

Quando o "... " importa: análise comparada de modalidades de transcrição de hesitações em entrevistas do Corpus Roda Viva

Valeria Vieira dos Santos

Programa de Pós-Graduação em Linguística (PPGL)
Universidade Federal de São Carlos (UFSCar)

valeriavsvsantos93@gmail.com

Abstract. *This paper investigates how hesitations — pauses, lengthenings, truncations, repetitions, and lexical fillers — are represented across different transcription modalities in Brazilian Portuguese journalistic interviews. Drawing on a pilot corpus of 95 speaker turns from the Roda Viva program (TV Cultura), we compare a faithful transcription (preserving hesitation markers) with a sanitized version and a conversational-protocol transcription (Jefferson; GAT2). Results show that each modality systematically suppresses, normalizes, or preserves hesitation markers in distinct ways, revealing the need for an explicit annotation protocol. We propose an initial functional typology of hesitations in five formal categories and discuss implications for oral corpus construction and NLP annotation tasks.*

Resumo. *Este trabalho investiga como as hesitações — pausas, prolongamentos, truncamentos, repetições e preenchimentos lexicais — são representadas em diferentes modalidades de transcrição de entrevistas jornalísticas do português brasileiro. A partir de um corpus piloto de 95 turnos de fala do programa Roda Viva (TV Cultura), comparamos uma transcrição fiel (com marcas hesitativas) com uma versão higienizada e com uma transcrição em protocolos conversacionais (Jefferson; GAT2). Os resultados mostram que cada modalidade trata sistematicamente as marcas hesitativas de forma distinta, revelando a necessidade de um protocolo de anotação explícito. Propomos uma tipologia funcional inicial em cinco categorias formais e discutimos implicações para a construção de corpora orais e tarefas de anotação em PLN.*

1. Introdução

A transcrição de fala espontânea é uma etapa fundamental na construção de corpora orais computacionalmente utilizáveis. Diferentes práticas de transcrição, como jornalísticas, automáticas e conversacionais, são orientadas por objetivos distintos e produzem

representações diversas de um mesmo evento de fala, com consequências diretas sobre a qualidade e a completude dos recursos gerados (Vale et al., 2024). Entre os fenômenos mais afetados por essas diferenças estão as *hesitações*: pausas, prolongamentos, truncamentos, preenchimentos lexicais (como *ééé*, *ahn*, *então*) e repetições. Elementos recorrentes da fala espontânea que representam aproximadamente 6% dos tokens em interações orais (Fox Tree, 1995; Shriberg, 2001).

A transcrição **jornalística** tem como objetivo a legibilidade pública: remove marcas interpretadas como ruído. A transcrição **automática**, gerada por sistemas como YouTube/Google, produz texto fluente em larga escala, mas com baixa sensibilidade a fenômenos disfluente. Já as transcrições **conversacionais**, com protocolos como Jefferson (2004) e GAT2 (Selting et al., 2009), registram com precisão a materialidade da fala para fins analíticos. O resultado é que as hesitações são apagadas em algumas modalidades, normalizadas em outras, e detalhadamente preservadas apenas nos protocolos especializados, criando heterogeneidade que compromete a consistência do corpus para uso em PLN.

Este trabalho parte de um corpus piloto de 95 turnos de fala coletados de uma entrevista do programa *Roda Viva* (TV Cultura), integrante do Corpus Roda Viva (Vale et al., 2024), e tem como objetivo: (i) comparar sistematicamente o tratamento das hesitações em duas modalidades de transcrição, fiel e higienizada; (ii) contrastar com exemplos em protocolos conversacionais; e (iii) propor uma tipologia funcional inicial das hesitações com base nas diferenças observadas. O objetivo final é contribuir com fundamentos linguísticos para a construção de um protocolo de anotação explícito e reutilizável para o Corpus Roda Viva.

2. Hesitação como fenômeno linguístico

As hesitações são fenômenos formalmente descritíveis: cada tipo possui propriedades distribucionais, prosódicas e funcionais que permitem categorização sistemática (Shriberg, 2001; Corley & Stewart, 2008). Não se trata de ruído ou falha de desempenho, mas de recursos da fala que cumprem funções específicas na organização do discurso oral: indicam planejamento linguístico em curso, sinalizam dificuldades lexicais ou estruturais, e participam da gestão de turnos em interações assimétricas como entrevistas jornalísticas.

Com base na literatura (Shriberg, 2001; Corley & Stewart, 2008; Zayats & Ostendorf, 2019; Marcuschi, 2003) e nos dados do corpus piloto, identificamos cinco categorias formais de hesitação:

Pausas, que são os intervalos de silêncio dentro do turno de fala, sem transferência de turno ao interlocutor. Podem ser breves (.) ou longas (..) e ocorrem em posições de planejamento discursivo, fronteiras sintáticas ou momentos de busca lexical (Shriberg,

2001). Em protocolos conversacionais como Jefferson (2004), sua duração é estimada; nas transcrições jornalísticas e automáticas, são sistematicamente apagadas.

Preenchimentos lexicais (fillers), vocalizações como ééé, ahn, e usos hesitantes de então e né funcionam como sinais de que o falante mantém o turno enquanto planeja o enunciado (Marcuschi, 2003; Corley & Stewart, 2008). São os marcadores mais frequentemente apagados em transcrições jornalísticas.

Truncamentos, representados por interrupções abruptas de palavras antes de seu término, geralmente seguidas de reformulação (ex.: Mina- Minas, dem- práticas). Indicam replanejamento imediato e são representados com hífen nos protocolos conversacionais, mas frequentemente omitidos ou corrigidos nas demais modalidades.

Prolongamentos, sendo as extensões anormais de sílabas, representadas por :: (ex.: Então::). Indicam hesitação prosódica e funcionam como estratégia para manter o turno. São raramente preservados em transcrições que não seguem protocolos especializados.

E Repetições, reiteraões de sílabas, palavras ou sintagmas em contexto de replanejamento discursivo (ex.: do do, para... para...). Marcuschi (2003) as distingue de repetições retóricas intencionais pela co-ocorrência com outros marcadores hesitativos e pela ausência de função enfática.

A Tabela 1 sintetiza as cinco categorias com seus marcadores e funções:

Tabela 1. Tipologia funcional inicial das hesitações — categorias e marcadores

Categoria	Marcador (ex.)	Função / descrição
Pausa	(.) (..) (...)	Intervalo de silêncio dentro do turno; pode ser breve (.) ou longa (..)
Preenchimento lexical	ééé, ahn, éé, então	Vocalizações que preenchem pausas e sinalizam continuidade de turno
Truncamento	palavra- (ex.: Mina-)	Interrupção abrupta de palavra, seguida de reformulação ou nova tentativa
Prolongamento	:: (ex.: Então::)	Extensão anormal de sílaba; indica hesitação prosódica e busca lexical
Repetição	do do, na na, para... para...	Reiteração de sílaba, palavra ou sintagma; marca replanejamento

3. Modalidades de transcrição e seus protocolos

Cada modalidade de transcrição é orientada por um objetivo diferente, o que determina diretamente o tratamento das hesitações. A transcrição **jornalística** prioriza a inteligibilidade do conteúdo: edita a fala para parecer fluente, apagando sistematicamente

preenchimentos, pausas e truncamentos. A transcrição **automática**, gerada por sistemas de reconhecimento automático de fala como os do YouTube/Google, otimiza velocidade e cobertura, produzindo texto normalmente sem marcas de oralidade. Essas limitações já foram documentadas para o português brasileiro por Lima & Campelo (2024) e Souza et al. (2025).

Já os protocolos conversacionais têm propósito analítico: o **sistema Jefferson** (2004) registra a estrutura interacional da fala (turnos, sobreposições, pausas cronometradas), e o **GAT2** (Selting et al., 2009) acrescenta detalhamento prosódico (entonação, ênfase, intensidade). Juntos, esses protocolos oferecem a representação mais fiel da hesitação como fenômeno da fala espontânea, mas são trabalhosos e raramente aplicados a larga escala.

A Tabela 2 sintetiza as diferenças entre as três modalidades em relação ao tratamento das hesitações.

Tabela 2. Tratamento das hesitações por modalidade de transcrição

Categoria	Jornalística	Automática	Conversacional
Pausa	Apagada	Apagada	Marcada e cronometrada
Preenchimento lexical	Apagado	Parcialmente mantido	Preservado e transcrito
Truncamento	Apagado ou corrigido	Parcialmente mantido	Marcado com hífen
Prolongamento	Apagado	Apagado	Marcado com ::
Repetição	Apagada	Apagada	Preservada

4. Corpus e metodologia

O corpus desta pesquisa é o Corpus Roda Viva, coordenado pelo Prof. Dr. Oto Araújo Vale (UFSCar), que organiza 713 entrevistas do programa *Roda Viva* (TV Cultura) com transcrições jornalísticas, arquivos audiovisuais e versões em Universal Dependencies (Vale et al., 2024). Para este trabalho, utilizamos um corpus piloto composto por **95 turnos de fala** extraídos de uma entrevista completa, com dois locutores principais e mais cinco participantes, totalizando sete locutores distintos. Por se tratar de uma única entrevista, os resultados aqui apresentados têm caráter exploratório; a ampliação da amostra, com seleção intencional de entrevistas por período, área temática e perfil de entrevistados, é etapa prevista da pesquisa em andamento.

O corpus piloto foi organizado em três colunas: (i) **FALA_FIEL**: transcrição com marcadores de hesitação preservados, seguindo convenções próximas ao sistema Jefferson; (ii) **FALA_HIGIENIZADA**: versão sem marcas hesitativas, equivalente a uma transcrição jornalística; e (iii) **NOTA_HUMANA**: score de 0 a 100 indicando o grau de semelhança entre as duas versões, atribuído manualmente (100 = idênticas; 0 = máxima divergência). Adicionalmente, para fins de exemplificação qualitativa, utilizamos transcrições nos protocolos Jefferson e GAT2 de uma entrevista publicada anteriormente (Vale et al., 2024).

A **NOTA_HUMANA** foi atribuída manualmente como medida de similaridade percebida entre a **FALA_FIEL** e a **FALA_HIGIENIZADA** de cada turno: valores próximos de 100 indicam que as duas versões são praticamente idênticas, ou seja, o turno não continha marcas hesitativas relevantes; valores próximos de 0 indicam máxima divergência, com apagamento extenso de marcadores. A nota foi atribuída por uma única anotadora, linguista, mestre em Linguística, com experiência em transcrição de fala e Análise da Conversação. Optou-se pela escala contínua de 0 a 100, em vez de escalas ordinais do tipo Likert, por sua maior granularidade e pela compatibilidade direta com métricas computacionais de similaridade normalizadas (0–1), que serão empregadas na etapa subsequente da pesquisa. Trata-se, ainda assim, de uma métrica exploratória, baseada em julgamento humano, que permite estratificar o corpus piloto por densidade hesitativa sem depender de contagem automática e que será substituída por esquemas de anotação validados por Acordo Inter-Anotadores (IAA).

As transcrições conversacionais foram produzidas com auxílio do software ELAN (Wittenburg et al., 2006), que permite o alinhamento entre anotação textual e os arquivos de áudio e vídeo das entrevistas, viabilizando a verificação direta das marcas hesitativas no sinal acústico original.

A análise quantitativa consistiu em: (i) contagem de ocorrências por categoria de hesitação na **FALA_FIEL**; (ii) verificação de supressão na **FALA_HIGIENIZADA**; e (iii) análise da distribuição do score **NOTA_HUMANA** para medir o impacto do apagamento.

5. Análise comparada

O corpus piloto contém **903 marcadores de hesitação** na **FALA_FIEL**, distribuídos conforme a Tabela 3. Todos são sistematicamente apagados ou normalizados na **FALA_HIGIENIZADA**.

Tabela 3. Frequência de marcadores hesitativos no corpus piloto (95 turnos)

Categoria	Ocorrências	%
Truncamento (palavra-)	289	32,0
Pausa breve (.)	264	29,2

Preenchimento (ééé, ahn)	247	27,4
Prolongamento (::)	75	8,3
Pausa longa (..)	28	3,1
Total	903	100

A distribuição do score NOTA_HUMANA revela o impacto do apagamento: 26 turnos (27,4%) receberam nota 0, ou seja, as duas versões foram consideradas completamente divergentes, indicando que praticamente todas as marcas hesitativas foram removidas. Outros 14 turnos (14,7%) receberam nota 100, correspondendo a falas sem hesitação registrada. Os demais 55 turnos (57,9%) apresentaram divergência parcial, com apagamento seletivo de marcadores específicos.

Esses resultados têm implicação direta para a construção do corpus: turnos com nota 0 representam os casos de maior perda informacional na passagem da modalidade fiel para a jornalística, e são também os mais relevantes para o treinamento de modelos de detecção automática, pois concentram a maior densidade de marcadores por turno. Já os turnos com nota 100 correspondem majoritariamente a falas curtas e formulaicas (cumprimentos, confirmações, agradecimentos), nas quais a ausência de hesitação é esperada pelo contexto interacional. A faixa intermediária (notas entre 40 e 90, com 40 turnos) revela os casos de apagamento parcial e inconsistente, onde a escolha de quais marcadores preservar ou remover não segue critério explícito, evidenciando precisamente a lacuna que um protocolo de anotação formal viria preencher.

O Exemplo 1 ilustra o apagamento seletivo de marcadores na transição da FALA_FIEL para a FALA_HIGIENIZADA (ID 16, locutor jornalista, nota 65):

Exemplo 1 — Fala fiel: *boa noite Luísa (.) boa noite a todos da roda ééé noite queria aproveitar esse gancho [...] ééé você fala tutela militar né (.) e no- nas entrevistas...*

Fala higienizada: *boa noite Luísa boa noite a todos da roda noite queria aproveitar esse gancho [...] você fala tutela militar né e no- nas entrevistas...*

Nota-se que os dois preenchimentos *ééé* e a pausa breve *(.)* foram completamente removidos. O truncamento *no-* foi preservado na versão higienizada neste caso, o que ilustra a **inconsistência** no tratamento de categorias distintas dentro de uma mesma versão, evidenciando a ausência de um protocolo explícito.

O Exemplo 2 demonstra como o mesmo trecho é representado nas três modalidades (a partir de entrevista disponível em Vale et al., 2024):

Exemplo 2 — Três modalidades do mesmo trecho de fala:

[Automática] "Eu quero dizer, talvez a gente... três eixos grandes. Um: nós temos que, as forças armadas têm que apresentar o país, e o país tem que discutir qual é o plano de defesa."

[Jefferson] "Eu. Eu quero dizer. . . talvez a gente – isso – 3, 3.. eixos grandes, um (.). Nós temos que /as forças armadas têm que apresentar o país.. e o país tem que discutir.. qual é o plano de defesa?"

[GAT2] "Eu (.) eu quero dizer. . . talvez a gente – isso (.) três, três.. EIXOS grandes (.) um (.) nós temos que /as forças armadas têm que apresentar o país (.) e o país tem que disCUTIR (.) qual é o plano de DE FESA?^"

A versão automática apaga as repetições (*Eu. Eu*), as pausas e os truncamentos. A versão Jefferson registra a estrutura interacional completa. A versão GAT2 acrescenta ênfase prosódica e direção de entonação. Cada modalidade disponibiliza um recorte diferente da mesma realidade linguística, e a escolha de modalidade determina quais fenômenos existem ou não no corpus.

6. Para uma tipologia funcional

Os dados apresentados sustentam a necessidade de uma tipologia funcional das hesitações que seja: (i) **formalmente descritiva**, com critérios explícitos para cada categoria, independentes de julgamentos subjetivos; (ii) **multimodalmente compatível**, mapeável entre as três modalidades de transcrição; e (iii) **computacionalmente operacionalizável**, exportável em schema legível por máquina para uso em ferramentas de anotação e treino de modelos de linguagem.

A Tabela 1, apresentada na Seção 2, constitui uma proposta inicial nessa direção. Sua validação metodológica, via Teste de Acordo Inter-Anotadores (IAA) com Kappa de Cohen e Alpha de Krippendorff, está prevista como etapa subsequente da pesquisa, assim como sua implementação em formato ELAN/Universal Dependencies para integração ao Corpus Roda Viva.

Vale destacar também a relevância da tipologia para tarefas de PLN que vão além da anotação de corpora. A detecção automática de hesitações tem aplicação direta em sistemas de legendagem automática, que hoje produzem legendas "limpas" sem marcas de oralidade, em ferramentas de análise de entrevistas jornalísticas em larga escala, e no desenvolvimento de modelos de linguagem mais sensíveis à fala espontânea. Trabalhos recentes demonstram que LLMs aplicados a transcrições em português brasileiro cometem erros sistemáticos justamente nas categorias que as transcrições automáticas apagam: truncamentos e preenchimentos lexicais (Lima & Campelo, 2024). Uma tipologia formalmente descrita e um corpus de referência anotado são, portanto, pré-requisitos para que esses modelos possam ser treinados e avaliados de forma consistente, e não apenas testados em dados que já não contêm o fenômeno que se quer detectar.

Do ponto de vista computacional, os dados quantitativos do corpus piloto (Tabela 3) já fornecem evidências sobre quais categorias são mais frequentes e, portanto, mais relevantes para treino de modelos de detecção automática. O alto índice de truncamentos (32,0%) e preenchimentos lexicais (27,4%) sugere que essas categorias constituem prioridade para a anotação de referência a ser desenvolvida.

7. Considerações finais

Este trabalho demonstrou, a partir de um corpus piloto de 95 turnos de fala do Corpus Roda Viva, que as modalidades de transcrição jornalística, automática e conversacional tratam as hesitações de forma sistematicamente distinta — apagando, normalizando ou preservando marcadores de acordo com seus objetivos. Em 27,4% dos turnos analisados, o apagamento foi total; em 57,9%, foi parcial e inconsistente, sem critério explícito.

A proposta de tipologia funcional em cinco categorias (pausa, preenchimento lexical, truncamento, prolongamento e repetição) oferece um ponto de partida para um protocolo de anotação explícito, replicável e compatível com múltiplas modalidades de transcrição. A consolidação dessa tipologia, incluindo validação por IAA e implementação em formato legível por máquina, está prevista como contribuição metodológica central da pesquisa em andamento, cujos produtos serão integrados ao repositório público do Corpus Roda Viva.

Os dados e exemplos apresentados reforçam que a escolha da modalidade de transcrição não é neutra: ela determina quais fenômenos linguísticos existem no corpus. Tornar essa escolha explícita e os critérios de anotação públicos é condição necessária para que corpora orais sejam utilizáveis de forma consistente em pesquisa linguística e em PLN.

Cabe destacar que o corpus piloto quantitativo é derivado de uma única entrevista, o que limita a generalização dos resultados — ainda que a comparação qualitativa entre modalidades (Exemplo 2) recorra a uma segunda entrevista do corpus. A ampliação da amostra, com seleção intencional por período, área temática e perfil de entrevistados, é etapa prevista da pesquisa.

Referências

- Corley, M. and Stewart, O. W. (2008) "Hesitation disfluencies in spontaneous speech: the meaning of um." *Language and Linguistics Compass*, v. 2, n. 4, p. 589–602.
- Fox Tree, J. E. (1995) "The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech." *Journal of Memory and Language*, v. 34, n. 6, p. 709–738.
- Jefferson, G. (2004) "Glossary of transcript symbols with an introduction." In: Lerner, G. H. (Ed.). *Conversation Analysis: Studies from the first generation*. Amsterdam: John Benjamins. p. 13–31.

- Lima, P. L. S. de and Campelo, C. E. C. (2024) "Disfluency Detection and Removal in Speech Transcriptions via Large Language Models." In: Anais do XIV STIL. Porto Alegre: SBC. Disponível em: <https://aclanthology.org/2024.stil-1.16>.
- Marcuschi, L. A. (2003) *Análise da conversação*. 5. ed. São Paulo: Ática.
- Selting, M. et al. (2009) "Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)." *Gesprächsforschung*, v. 10, p. 353–402.
- Shriberg, E. (2001) "To 'errrr' is human: ecology and acoustics of speech disfluencies." *Journal of the International Phonetic Association*, v. 31, n. 1, p. 153–169.
- Souza, K. D. F. de, Pereira, D. E., Campelo, C. E. C. and Vasconcelos, L. L. (2025) "DEBISS: a Corpus of Individual, Semi-structured and Spoken Debates." arXiv, 2603.05459. Disponível em: <https://arxiv.org/abs/2603.05459>.
- Vale, O. A. et al. (2024) "Roda Viva Corpus: An Audiovisual Brazilian Portuguese Interview Corpus." In: *Proceedings of PROPOR 2024*. p. 204–210. Disponível em: <https://aclanthology.org/2024.propor-2.22>.
- Wittenburg, P. et al. (2006) "ELAN: a professional framework for multimodality research." In: *Proceedings of LREC 2006*. Genova: ELRA. p. 1556–1559.
- Zayats, V. and Ostendorf, M. (2019) "Giving attention to the unexpected: using prosody innovations in disfluency detection." In: *Proceedings of NAACL 2019*. Minneapolis: ACL. p. 2162–2172.