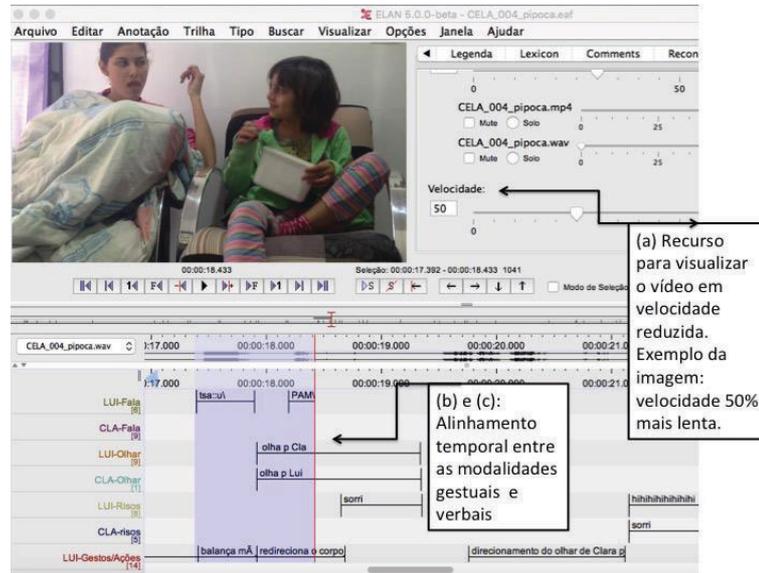
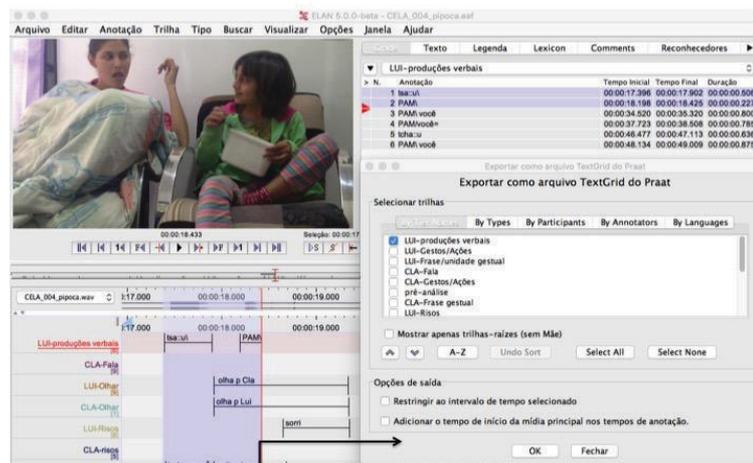


Figura 5 – Captura de tela do ELAN contendo sistema de notação aplicado ao segmento audiovisual referente ao Excerto 4. Destaque aos procedimentos (a), (b), (c).



Captura de tela: função “Exportar como arquivo Text Grid do Praat” aplicada às produções verbais de Luiza.

Figura 6 – Captura de tela ELAN com destaque para o procedimento (d): função de extração em formato Text Grid dos contornos entonacionais das produções verbais de Luiza alinhadas às notações não-verbais para leitura em *Praat*.

O trabalho técnico-metodológico e analítico com dados interacionais audiovisuais

(d) Extração dos contornos entonacionais das produções verbais *tsau* e *pam* de Luiza relacionadas às produções gestuais a partir da interface entre ELAN e Praat. Linhas 01-07 da transcrição multimodal do excerto 4 apresentada em formato textual.

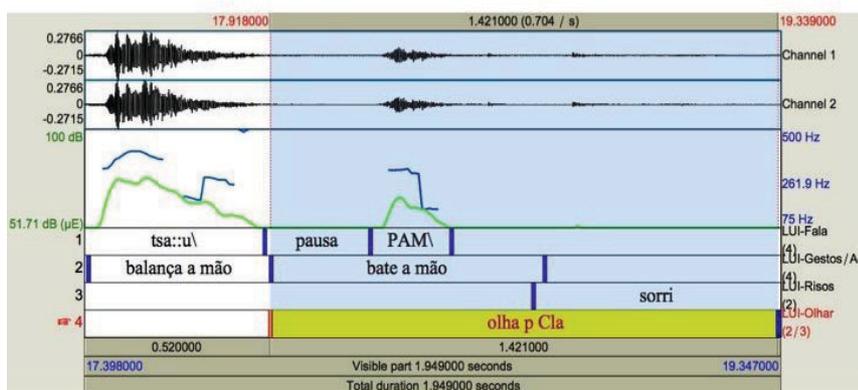


Figura 7 – Captura de tela da janela do software *Praat* contendo informações extraídas do ELAN.

Posteriormente, o resultado desse alinhamento temporal realizado com ELAN é a base para a realização da transcrição multimodal apresentada acima em formato textual (convenção Mondada 2014d).

Essa última análise leva-nos à possibilidade de sistematização de algumas camadas de visualizações imbricadas na prática de tratar analiticamente um vídeo na condução de estudos linguístico-interacionais. Equipamentos como câmeras e gravadores e as várias possibilidades reconstitativas descritas aqui, que vão desde práticas de transcrição, escolhas entre descrever ou transcrever temporal e sequencialmente modalidades não-verbais até os recursos oferecidos por softwares como o ELAN e o *Praat*, passam a ser incorporados como elementos que preservam e ao mesmo tempo configuram as análises empreendidas. No que diz respeito especificamente às inovações tecnológicas e seu acesso por parte do/a analista, podemos pensar que se o uso de uma ferramenta técnico-tecnológica pode influenciar significativamente o trabalho de transcrição e representação dos dados, podemos então depreender que a relação entre saber técnico-tecnológico e geração de

dados multimodais transcritos não é apenas um detalhe do tema, mas senão mais um aspecto sobre a própria produção de conhecimento no campo dos estudos interacionais (Cruz, 2018).

Considerações finais

Os quatro excertos discutidos ao longo deste artigo nos permitiram reconstruir descritivamente nosso exercício de visualização, transcrição, análise e apresentação multimodal no qual uma mentalidade visual do/a analista transcritor/a emerge ou destaca-se no curso da própria atividade de transcrever e analisar. As pesquisas que temos desenvolvido sobre a interação social ancoradas em mentalidade multimodal oferecem a possibilidade de sistematizarmos várias camadas de visualização discutidas aqui: (a) a prática do registro visual; (b) o resultado do registro; (c) os tratamentos para alinhamento temporal das modalidades feito com softwares como ELAN e *Praat* (d) e as próprias escolhas de transcrição envolvidas na investigação e exploração da forma como construímos nossas interações em contextos diversos e ecologicamente complexos.

A sistematização de questões oriundas de videoanálises mostra que o trabalho com vídeo é em si uma temática importante de reflexão teórico-analítica, como sugerem Knoblauch *et al.* (2006). Mostramos ainda como essas várias camadas sistematizadas não são periféricas, menos ainda marginais à condução dos procedimentos de investigação em torno das interações sociais e de sua organização. Ao contrário, apresentam-se como fundamentais para conduzir a reflexão sobre *como* efetivamente os/as participantes exploram intersubjetivamente o mundo físico em suas interações cotidianas.

Para finalizar, defendemos que cada detalhe supostamente “mínimo” passa a ser visível por meio de um trabalho de reconstrução dos detalhes multimodais feito pelo/a analista. Contudo, há que se reiterar que o/a analista não apenas vê *através* da câmera, mas vê *com* ela e vê, principalmente, *com* os movimentos corporais realizados pelos/as participantes. Assim, a visibilidade de um detalhe não se resume apenas ao que um vídeo é capaz de “captar” em termos de detalhes visuais da cena, mas se constitui principalmente do trabalho de reconstrução

temporal e sequencial de tais detalhes. Esse trabalho, por sua vez, é feito a partir de um exercício minucioso de identificar e medir as unidades de ações que emergem no vídeo, um pouco como o trabalho do médico diante da tela do ultrassom, que frente a movimentos do feto, congela unidades anatômicas que emergem e aí lança-se a medi-las. Para fazer ver as unidades analíticas, congelamos, por meio de práticas de visualização repetidas, de transcrição e de representação por imagens aquilo que o vídeo ainda carrega de natural das interações: um constante movimento entre os/as participantes.

Referências

- ANDRADE, Daniela Negraes Pinheiro. 2016. Recomendações e prescrições para cuidados de saúde no pós-alta: a investigação de um programa educativo a pacientes cardiopatas sob uma perspectiva interacional. Tese (Doutorado em Programa de Pós-Graduação em Linguística Aplicada) - Universidade do Vale do Rio dos Sinos (Unisinos). Disponível online: <<http://www.repositorio.jesuita.org.br/handle/UNISINOS/5267>>.
- AUER, Peter. 2009. Online-syntax: thoughts on the temporality of spoken language. *Language Sciences*, 31(1): 1-3.
- BERGMANN, Jörg. 1985. Flüchtigkeit und methodische Fixierung sozialer Wirklichkeit: Aufzeichnungen als Daten der interpretativen Soziologie. In: BONSS, Wolfgang; HARTMANN, Heinz (Eds.). *Entzauberte Wissenschaft*. Göttingen: Otto Schwarz.
- BOERSMA, Paul; WEENINK, David. *Praat: doing phonetics by computer*. Versão 6.0.28, de 2017 [1992].
- BÜSCHER, Monika. 2005. Social life under the microscope? *Sociological Research Online*. 10. Disponível em: <<http://www.socresonline.org.uk/10/1/buscher.html>>. Acesso em: 12 nov. 2015.
- COTS, Caroline Paola. 2018. Recursos interacionais multimodais mobilizados por uma criança com Transtorno do Espectro Autista (TEA) em brincadeiras familiares. Dissertação (Mestrado em Programa de Pós-Graduação, Escola de Filosofia, Letras e Ciências Humanas, em Letras, Universidade Federal de São Paulo. (UNIFESP).
- CRUZ, Fernanda Miranda da. 2017. Elementos para uma análise multimodal da interação: um exemplo de correlação linguístico-gestual no autismo. In: GONÇALVES-SEGUNDO, Paulo R.; MODOLO, Artur R.; SOUSA, Daniel, R.; FERREIRA, Filipe M.;

- COAN, Giovanna I.; BRITTO-COSTA, Leticia F. (Orgs.). *Texto, discurso e multimodalidade: perspectivas atuais*. São Paulo: Editora Paulistana, p. 158-179.
- _____. 2018. Documentação e investigação multimodal de interações envolvendo crianças com autismo: corpo, linguagem e mundo material. *Revista Calidoscópio*, 16 (2): 179-193.
- CRUZ, Fernanda Miranda da; COTS, Caroline Paola; LUIZ, Reuel Alves. 2017. Linguagem em Micro- Acontecimentos: corpo, gestos e fala explorados em uma análise multimodal de interações envolvendo uma criança autista. *Revista Intercâmbio*, v. XXXIV: 34-57, 2017. São Paulo: LAEL/Pontífica Universidade Católica de São Paulo –PUC-SP.
- ERICKSON, Frederick; SHULTZ, Jeffrey. 2013[1981]. “O quando” de um contexto: questões e métodos na análise da competência social. In: RIBEIRO, Branca T.; GARCEZ, Pedro M. (Orgs). *Sociolinguística Interacional*. Tradução de Pedro M. Garcez e Larissa Surek-Clark. São Paulo: Edições Loyola. p. 215-246.
- GOODWIN, Charles. 1981. *Conversational organization: interaction between speakers and hearers*. New York: Academic Press.
- _____. 2010. Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32(10): 1489-1522.
- GOODWIN, Marjorie H. 1980. Processes of mutual monitoring implicated in the production of description sequences. *Sociological Inquiry*, 50: 303-317.
- HEPBURN, Alexa; BOLDEN, Galina. 2017. *Transcribing for social research*. Los Angeles: Sage Publications Ltd.
- HERITAGE, John. 1984. A change-of-state token and aspects of its sequential placement. In: ATKINSON, J.; HERITAGE, J. *Structures of social action: studies in conversation analysis*. New York: Cambridge University Press. p. 299-345.
- HERITAGE, John; LINDSTRÖM, Anna. 2012. Advice-giving: terminable and interminable. In: LIMBERG, Holger; LOCHER, Miriam (Eds.). *Advice in discourse*. Amsterdam: John Benjamins Company. p. 169-194.
- HINDMARSH, Jon; HEATH Christian. 2000. Embodied reference: a study of deixis in workplace interaction. *Journal of Pragmatics*, 32: 1855-1878.
- KENDON, Adam. 2004. *Gesture: visible action as utterance*. New York: Cambridge University Press.
- KNOBLAUCH, Hubert; SCHNETTLER, Bernt; RAAB, Jürgen; SOEFFNER, Hans-Georg (Orgs.). 2006. *Video analysis methodology*

and methods: qualitative audiovisual data analysis in sociology, Berlin: Peter Lang.

- LEVINSON, Stephen. 2013. Action formation and ascription. In: SIDNELL, Jack; STIVERS, Tanya (Eds). *The handbook of conversation analysis*. Boston: Wiley-Blackwell. p. 103-130.
- MONDADA, Lorenza. 2013. Interactional space and the study of embodied talk-in-interaction. In: AUER, Peter; HILPERT, Martin; STUKENBROCK, Anja; SZMRECSANYI, Benedikt (Orgs.). *Space in language and linguistics: geographical, interactional and cognitive perspectives*. Berlin, Boston: De Gruyter. p. 247-275.
- _____. 2003a. Observer les activités de la classe dans leur diversité: choix méthodologiques et enjeux théoriques. In: PEREIRA Joan; NUSSBAUM, Luci; MILIAN, Marta (Orgs). *L'educacio linguistica en situacions multiculturals i multilingues*. Barcelona: Institut de Ciència e de la education de la Universitat de Barcelona. p. 49-70.
- _____. 2003b. Working with video: how surgeons produce video records of their actions, *Visual Studies*, 18(1): 58-72.
- _____. 2006. Video Recording as the reflexive Preservation of Fundamental Features for Analysis. In: KNOBLAUCH, Hubert; SCHNETTLER, Bernt; RAAB, Jürgen; SOEFFNER, Hans-Georg (Eds). *Video Analysis*. Bern: Lang, 51-68.
- _____. 2007a. Multimodal resources for turn-taking: pointing and the emergence of possible next speakers. *Discourse Studies*, 9(2): 195-226.
- _____. 2007b. Commentary: transcript variations and the indexicality of transcribing practices. *Discourse Studies*, 9(6): 809-821.
- _____. 2008a. Using video for a sequential and multimodal analysis of social interaction: Videotaping institutional telephone calls. *FQS - Forum: Qualitative Social Research*, 9 (3).
- _____. 2008b. L'analyse de "collections" de phénomènes multimodaux en linguistique interactionnelle: à propos de l'organisation systématique des ressources gestuelles en début de tour. *Cahiers de praxématique*, 50: 21-66.
- _____. 2008c. A Documenter l'articulation des ressources multimodales dans le temps: la transcription d'enregistrements vidéos d'interactions. In: BILGER, Mireille (Org.). *Données orales: les enjeux de la transcription*. Perpignan: Presses Universitaires de Perpignan. p. 127-156.
- _____. 2009. Video recording practices and the reflexive constitution of the interactional order: some systematic uses of the split-screen technique. *Human Studies*, 32(1): 67-99.

- _____. 2012. Organisation multimodale de la parole-en-interaction: pratiques incarnées d'introduction des référents. *Langue française*, 175 (3): 129-147.
- _____. 2014a. Pointing, talk, and the bodies: reference and joint attention as embodied interactional achievements. In: SEYFEDDINIPUR, Mandana; GULLBERG, Marianne (Orgs.). *From gesture in conversation to visible action as utterance: Essays in honor of Adam Kendon*, Londres: John Benjamins Publishing Company. p. 95-124.
- _____. 2014b. The local constitution of multimodal resources for social interaction. *Journal of Pragmatics*, 65: 137-156.
- _____. 2014c. The surgeon as a camera director: maneuvering video in the operating theatre. In: BROTH, Matias; LAURIER, Eric; MONDADA, Lorenza (Eds). *Video at Work*. London: Routledge. p. 97-132.
- _____. 2014d. Conventions for multimodal transcription, versão 3.0.1. Disponível online: <https://franz.unibas.ch/fileadmin/franz/user_upload/redaktion/Mondada_conv_multimodality.pdf>.
- _____. 2016. Challenges of multimodality: language and the body in social interaction. *Journal of Sociolinguistics*, 20(2): 194-225.
- OSTERMANN, Ana Cristina; FREZZA, Minéia. 2018a. Understanding what no longer is: Challenges to the (re)establishment of intersubjectivity when talking about referents that move (and disappear). 2018. Apresentação de Trabalho *5th International Conference of Conversation Analysis – ICCA-18*, Universidade de Loughborough, Inglaterra.
- _____. 2018b. Zooming into what matters: The double bind affordance of fetal ultrasound technology. (Apresentação de Trabalho no Congresso *VALS-ASLA 2018: A Video Turn in Linguistics? Methodologie – Analisi – Applications*, Universidade de Basel, Suíça.
- _____. 2018c. Ephemeral visual referents and their consequences to intersubjectivity in fetal ultrasound scans. (Apresentação de Trabalho no *Colloque IMPEC 2018: “Interactions Multimodales Par ECran”*, Lyon 2, França.
- ROSSANO, Federico. 2012. *Gaze behavior in face-to-face interaction*. Unpublished PhD Dissertation. Radboud University Nijmegen, Nijmegen.
- SCHEGLOFF, Emmanuel. 2007. *Sequence organization in interaction: a primer in conversation analysis*. Cambridge: Cambridge University Press, v. 1.
- SELTING, Margret; AUER, Peter. 2011. A system for transcribing talk-in-interaction: GAT2 translated and adapted for English by Elizabeth

O trabalho técnico-metodológico e analítico com dados interacionais audiovisuais

- Couper-Kuhlen and Dagmar Barth-Weingarten. *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion*, 12: 1-51.
- STEVANOVIC, Melisa; PERÄKYLÄ, Anssi. 2012. Deontic authority in interaction: the right to announce, propose, and decide. *Research on Language and Social Interaction*, 45(3): 297-321.
- STREECK, Jürgen; GOODWIN, Charles; LeBARON, Curtis. 2011a. *Embodied Interaction: language and body in the material world*. New York: Cambridge University Press.
- _____. 2011b. Embodied interaction: language and body in the material world: an introduction. In: STREECK, Jürgen; GOODWIN, Charles; LeBARON, Curtis. *Embodied interaction: language and body in the material world*. Cambridge: Cambridge University Press. p. 1-26.
- WITTENBURG, Peter; BRUGMAN, Hennie; RUSSEL, Albert; KLASSMANN, Alex; SLOETJES, Han. 2006. ELAN: a professional framework for multimodality research. In: *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*. p. 1556-1559.

Recebido em: 24/01/2018

Aprovado em: 18/09/2019

Anexo: Quadros de convenções de transcrição

Quadro 1 – Convenção de transcrição multimodal baseada em Mondada (2014d)

I. Aspectos linguístico-verbais	
Xxx	Segmento ininteligível
.	Micro pausas (inferiores a 0,3 segundos)
(0.4)	Pausas medidas com ELAN
:	Alongamento silábico
.h	Marca inspiração do locutor
/ e \	Entoações ascendente e descendente
//	Entonação ascendente acentuada (de pergunta)
<u>sublinhado</u>	Ênfase particular (intensidade, acento)
maÍSCULA	Volume forte de voz
o o	Volume baixo de voz
II. Trajetória dos gestos	
+----+	Delimitação de início e fim do gesto

+ , *	A fala transcrita comporta os símbolos gráficos indicadores de gestos (+, *, por exemplo) posicionados no momento em que são realizados com relação à fala.
(1.0) +* (0.2)	Se um gesto começa no meio de uma pausa, segmenta-se a pausa. (ex. 1.2 segundos= 1,0 + 0,2).
---->121	Continuação do gesto até a linha 21 (exemplo) do excerto.
----->+	Se um gesto continua nas linhas seguintes, sua descrição é seguida de uma flecha que remete ao símbolo que delimita sua finalização.
----->>	Continuação do gesto até o fim do excerto
III. Trajetória temporal dos gestos	
.....	Emergência
.....>	Finalização ou retração
-----	Manutenção
IV. Inserção de imagens	
ima	Imagens reproduzidas no texto da transcrição extraídas por captura de tela do vídeo.
#	Indica o exato momento que corresponde à imagem (ima) sincronizado ao turno de fala

Quadro 2 – Convenção para riso baseada no sistema GAT2 (Selting et al. 2011) para o Excerto 1

{{rindo} texto}	Turnos ou palavras pronunciadas rindo
-----------------	---------------------------------------