

O c-oral-esq: corpus oral de esquizofrênicos

*The c-oral-esq:
a brazilian corpus of schizophrenics*

Bruno Rocha¹ 

Tommaso Raso² 

Átila Augusto Soares Vital³ 

Gabriela Dias Toledo⁴ 

Ísis Beber Fiorilo Rocha⁵ 

¹ Universidade Federal de Minas Gerais, Faculdade de Letras. Belo Horizonte, MG, Brasil

² Universidade Federal de Minas Gerais, Faculdade de Letras. Belo Horizonte, MG, Brasil

³ Universidade Federal de Minas Gerais, Faculdade de Letras. Belo Horizonte, MG, Brasil

⁴ Universidade Federal de Minas Gerais, Faculdade de Letras. Belo Horizonte, MG, Brasil

⁵ Universidade Federal de Minas Gerais, Faculdade de Letras. Belo Horizonte, MG, Brasil

E-mail: bbruno791@gmail.com

E-mail: tommaso.raso@gmail.com

E-mail: 4tilavital@gmail.com

E-mail: gabidtoledo@gmail.com

E-mail: isisfiorilo@gmail.com

Editora-chefe

Marcia dos Santos
Machado Vieira

Editoras convidadas

Manuella Carnaval e
Carolina Gomes da Silva

Recebido: 02/06/2025

Aceito: 05/11/2025

Como citar: ROCHA, Bruno; RASO, Tommaso; VITAL, Átila Augusto Soares; TOLEDO, Gabriela Dias; FIORILO, Ísis. O c-oral-esq: corpus oral de esquizofrênicos. *Revista Diadorim*, v.27, n.3, e68619, 2025. DOI: 10.35520/diadorim.2025.v27n3a68619

Resumo

Este artigo apresenta o C-ORAL-ESQ, corpus oral de interações espontâneas entre pacientes diagnosticados com esquizofrenia e seus psiquiatras, coletadas em instituições públicas de Belo Horizonte (MG). O corpus documenta 43 consultas psiquiátricas reais — 19 gravadas em áudio e vídeo e 24 apenas em áudio —, totalizando 103.403 palavras, das quais 47.722 foram produzidas por pacientes. As gravações foram transcritas com segmentação prosódica e alinhadas aos sinais

acústico e visual. Descrevemos as etapas metodológicas de compilação do corpus, incluindo transcrição, anonimização, e anotação linguística e multimodal, além das características sociodemográficas e clínicas dos participantes. Por fim, sintetizamos resultados de estudos preliminares que apontam diferenças significativas entre pacientes e grupos controle em aspectos como estrutura informacional, proeminência prosódica, pausas silenciosas, expressividade facial e correlações entre medidas textuais e sintomas negativos.

Palavras-chave:

C-ORAL-ESQ; Corpus oral; Fala espontânea; Esquizofrenia.

Abstract

This article presents C-ORAL-ESQ, an oral corpus of spontaneous interactions between patients diagnosed with schizophrenia and their psychiatrists, collected in public institutions in Belo Horizonte, Brazil. The corpus documents 43 real psychiatric consultations — 19 recorded in audio and video, and 24 in audio only — totaling 103,403 words, of which 47,722 were produced by patients. The recordings were transcribed with prosodic segmentation and aligned with both the audio and visual signals. We describe the methodological steps involved in compiling the corpus, including transcription, anonymization, linguistic and multimodal annotation, as well as the sociodemographic and clinical profiles of the participants. Finally, we summarize preliminary research findings that reveal significant differences between patients and control groups in aspects such as information structure, prosodic prominence, silent pauses, facial expressiveness, and correlations between textual measures and negative symptoms.

Keywords:

C-ORAL-ESQ; Oral corpus; Spontaneous speech; Schizophrenia.

Introdução

O presente trabalho apresenta o C-ORAL-ESQ (ROCHA et al., 2025), Corpus Oral de Esquizofrênicos, já finalizado e disponível no site do projeto¹. O corpus documenta 43 interações reais entre pacientes com esquizofrenia e seus respectivos psiquiatras durante sessões regulares de atendimento oferecidas por duas instituições públicas da cidade de Belo Horizonte, o Instituto Raul Soares (IRS/FHEMIG) e o Hospital das Clínicas da UFMG (HC, UFMG/EBSERH). O projeto foi apro-

¹ Corpus Oral de Esquizofrênicos (C-ORAL-ESQ). Disponível em: <<https://www.c-oral-brasil.org/c-oral-esq.php>>.

vado pelos comitês de ética da FHEMIG e da UFMG (respectivamente, CAAE 79107817.6.0000.5149 e CAAE 79107817.6.3002.5119). Das 43 consultas que compõem o C-ORAL-ESQ, 19 possuem registros em áudio e vídeo, e as outras 24 foram gravadas somente em áudio. O corpus possui um total de 103.633 palavras, sendo 47.828 produzidas por pacientes, 48.069 por médicos, 7.090 por acompanhantes e 646 por intervenientes.

O C-ORAL-ESQ integra a família C-ORAL, conjunto de corpora de fala espontânea realizados com a mesma metodologia, a qual prevê a segmentação prosódica das transcrições e o alinhamento do sinal sonoro (e do vídeo, quando disponível) ao texto. No caso do C-ORAL-ESQ, o alinhamento foi feito no software ELAN (Witterburg et al., 2006).

O presente trabalho apresenta as principais características do corpus e os resultados de estudos preliminares conduzidos com base nele. A primeira parte traz informações sobre o local e o período das gravações, o *setting*, os equipamentos empregados para o registro de áudio e vídeo e informações relativas ao pré-processamento dos arquivos. A segunda parte explica em maiores detalhes as etapas de compilação do corpus. A terceira parte apresenta uma síntese das características dos pacientes registrados pelo C-ORAL-ESQ. A quarta parte fornece um panorama das medidas descritivas do corpus. Por fim, se apresentam de maneira sucinta os resultados dos primeiros estudos realizados com dados do C-ORAL-ESQ.

As gravações

Local e período das gravações

O C-ORAL-ESQ foi gravado em duas instituições públicas em Belo Horizonte (MG): o Instituto Raul Soares, subordinado à Fundação Hospitalar do Estado de Minas Gerais (FHEMIG), e o Hospital das Clínicas (HC), parte da Universidade Federal de Minas Gerais (UFMG) e administrado em conjunto à Empresa Brasileira de Serviços Hospitalares (EBSERH). Ambas as instituições fornecem atendimento ao Serviço Único de Saúde. Para o IRS, a referência foram os doutores João V. Salgado e Lucas Machado Mantovani, enquanto no HC a referência foi o doutor Breno Fiuza Cruz. Das 43 gravações que entraram nessa versão do corpus, 31 foram gravadas no IRS e 12 no HC.

As gravações do C-ORAL-ESQ ocorreram em duas fases distintas. A primeira vai de maio de 2018 até o início de 2020, quando as gravações foram suspensas em função da pandemia de COVID-19. Nessa etapa, foram feitas gravações em áudio de consultas no IRS. A segunda fase se inicia em março de 2022 e prossegue até maio de 2024. Nessa etapa, foram feitas gravações em áudio e vídeo nas duas instituições.

Ao total, foram realizadas 88 gravações, tendo sido selecionadas 43 para compor a primeira versão do corpus.

Setting das gravações

As gravações foram feitas sempre em consultas agendadas por um médico residente, como parte de um tratamento em curso, nos mesmos consultórios em que as consultas normalmente ocorrem. Além do médico e do paciente, algumas consultas contam com a presença de um acompanhante do paciente (na maior parte dos casos, um familiar), que pode interagir seja com o médico, seja com o paciente. Ocasionalmente, há a presença de intervenientes (normalmente funcionários do hospital que, por motivos diversos, entram no consultório com a consulta em curso e falam brevemente com o médico e/ou o paciente).

Para garantir a validade ecológica das gravações, foram realizadas reuniões prévias entre a equipe médica e a equipe de compilação do corpus, em que os médicos residentes foram explicitamente instruídos a conduzirem as consultas normalmente, sem alterarem a sua estrutura em função delas estarem sendo gravadas. Esse procedimento foi necessário pois tanto no IRS quanto no HC, é comum que residentes e pacientes participem de pesquisas científicas cujos dados são coletados antes, durante ou após uma consulta, por meio da realização de procedimentos adicionais de natureza diversa àqueles da consulta. Em face disso, foi necessário explicar aos médicos residentes que o corpus tem o objetivo de registrar as consultas da maneira mais próxima possível da que elas acontecem quando não estão sendo gravadas, para capturar a fala espontânea produzida naquele contexto. Desse modo, foi explicado que, ainda que um paciente estivesse falando pouco em uma dada consulta, os médicos não deveriam tentar fazer com que falassem mais *peelo simples fato de estarem sendo gravados*, a menos que isso fosse necessário para o bom andamento da consulta. Assim, o corpus admitia a possibilidade de incluir gravações de consultas em que, embora o paciente estivesse presente e participasse da mesma, a principal interação se desse entre o médico e o acompanhante, como efetivamente pode ocorrer em contexto natural.

O *setting* padrão das consultas (e, portanto, das gravações) prevê que paciente e médico estejam sentados um de frente para o outro, de lados opostos da mesa do consultório. O médico está diante de um computador e frequentemente usa o mesmo para fazer anotações. O acompanhante, quando presente, é colocado em uma cadeira ao lado do paciente. A equipe de gravação, composta por dois alunos de graduação e/ou pós-graduação do projeto, entra no consultório após o médico já ter conversado brevemente com o paciente sobre a gravação e ter pedido uma autorização prévia a ele e ao eventual acompanhante. A equipe então se apresenta, coleta o Termo de Consentimento Livre e Esclarecido de todos os participantes e coloca um microfone de lapela no paciente e outro no médico, ambos conectados a um

gravador digital, posicionado fora do consultório. O acompanhante permanece sem microfone, mas sua voz é captada indiretamente pelos microfones do médico e do paciente. Nas consultas com gravação de vídeo, são colocadas uma ou duas câmeras em cima da mesa direcionadas ao paciente, sendo uma frontal e uma levemente lateral. A câmera frontal destina-se a capturar sobretudo as expressões faciais do paciente, mas também seus gestos. A câmera lateral tem a função primária de capturar os gestos do paciente nos casos em que ele gesticule em posições ou de formas que a câmera frontal não consiga capturar adequadamente. Na impossibilidade de serem usadas duas câmeras, a única câmera disponível é colocada na posição frontal. Após posicionar o equipamento, a equipe sai do consultório e começa a gravar. Assim, todas as gravações foram conduzidas invariavelmente com a equipe de gravação fora do consultório.

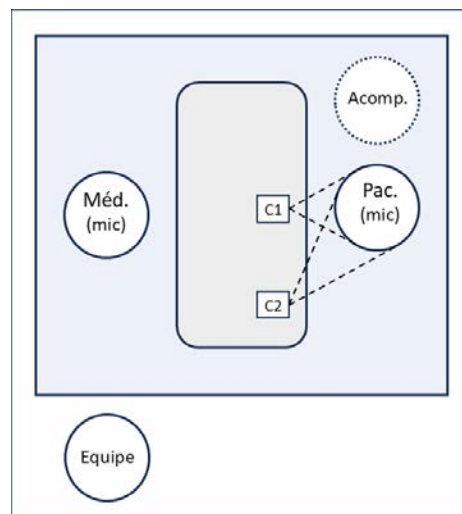


Figura 1. Ilustração do setting de gravação das consultas, indicando a posição do médico (Méd.), paciente (Pac.), acompanhante (Acomp.) e equipe de gravação (Equipe), bem como dos microfones de lapela (mic) e das câmeras de vídeo (C1 e C2). O quadrado azul indica a área do consultório, e o retângulo cinza ilustra a posição da mesa do médico

Fonte: Elaborada pelos próprios autores

No HC, em particular, percebeu-se após fazer algumas gravações que era necessário colocar uma folha de papel pardo bloqueando a janela dos consultórios em que as gravações de vídeo seriam feitas. Os consultórios dessa instituição possuem janelas sem cortinas ou estruturas que bloqueiam a luz. Como as consultas eram gravadas no período da tarde, muitos dos primeiros registros foram descartados porque a incidência de forte luz solar criava um contraste tão forte com o rosto do paciente a ponto de inviabilizar o exame visual e/ou automático de suas expressões faciais. Assim, a partir de quando passou-se a adotar essa solução, ao chegar no HC, a equipe colocava o papel pardo antes do paciente chegar na sala em que a consulta estava marcada.

A maior parte das consultas foi gravada integralmente, mas há casos em que por motivos diversos (por exemplo, por falha dos equipamentos ou porque a consulta

começou abruptamente) a gravação se inicia logo após o início da sessão. Essas informações são registradas por uma tabela disponível no site do projeto. As consultas têm duração muito variável, indo de 00:05:24 a 00:37:31, com média de 00:16:42 (cf. Anexo II).

Equipamento de gravação e pré-processamento dos arquivos de áudio e vídeo

Para a gravação do áudio, foram usados dois conjuntos de microfones de lapela sem fio (um para o médico e o outro para o paciente) conectados a um gravador digital de alta resolução. Para a gravação do vídeo, foram usadas duas câmeras direcionadas ao paciente para capturar sua imagem. Os equipamentos são:

1. Microfones: As gravações foram feitas com conjuntos de microfones de lapela omnidirecionais Sennheiser ME 2 conectados a *transmitters* e *receivers* Sennheiser EK100 e SK100.
2. Gravadores: A maior parte das gravações de áudio foi feita com gravadores Tascam DR-100 MK II. Esporadicamente, foram usados gravadores Tascam DR-100 MK III e Marantz PMD 660. Todas as gravações foram feitas em dois canais (estéreo) com taxa de amostragem de 44.100 Hz e salvas em arquivo *.wav*.
3. Câmeras: Todas as gravações de vídeo foram feitas com uma ou duas câmeras GoPro Hero7 White. Os vídeos foram gravados na resolução 1920x1440 px, com 30 fps e áudio estéreo com taxa de amostragem de 48.000 Hz.

Os arquivos de áudio, originalmente gravados em 44.100 Hz, foram reamostrados para a taxa de 22.050 Hz, para que se tornassem mais leves, e mantidos no formato estéreo. Ainda assim, todos os arquivos originais são preservados pela equipe de compilação do corpus. Há algumas gravações em que o canal do médico se corrompeu ou possuía qualidade acústica muito inferior com relação ao do paciente e foi removido, preservando somente o canal do paciente, transformando a gravação em mono. As informações sensíveis (informações textuais que permitem identificar a identidade do paciente, como seu nome) foram bipadas no *software* Audacity² com uma onda sinusoidal de 300 Hz com amplitude de 0,3, de duração variável.

Os vídeos foram processados com o codec H.264 para diminuição de tamanho, mantendo a resolução 1920x1440 px e a taxa de 30 fps. Os parâmetros para o processamento dos arquivos foram Progressive, 203 (75% HLG, 58% PQ), VBR, 1 pass, Target 10,00 Mbps; AAC, 320 kbps, 48.000 Hz, stereo. A anoni-

² Audacity Team (2014). Audacity(R): Free Audio Editor and Recorder [Computer program]. Version 2.0.0. Disponível em: <<http://audacity.sourceforge.net/>>. Acesso em: 3 de maio de 2025.

mização dos vídeos frontais foi realizada pelo FaceFusion³, que substitui automaticamente os rostos por meio do processador *face_swapper*. Já nos vídeos laterais, foi utilizado o DeFace⁴, que aplica um efeito de desfoque nos rostos detectados. Todos os vídeos passaram por uma etapa de verificação manual para assegurar a eficácia da anonimização. Nos trechos em que a anonimização automática se mostrou insuficiente, foram realizadas edições manuais utilizando o software Adobe Premiere Pro 2022. Nesses casos, aplicou-se o efeito Gaussian Blur, com intensidade de desfoque (blurriness) variando entre 100 e 150, para ocultar adequadamente o rosto do paciente sempre que este estivesse visível devido a falhas na anonimização automática.

O arquivo de áudio usado para o alinhamento texto-som-vídeo no Elan (Witterburg et al., 2006) foi sempre aquele captado pelo gravador, via de regra de qualidade superior ao áudio da câmera. Nas consultas com vídeo, o áudio das câmeras foi usado exclusivamente em momentos em que o áudio do gravador não estava disponível por algum problema técnico (por exemplo, casos em que o gravador desligou temporariamente durante a consulta, mas as câmeras continuaram gravando). Assim, em algumas gravações, há oscilação da qualidade do áudio do alinhamento.

Além disso, considerando que as gravações foram feitas em contexto natural (consultas regulares ocorridas nos mesmos consultórios em que elas ocorreriam caso não estivessem sendo gravadas, em instituições que prestam atendimento simultâneo a inúmeras pessoas), é comum que se ouçam ruídos de fundo diversos (pessoas conversando ao fundo, cadeiras arrastando, portas abrindo, etc.), os quais todavia não comprometem a qualidade da gravação como um todo. Diante disso, foram selecionadas para compor o corpus as gravações cuja qualidade acústica é igual ou superior à faixa B segundo a classificação do C-ORAL-BRASIL (Raso, 2012): boa ou média resposta dos microfones; muitas partes da gravação possuem qualidade apropriada para análise fonética; computação da f_0 possível em grande parte do arquivo; não há excesso de sobreposições e de ruído de fundo.

Etapas de compilação do corpus

Todas as gravações do C-ORAL-ESQ foram submetidas ao mesmo protocolo, composto das seguintes etapas, as quais serão explicadas a seguir:

³ Henry Ruhs. FaceFusion (versão 3.0.0). GitHub, 2024. Disponível em: <<https://github.com/facefusion/facefusion>>. Acesso em: 3 de maio de 2025.

⁴ Martin Drawitsch. DeFace (versão 1.5.0). GitHub, 2023. Disponível em: <<https://github.com/ORG-HD/deface>>. Acesso em: 3 de maio de 2025.

1. Elaboração de um arquivo de metadados da gravação e dos participantes;
2. Transcrição e segmentação prosódica do áudio;
3. Primeira revisão da transcrição e segmentação;
4. Alinhamento do áudio ao texto no *software* Elan (Witterburg et al., 2006);
5. Segunda revisão da transcrição e segmentação;
6. Terceira revisão da transcrição e segmentação, com foco na revisão da aplicação dos critérios não ortográficos;
7. Anonimização dos arquivos de áudio;
8. Edição dos arquivos de vídeo;
9. Anonimização dos arquivos de vídeo;
10. Alinhamento do vídeo ao áudio e ao texto no *software* Elan;
11. Quarta revisão da transcrição e segmentação, com foco na revisão da segmentação prosódica;
12. Anotação léxico-morfossintática com o *software* PALAVRAS (Bick, 2012);
13. Anotação de expressões faciais com o *software* OpenFace (Baltrušaitis et al., 2016).

Para todas as gravações, é produzido um arquivo de texto, em formato.txt, com metadados da sessão de gravação e de seus participantes. O Anexo I exibe os metadados de uma das gravações do C-ORAL-ESQ.

A transcrição e a segmentação prosódica seguem os mesmos critérios elaborados para a família C-ORAL e em particular aqueles elaborados para o corpus C-ORAL-BRASIL (Mello et al., 2012), cuja metodologia é discutida por Mello (2014). A segmentação prevê a anotação de fronteiras prosódicas terminais e não terminais (Barbosa; Raso, 2018; Izre'el et al., 2020), enunciados interrompidos e fenômenos de *retracting* (reformulações parciais ou totais do conteúdo locutivo de uma unidade informacional), de acordo com o formato CHILDES (Mcwhinney, 2000) adaptado para a segmentação prosódica por Moneglia e Cresti (1997) e adotado pelo C-ORAL-ROM (Cresti; Moneglia, 2005). A transcrição é semiortográfica e registra fenômenos da fala em vias de gramaticalização ou lexicalização, como perda da morfologia verbal, cliticização dos pronomes sujeito e aférese (Mello et al., 2012). As formas ortográficas são transcritas de acordo com o Vocabulário Ortográfico da Língua Portuguesa (VOLP)⁵, em sua versão online vigente no período de compilação do corpus.

As tarefas de transcrição e segmentação foram realizadas concomitantemente e de maneira manual por alunos de graduação e pós-graduação que receberam treinamento especializado por pelo menos um semestre. Antes do início dessa

⁵ VOLP - Vocabulário Ortográfico da Língua Portuguesa. Academia Brasileira de Letras. Acesso em: 05 de maio de 2025. Disponível em: <<https://www.academia.org.br/nossa-lingua/busca-no-vocabulario>>. Acesso em: 3 de maio de 2025.

etapa, os alunos envolvidos nela foram submetidos ao teste Kappa (Fleiss, 1971) para medir o grau de concordância na anotação de fronteiras terminais e não terminais. As transcrições e segmentações foram realizadas apenas por alunos que alcançaram um Kappa de 0.8 com um transcritor expert. Posteriormente, todos os textos foram submetidos a três revisões pelos anotadores que obtiveram o Kappa mais alto entre si. Durante o período em que foi feita a terceira revisão, cujo foco foi a aplicação dos critérios não ortográficos, a equipe passou por oficinas semanais de *follow-up* sobre os critérios de transcrição do C-ORAL-BRASIL. Por fim, todas as transcrições passaram por uma quarta revisão, com foco na marcação de quebras terminais e não terminais, que foram realizadas pelo mesmo pesquisador.

O alinhamento do áudio ao texto foi feito pela mesma equipe no *software* ELAN (Witterburg et al., 2006), após a primeira etapa de revisão. Como explicado anteriormente, o áudio usado no alinhamento era sempre o áudio captado pelo gravador, que possui qualidade superior ao da câmera. O áudio dos arquivos de vídeo foi usado somente em trechos em que o áudio do gravador foi corrompido ou não estava disponível. Assim, as demais revisões eram realizadas já com o texto alinhado. O acréscimo do vídeo ao alinhamento só foi feito após a etapa de edição dos mesmos.

O corpus inteiro é anotado léxico-morfossintaticamente através do *software* Palavras (Bick, 2000) adaptado para os critérios do C-ORAL-BRASIL (Bick, 2012), que classifica as PoS e as estruturas sintáticas na perspectiva da Gramática de Dependência (Bick, 2023). Essa anotação é disponibilizada em duas versões, ambas com anotação de PoS e semântica, em arquivos de texto nos formatos.dep (com anotação de dependência sintática) e.syn (sem anotação de dependência).

Além disso, os vídeos possuem a anotação automática das *action units* (Ekman; Friesen, 1978) que formam as expressões faciais, feitas no *software* OpenFace (Baltrušaitis et al., 2016). Essa anotação foi feita somente para os *frames* que possuíam o grau mínimo de confiabilidade de 85%. Adicionalmente, cerca de três minutos de cada vídeo será futuramente submetido à anotação manual de *co-speech gestures*, com base em Bressem (2013) e Kendon (2004).

Por fim, além dos procedimentos explicados anteriormente, um subconjunto de gravações do corpus será anotado informacionalmente com base na Language into AcT Theory (Cresti, 2000; Raso, 2012b; Moneglia; Raso, 2014).

A Figura 2, a seguir, exibe a captura de tela do alinhamento da gravação *besqmm39* no Elan.

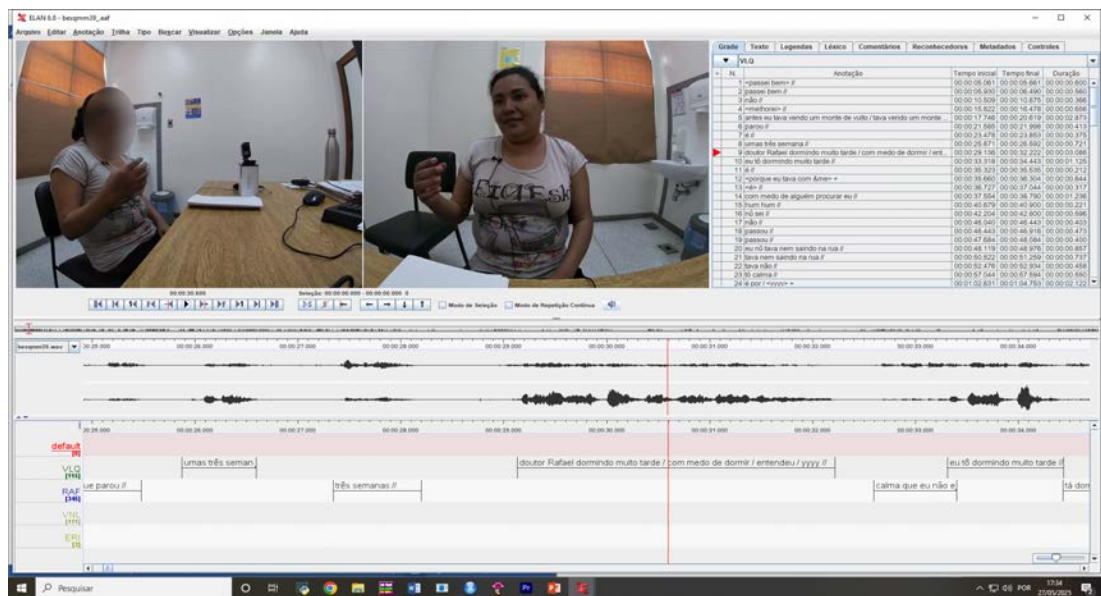


Figura 2. Captura de tela do alinhamento no software Elan (WITTERBURG et al., 2006) da gravação besqmm39. Na parte superior à esquerda, observam-se as imagens de vídeo e na parte inferior da imagem, estão o oscilograma da onda sonora e os tiers de transcrição da fala dos participantes. No vídeo da esquerda, o rosto da paciente foi anonimizado com o FaceFusion, *software* que cria um rosto artificial mantendo as expressões faciais do original. No vídeo da direita, foi usado o *software* DeFace para a anonimização aplicar um efeito de desfoque

Fonte: Elaborada pelos próprios autores

Os pacientes

O C-ORAL-ESQ registra interações de 43 indivíduos diagnosticados com esquizofrenia segundo os critérios estabelecidos pelo DSM-V e submetidos à confirmação diagnóstica de acordo com o SCID-5 (FIRST). São 27 indivíduos do sexo masculino e 16 do sexo feminino, com idade média de 43,4 anos (DP 12,8) e escolaridade média de 9,3 anos (DP 3,4) – Figuras 3 e 4. Todos estão estabilizados e em tratamento em curso no IRS e HC. Dos 43 pacientes, 1 faz tratamento medicamentoso com antipsicóticos típicos, 29 com antipsicóticos atípicos e 13 com medicamentos de ambas as classes. Essas características refletem o público atendido pelas duas instituições participantes do projeto, que prestam atendimento pelo SUS (Sistema Único de Saúde).

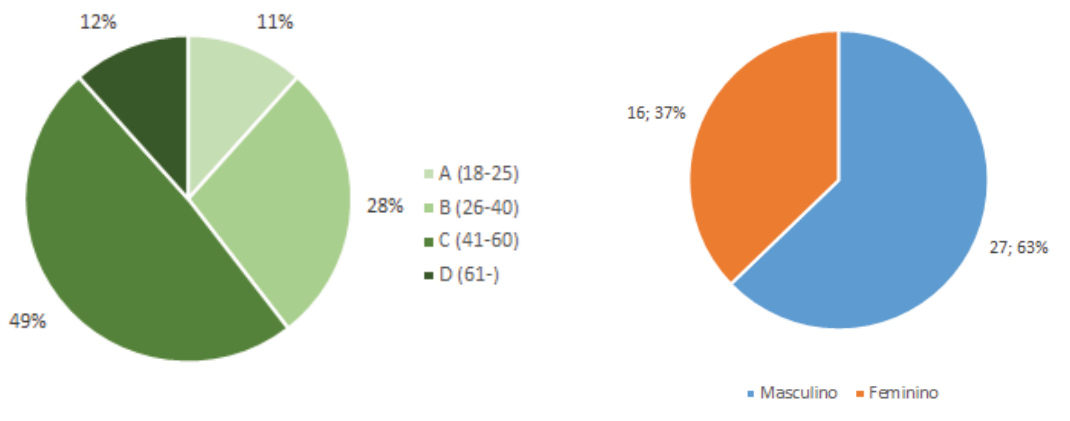


Figura 3. Sexo (esquerda) e faixa etária dos pacientes (direita)

Fonte: Elaborada pelos próprios autores

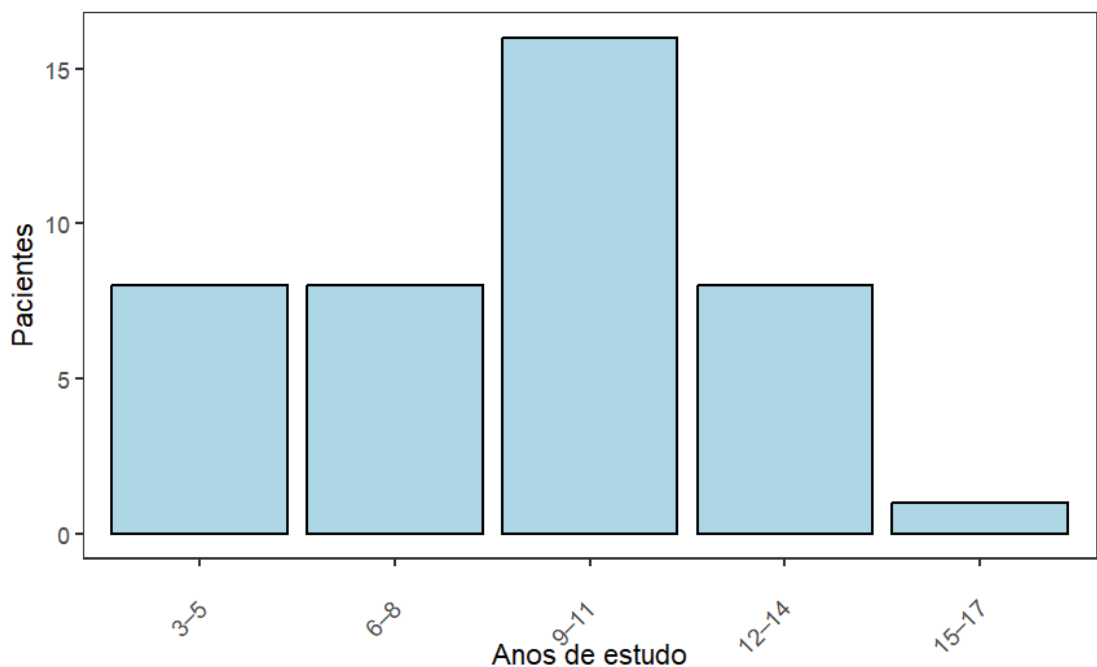


Figura 4. Escolaridade dos pacientes, em faixas de 3 em 3 anos de estudo

Fonte: Elaborada pelos próprios autores

Para todos os pacientes é disponibilizada uma grande quantidade de informações sociodemográficas e clínicas, coletadas previamente pela equipe médica como parte do tratamento oferecido pelas instituições. Essas informações estão sintetizadas nas tabelas a seguir: a Tabela 1 apresenta a média, mediana e desvio-padrão das principais informações sobre os pacientes. Os valores de cada paciente para idade, sexo, anos de estudo, duração da doença, classe de antipsicótico, equivalente de clorpromazina, PANSS_P e PANSS_N e outras informações encontram-se no site do C-ORAL-ESQ.

Tabela 1. Média, mediana e desvio-padrão da idade, anos de estudo, duração da doença, equivalente de clorpromazina, PANSS_P e PANSS_N dos pacientes

Variável	Média	Mediana	Desvio-padrão
Idade (anos)	43,4	44,0	12,8
Anos de estudo	9,3	11,0	3,4
Duração da doença (anos)	16,4	12,0	13,8
Equivalente de clorpromazina (mg)	648,9	600,6	372,4
PANSS_P (positiva)	11,3	11,0	3,6
PANSS_N (negativa)	18,8	18,0	6,4

Fonte: Elaborada pelos próprios autores

Além disso, o corpus disponibiliza os dados relativos aos seguintes testes e escalas:

1. PANNS (Positive and Negative Syndrome Scale; Kay et al., 1987): avalia a sintomatologia positiva, negativa e geral da doença;
2. Simpson-Angus (Simpson; Angus, 1970): avalia os EPS (Extra Piramidal Symptoms), efeitos colaterais dos remédios relativos ao parkinsonismo;
3. BACS (The Brief Assessment of Cognition in Schizophrenia; Keefe, 2004): mede os prejuízos cognitivos que têm correlação mais forte com a esquizofrenia;
4. UPSA (UCSD Performance-Based Skill Assessment; Patterson et al., 2001): mede a funcionalidade dos pacientes em seis domínios da vida cotidiana (planejamento doméstico, comunicação, finanças, transporte, lazer e planejamento, gerenciamento de medicação);
5. HINTING Task (Corcoran et al., 1995): mede a capacidade de inferência das intenções dos outros (Teoria da Mente);
6. BLERT (Bell Lysaker Emotion Recognition Task; Bell; Lysacker, 2000): mede a habilidade de identificar os sete estados emocionais básicos (alegria, tristeza, raiva, medo, surpresa, desprezo, neutro).

Medidas gerais do corpus

O C-ORAL-ESQ é composto por 43 registros de consultas psiquiátricas, sendo 24 consultas gravadas em áudio e 19 em áudio e vídeo. Cada gravação documenta a interação entre um médico residente e um paciente, sem a repetição de pacientes em gravações diferentes. Alguns médicos, por outro lado, se repetem ocasionalmente. Ao total, as consultas foram realizadas por 24 médicos diferentes, dos quais 16

atuaram em uma única consulta e os outros 8 conduziram mais de uma consulta. A Tabela 2 mostra o número de consultas de cada médico. Nela, lê-se que 14 dos 24 médicos realizaram somente 1 consulta. Por outro lado, há 1 médico que está presente em 6 das 43 consultas do corpus.

Tabela 2. Número de consultas gravadas pelos médicos

Médicos	Consultas
14 médicos	1 consulta
2 médicos	2 consultas
2 médicos	3 consultas
2 médicos	4 consultas
1 médico	5 consultas
1 médico	6 consultas

Fonte: Elaborada pelos próprios autores

O corpus é composto por 103.633 palavras, para uma duração total de 11:58:13. Cada gravação tem em média 2.410 palavras (DP 1.085), com 00:16:42 de duração (DP 00:07:02). A gravação mais curta possui 801 palavras e 00:05:24, e a mais longa, 4.485 palavras e 00:37:31. Essas informações podem ser observadas na Tabela 3. A distribuição de duração das gravações em faixas de 5 minutos e da extensão das gravações em faixas de 500 palavras podem ser vistas na Figura 5.

Tabela 3. Medidas relativas ao número de palavras e duração das gravações

Parâmetro	Total	Média	Mediana	DP	Mínimo	Máximo
<i>Palavras</i>	103.633	2.410	2.379	1.085	801	4.485
<i>Duração</i>	11:58:13	00:16:42	00:16:17	00:07:02	00:05:24	00:37:31

Fonte: Elaborada pelos próprios autores

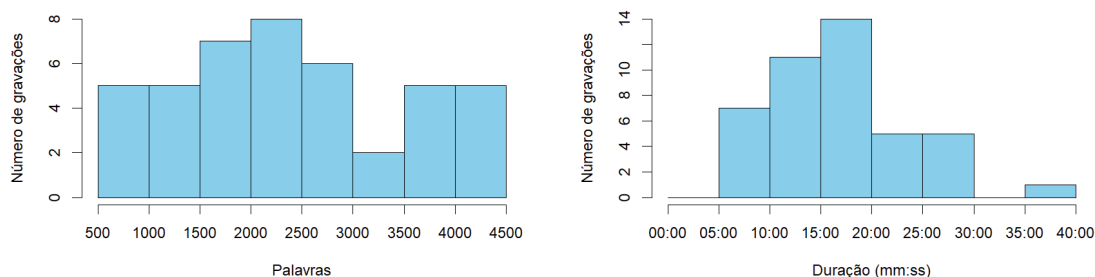


Figura 5. Histogramas de duração das consultas (esquerda), em subdivisões de 5 minutos e de número de palavras (direita), em subdivisões de 500 palavras

Fonte: Elaborada pelos próprios autores

A média de palavras de pacientes e de médicos por gravação é semelhante (1.112 e 1.118, respectivamente), com um desvio padrão superior para as palavras de pacientes (781) do que para as de médicos (546), como mostra a Tabela 4.

Tabela 4. Medidas de palavras por cada tipo de participante das consultas (paciente, médico, acompanhante e interveniente)

Participante	Palavras	Percentual	Média	Mediana	DP	Mínimo	Máximo
<i>Paciente</i>	47.828	46,2%	1.112	868	781	190	2.608
<i>Médico</i>	48.069	46,4%	1.118	995	546	264	2.322
<i>Acompanhante</i>	7.090	6,8%	165	0	273	0	978
<i>Interveniente</i>	646	0,6%	15	0	36	0	168

Fonte: Elaborada pelos próprios autores

Os dados de cada consulta podem ser observados no Anexo II. O código da gravação, na primeira coluna na tabela, é formado por uma sequência alfanumérica que indica o idioma da consulta (“b” para Português Brasileiro), a tipologia (“esq” para consultas médicas de pacientes com esquizofrenia), a modalidade de gravação (“au” para gravações em áudio, “mm” para gravações multimodais, em áudio e vídeo), o número da gravação dentro do corpus (indo de “01” a “43”).

A grande variação no número de palavras de pacientes e médicos, observável pelo seu alto desvio-padrão (781 e 546, respectivamente), é consequência direta do contexto de fala em que as gravações foram feitas e depõe a favor da validade ecológica dos dados: as consultas psiquiátricas registradas pelo corpus não possuem uma estrutura fixa, ainda que alguns temas sejam recorrentemente objeto de discussão (rotina do paciente, sintomas psiquiátricos, remédios, relação com a família, dentre outros). A isso, soma-se o fato de que a disposição ao diálogo varia expressivamente de paciente para paciente, o que é inclusive consequência da própria sintomatolo-

gia da doença. Há também de se considerar que muitas vezes os pacientes vão às consultas junto de acompanhantes, os quais podem falar por iniciativa própria ou respondendo a perguntas dos demais participantes. Enquanto alguns acompanhantes permanecem em silêncio na maior parte do tempo, outros engajam em longos diálogos com o médico, o que varia não somente em função do quão propenso o paciente está para o diálogo, mas também em face do grau de contato entre paciente e acompanhante. Esses são somente alguns dos fatores que contribuem para a variação do número de palavras dos participantes e da consulta como um todo.

Os gráficos e tabelas a seguir permitem entender melhor essa variação. A Figura 6 apresenta o número absoluto de palavras do paciente em cada consulta, em ordem decrescente. Na primeira coluna da esquerda, está uma consulta em que o paciente produziu 2.608 palavras, enquanto na última coluna da direita está uma em que o paciente produziu 190.

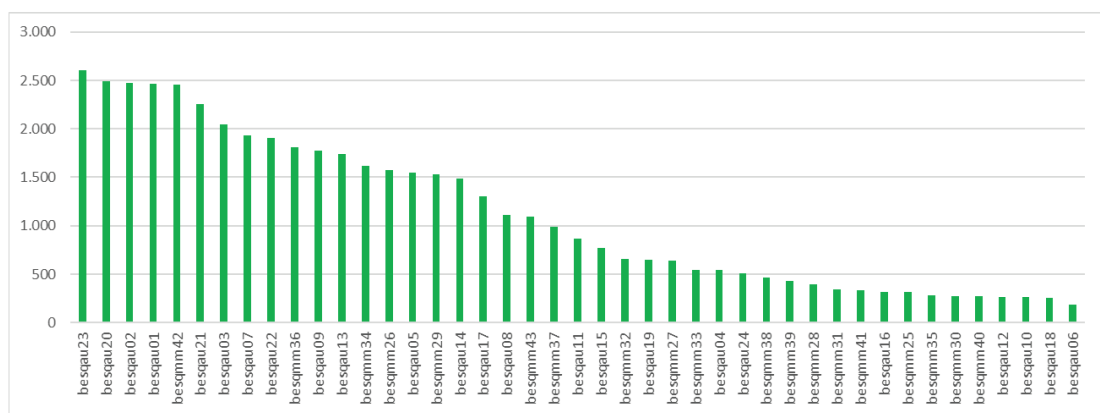


Figura 6. Número de palavras do paciente nas consultas, em ordem decrescente a partir da consulta em que o paciente produz o maior número de palavras

Fonte: Elaborada pelos próprios autores

A Figura 7, por sua vez, mostra a proporção entre o número de palavras de pacientes e o número de palavras de médicos em cada consulta, ordenada de maneira decrescente no eixo x (portanto, uma ordem diferente daquela dos gráficos anterior e seguinte). Na extremidade esquerda, está uma consulta em que 86% das palavras foram produzidas pelo paciente. No lado oposto, há uma consulta em que o paciente produziu 16% das palavras da consulta. Assim, há 16 gravações em que o paciente fala mais que o médico (aquelas em que o paciente tem mais de 50% das palavras) e 27 em que o paciente fala menos com relação ao médico. Vale notar que não há uma correlação sistemática entre o número de palavras produzidas pelo paciente (Figura 6) e a proporção das palavras do paciente com relação às do médico: as setas na Figura 7 indicam as gravações em que o paciente produz a maior quantidade de palavras (besqau23), a quantidade mínima (besqau06) e a mediana de palavras (besqau11). Essa marcação permite inferir que o número de palavras do paciente em uma gravação, retratado pela Figura 6, também é função da duração da consulta, não somente da disposição do paciente para falar.

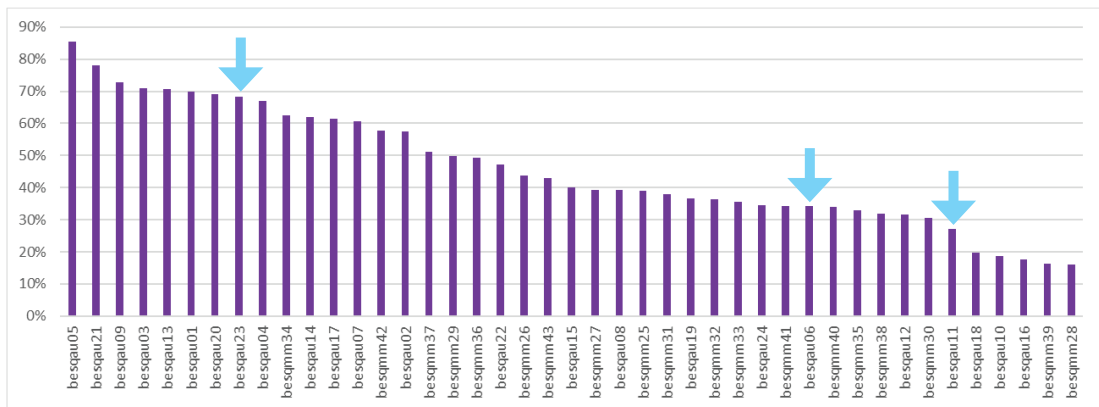


Figura 7. Proporção entre as palavras produzidas por pacientes e palavras produzidas por médicos em cada consulta, em ordem decrescente a partir da consulta em que o paciente produz proporcionalmente a maior quantidade de palavras
Fonte: Elaborada pelos próprios autores

A figura anterior baseia-se somente na quantidade de palavras produzidas por médicos e por pacientes, desconsiderando a contribuição dos demais participantes. Para entender a contribuição dos acompanhantes nas diferentes consultas, responsáveis por 6,8% do total de palavras do corpus, é preciso considerar primeiramente que em 20 das 43 gravações do corpus (46,5% delas) há a presença de um acompanhante que produz pelo menos 1 palavra. Nas demais 23 gravações, o paciente está sozinho ou seu acompanhante permanece em silêncio durante toda a consulta. Isso pode ser observado na Tabela 5.

Tabela 5. Número de gravações com e sem a presença de acompanhantes que produzem ao menos 1 palavra

Tipo de gravação	Quantidade	Percentual
<i>Sem acompanhante</i>	23	53,5%
<i>Com acompanhante</i>	20	46,5%
<i>Total</i>	43	100%

Fonte: Elaborada pelos próprios autores

A Figura 8, a seguir, mostra o número absoluto de palavras de pacientes, médicos e acompanhantes por consulta, em ordem decrescente segundo as palavras de pacientes. As setas em azul destacam as gravações em que o acompanhante produz um número igual ou superior de palavras com relação ao paciente (besqau11, besqmm33, besqmm39, besqau16, besqmm30, besqau18 e besqau06), que correspondem a 6 das 20 gravações com acompanhantes. Aqui, é importante notar que todas elas estão na parte à direita do gráfico, região que exibe as gravações em que pacientes produziram o menor número de palavras. As outras 14 gravações com acompa-

nhantes estão dispersas em todas as regiões do gráfico (algumas à esquerda, como besqau20 e besqau01; outras ao centro, como besqmm37 e besqau19; outras à direita, como besqmm41 e besqau12). Em suma, o gráfico mostra que as consultas em que os acompanhantes têm o papel mais relevante são justamente aquelas em que os pacientes produzem menos palavras em valores absolutos.

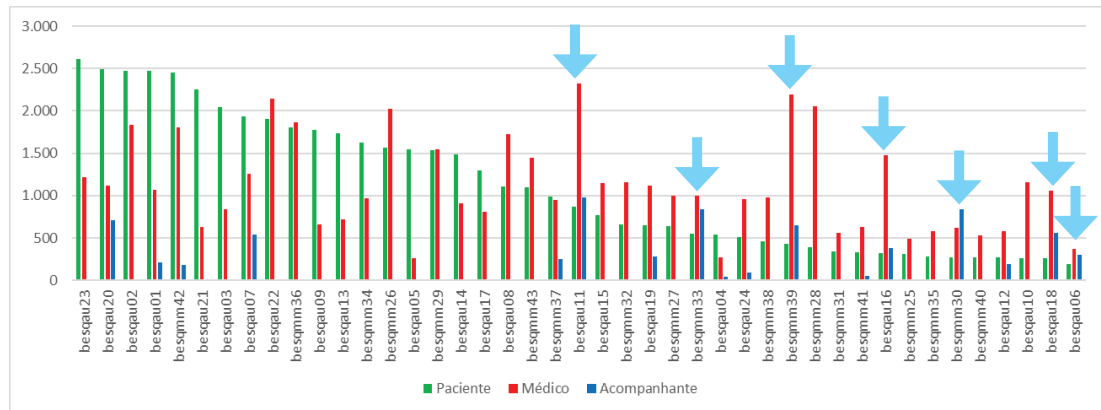


Figura 8. Número de palavras de pacientes (verde), médicos (vermelho) e acompanhantes (azul) por gravação, em ordem decrescente com relação ao número de palavras de pacientes, com destaque (setas azuis) nas gravações em que o acompanhante produz número igual ou superior de palavras que o paciente

Fonte: Elaborada pelos próprios autores

Ainda há de se perguntar se, nos casos em que o mesmo médico foi gravado várias vezes, existe uma correlação forte entre o seu número de palavras e aquele do paciente. A Figura 9 permite avaliar essa questão. Cada gráfico presente nela diz respeito às diferentes gravações realizadas por um mesmo médico (identificado na parte superior dos gráficos pela sigla de três letras maiúsculas e pelo código numérico). Assim, o gráfico presente na parte superior à esquerda da figura exhibe os dados das três gravações em que o médico M20 aparece. Na primeira, M20 produz pouco menos de 1.800 palavras, na segunda, pouco menos de 1.000, na terceira, cerca de 500 e, na quarta, cerca de 600. Comparando esses gráficos, é possível notar uma correlação direta entre o número de palavras dos médicos e o número de palavras do paciente nas consultas conduzidas por alguns profissionais, mas não por todos: nas consultas de M20 e M14, o médico produz mais palavras que o paciente, ao passo que nas consultas de M04 e M10, observa-se a tendência inversa. Para M16 e M19, o número de palavras parece não variar expressivamente em função do número de palavras produzidas pelo paciente. A variação do comportamento dos médicos entre si e de um médico em suas diferentes consultas mostra que não existe uma única condicionante para explicar o número de palavras do médico, sugerindo uma adaptação do comportamento do profissional em face das exigências de suas consultas, além de tendências pessoais. Assim, esse conjunto de dados sugere, mais uma vez, que as gravações do corpus conseguem capturar adequadamente a realidade das consultas, que é por si só bastante heterogênea.



Figura 9. Número de palavras de pacientes (verde), médicos (vermelho) e acompanhantes (azul) agrupadas segundo o médico presente na consulta. A sigla alfanumérica de três dígitos na parte superior de cada gráfico indica o médico das consultas

Fonte: Elaborada pelos próprios autores

Por fim, para todos os textos, foram extraídas as seguintes medidas linguísticas, que são disponibilizadas com o corpus:

- Medidas das gravações:
 - Duração
 - Número de palavras de cada participante (paciente, médico e acompanhante)
 - Número de palavras por turno
 - Número de palavras por unidade terminada
 - Número de palavras por unidade prosódica
 - Número de unidades terminadas com uma unidade prosódica
 - Número de unidades terminadas com mais de uma unidade prosódica
 - Número de unidades prosódicas por unidades terminadas
 - Número de unidades não terminadas
 - Número de palavras retratadas
 - Número de palavras interrompidas
 - Número de tomadas de tempo
 - Frequência, intensidade de duração das *action units* das expressões faciais de cada paciente
- Medidas gerais:
 - Lista de frequência de palavras (*types e tokens*) geral e por paciente
 - Palavras chave

Síntese de resultados de pesquisas preliminares

Até o momento foram realizados alguns estudos piloto em parte das gravações do corpus. Os primeiros estudos, realizados anteriormente ao início da compilação do C-ORAL-MED, corpus de controle do C-ORAL-ESQ, foram feitos usando como controle alguns textos do C-ORAL-BRASIL (Raso; Mello, 2012), corpus de referência do Português Brasileiro, que registra a fala espontânea em uma ampla variedade de contextos comunicativos.

Os estudos mais atuais usam como grupo controle os pacientes do C-ORAL-MED, corpus em fase inicial de compilação projetado para ser o controle do C-ORAL-ESQ, que está sendo gravado no ambulatório da Igreja do Carmo (Belo Horizonte-MG), subordinado ao Hospital das Clínicas da UFMG (HC-UFMG/EBSERH). Nesse ambulatório, alunos de graduação do curso de Medicina da UFMG (do VI e VIII períodos) prestam atendimento à população, com a supervisão de Médicos da Faculdade de Medicina da UFMG, como parte das atividades de estágio clínico. Para o C-ORAL-MED, estão sendo gravadas consultas de atendimento a pacientes com doenças crônicas (cardiopatias, diabetes, etc.) que não possuem em idade, gênero e escolaridade ao C-ORAL-ESQ. Suas gravações se iniciaram no segundo semestre de 2024 e, mesmo que ainda não tenham sido submetidas a todo o tratamento metodológico das gravações do C-ORAL-ESQ (detalhado na seção *Etapas de compilação do corpus*), já podem ser usadas para alguns estudos que precisam apenas dos dados brutos.

Estrutura informacional

Costa Jr. (2022) analisou a estrutura informacional de nove pacientes e a comparou com aquela de nove falantes do corpus C-ORAL-BRASIL. Para viabilizar esse estudo, que se baseia na comparação de corpora fundamentalmente diferentes, foi desenvolvida uma metodologia específica que é detalhada em Rocha et al. (2022) e Raso et al. (2024). Com base nela, Costa Jr. observa que a estrutura informacional dos esquizofrênicos é constantemente menos complexa daquela dos controles segundo vários parâmetros de análise. Como mostra o autor, pacientes realizam significativamente menos unidades textuais que o grupo controle e sua fala é caracterizada, em particular, por uma porcentagem significativamente menor de unidades de Tópico. Em uma amostra de 113 sequências terminadas (STs) com 2 unidades ilocucionárias e 48 STs com 3 unidades ilocucionárias extraídas de cada corpus, falantes do C-ORAL-BRASIL produziram 68 Tópicos (41 em STs com 2 unidades ilocucionárias e 27 em STs com 3 unidades ilocucionárias), enquanto pacientes do C-ORAL-ESQ realizam somente 20 (11 em STs com 2 unidades ilocucionárias e 9 em STs com 3 unidades ilocucionárias). Essa diferença, que é estatisticamente significativa segundo o teste U de Mann Whitney (Mann; Whit-

ney, 1947), é coerente com o prejuízo cognitivo dos pacientes, já que a unidade de Tópico é aquela que requer um maior esforço tanto cognitivo quanto prosódico. Esses dados podem ser observados na Tabela 6.

Tabela 6. Número de Tópicos presentes em unidades terminadas com 2 e 3 unidades ilocucionárias dos corpora C-ORAL-BRASIL e C-ORAL-ESQ

Corpus	Unidades ilocucionárias	Unidades terminadas analisadas	Tópicos	Teste U (Mann Whitney)
C-ORAL-BRASIL	2	113	41	p= 1.64025e-05
C-ORAL-ESQ	2	113	11	
C-ORAL-BRASIL	3	46	27	p = 0.005
C-ORAL-ESQ	3	46	9	

Fonte: Adaptado de Costa Jr. (2022).

Proeminências de tópico

Fiorilo e Rocha (em preparação) aplicaram uma metodologia inspirada em Martinez Sanchez et al. (2015) para verificar a capacidade dos pacientes com esquizofrenia em realizar a proeminência de Tópico em comparação com falantes do corpus C-ORAL-BRASIL. Os autores mediram em cerca de 100 Tópicos extraídos de cada corpus o percentual de sílabas com variação ≥ 4 ST, a variação de f_0 (máx.-mín.) e a taxa de variação de f_0 (variação/tempo) em sílabas nucleares e não nucleares. O teste de Shapiro-Wilk (1965) indicou desvio da normalidade ($p < 0,05$), portanto foram usados testes não paramétricos. O teste do qui-quadrado de Pearson com correção de continuidade de Yates (1934) revelou diferenças significativas nos percentuais de sílabas com variação de $f_0 \geq 4$ ST: nas sílabas nucleares, os falantes do C-ORAL-BRASIL apresentaram uma porcentagem de variação de $f_0 \geq 4$ ST em 41% dos casos contra 19% falantes do C-ORAL-ESQ ($X\text{-squared} = 12.145$, $df = 1$, $p\text{-value} = 0.0004921$). Nas sílabas não nucleares, a diferença é menor, mas ainda com significância estatística: 18% dos casos para falantes do C-ORAL-BRASIL e 9% para o C-ORAL-ESQ ($X\text{-squared} = 4.8056$, $df = 1$, $p\text{-value} = 0.02837$). Além disso, o teste de Wilcoxon (1945) indicou diferenças na variação e na taxa de variação de f_0 tanto em sílabas nucleares ($W = 4310$, $p = 0,0028$; $W = 4197$, $p = 0,0012$) quanto em não nucleares ($W = 10.292$, $p = 0,00034$; $W = 9168$, $p = 0,000001$). A Tabela 7 exibe as estatísticas descritivas analisadas.

Tabela 7. Variação de f0 (em ST), taxa de variação de f0 (em ST/s) e número de sílabas com variação igual ou superior a 4 ST em sílabas nucleares e não nucleares nos corpora C-ORAL-ESQ (ESQ) e C-ORAL-BRASIL (BR)

Tipo de sílaba	Corpus	Quantidade	Variação de f0 (ST)	Variação >=4 ST	Taxa de variação de f0 (ST/s)
<i>Nuclear</i>	ESQ	123	3.1 (DP 2.1)	23 (19%)	18.6 (DP 15.8)
	BR	92	3.5 (DP 2.5)	38 (41%)	19.9 (DP 18.0)
<i>Não-nuclear</i>	ESQ	158	2.2 (DP 1.8)	14 (9%)	16.6 (DP 21.2)
	BR	169	2.3 (DP 1.4)	30 (18%)	15.9 (DP 12.1)

Fonte: Adaptado de Fiorilo e Rocha (em preparação)

Pausas silenciosas

Fiorilo et al. (em preparação) investigaram a realização de pausas silenciosas na fala de pacientes com esquizofrenia, utilizando amostras do corpus C-ORAL-ESQ, e comparando-as àquelas produzidas por indivíduos do grupo controle, representado pelo corpus C-ORAL-MED. A metodologia adotada seguiu os procedimentos descritos por Saccone & Trillocco (2022), com análise das pausas identificadas nos dois primeiros minutos de quatro gravações de cada grupo, balanceadas por sexo. As pausas com duração mínima de 150 ms foram classificadas em quatro tipos: mudanças de turno (T), mudança de unidade terminada (UT), fronteiras de unidades informacionais (IU) e pausas em fronteira de unidade prosódica dentro de unidade informacional (SCA). Um total de 165 pausas foi identificado no grupo com esquizofrenia, frente a 145 no grupo controle. Os resultados indicam uma maior incidência de pausas longas (>1000 ms) no grupo com esquizofrenia em todos os tipos de pausa analisados. O teste de Shapiro-Wilk revelou desvio da normalidade nos quatro tipos de pausas ($p < 0,05$), motivo pelo qual foi aplicado o teste de Mann-Whitney para comparar os grupos com e sem esquizofrenia. Observou-se uma diferença estatisticamente significativa nas pausas UT ($U = 1124$, $p = 0.011$, $r = 0.33$), o que indica que a mediana do grupo com esquizofrenia ($Mdn = 762,53$ ms) foi significativamente maior do que a do grupo controle ($Mdn = 470,03$ ms). As diferenças observadas para os outros tipos de pausa não foram estatisticamente relevantes, embora a mediana também seja maior para pausas T (518,38 ms vs. 470,54 ms) e para pausas IU (456 ms vs. 416,76 ms) na esquizofrenia. Os dados confirmam tendências previamente observadas por Saccone & Trillocco (2022) e apontam para uma possível relação entre a duração das pausas e a monotonia percebida na fala de indivíduos com esquizofrenia.

Expressões faciais

Toledo (em preparação), em uma pesquisa em andamento, analisou diferenças nas expressões faciais entre os pacientes do C-ORAL-ESQ e do C-ORAL-MED. A pesquisadora selecionou 19 e 17 gravações, respectivamente, e extraiu 30 segundos de fala contínua de cada paciente. Pelo *software* OpenFace (Baltrušaitis et al., 2016), os trechos foram anotados automaticamente em relação às *action units* (Ekman; Friesen, 1978) – movimentos musculares breves da face, como erguer as sobrancelhas, que compõem em conjunto as expressões faciais. A partir das anotações, considerando apenas os frames de vídeo com ao menos 85% de confiabilidade na leitura facial, foram extraídas as medidas descritivas. Análises preliminares apontam para a hipótese de que pessoas com esquizofrenia realizam menos expressões faciais, com maior duração e menor intensidade, em comparação com pessoas sem essa patologia.

Medidas textuais e sintomas negativos

Lopes (2025), em pesquisa em desenvolvimento, analisa a existência de correlações entre medidas descritivas do corpus e o resultado de escalas psicométricas e clínicas aplicadas aos pacientes do C-ORAL-ESQ. Dentre outros resultados, o autor observou correlações significativas entre o número de palavras, enunciados, unidades prosódicas, palavras por turno e palavras retratadas produzidas por pacientes e diversos itens da subescala de sintomas negativos da PANSS (Kay et al., 1987). A Tabela 8 mostra as correlações significativas encontradas para a PANSS_N por meio do teste de correlação de Pearson. Vale notar que são todas correlações negativas, de modo que, quanto maior é o escore nos itens da subescala, menor é o número de cada variável (palavras, enunciados, unidades prosódicas, etc.) na fala do paciente. Em especial, os itens N3 (Fraco relacionamento) e N6 (Falta de espontaneidade e fluidez da conversa) correlacionam com todos os parâmetros, e a PANSS_N Total não correlaciona apenas com o número de palavras retratadas.

Tabela 8. Correlações estatisticamente significativas entre medidas textuais e escore dos itens da subescala negativa da PANSS

	Estatísticas	Total	N1	N2	N3	N4	N6
<i>Palavras</i>	Pearson's r	-0,368*			-0,393*		-0,573**
	2-tailed	0,023			0,015		0
	N	38			38		38
<i>Enunciados</i>	Pearson's r	-0,427**	-0,385*	-0,411**	-0,433**		-0,666**
	2-tailed	0,002	0,012	0,007	0,008	0,048	0,007
	N	42	42	42	42	42	42
<i>Unidades tonais</i>	Pearson's r	-0,467*	-0,316*	-0,439**	-0,432**	-0,343*	-0,554**
	2-tailed	-2	0,041	0,004	0,004	0,026	0
	N	42	42	42	42	42	42
<i>Palavras / turno</i>	Pearson's r	-0,337	-0,381*		-0,405**		-0,644**
	2-tailed	0,029	0,013		0,008		0
	N	42	42		42		42
<i>Palavras retratadas</i>	Pearson's r			-0,403**	-0,384*		-0,42
	2-tailed			0,008	0,018		0,006
	N			42	42		42

Fonte: Adaptado de Lopes (2025).

Conclusão

O trabalho apresentou um corpus de fala de indivíduos diagnosticados com esquizofrenia durante as consultas habituais com seus médicos, portanto em um contexto natural. O artigo descreve o corpus e todas as etapas metodológicas de compilação. O corpus é disponibilizado online aos pesquisadores. Foram também apresentados alguns estudos piloto. Esses estudos, ainda preliminares, permitem algumas comparações com grupos de controle: na fala esquizofrênica observa-se uma organização informacional significativamente menos complexa do que aquela dos controles, com uma porcentagem significativamente maior de pausas silenciosas longas, uma menor capacidade de realizar proeminências prosódicas e *action units* nas expressões faciais; essas últimas são em número menor e com duração maior. Foi também apresentado um trabalho que observa diversas correlações negativas entre as diversas medidas do corpus e itens da subescala negativa PANNS.

Referências

- BALTRUŠAITIS, T.; ROBINSON, P.; MORENCY, L.-P. OpenFace: an open source facial behavior analysis toolkit. In: IEEE WINTER CONFERENCE ON APPLICATIONS OF COMPUTER VISION (WACV), 2016, Lake Placid. *Proceedings [...]*. New York: IEEE, 2016. p. 1-10. DOI: 10.1109/WACV.2016.7477553.
- BARBOSA, P. A.; RASO, T. Spontaneous Speech Segmentation: Functional and Prosodic Aspects With Applications for Automatic Segmentation / A segmentação da fala espontânea: aspectos prosódicos, funcionais e aplicações para a tecnologia. *Revista de Estudos da Linguagem*, v. 26, p. 1361-1396, 2018.
- BELL, M. D.; LYSACKER, P. H. Clinical benefits of paid work activity in schizophrenia: 1-year follow-up. *Schizophrenia Research*, v. 45, n. 1-2, p. 109-120, 2000.
- BICK, E. *VISL & CG-3: Constraint Grammar on the Move: An application-driven paradigm*. In: HURSKAINEN, A.; KOSKENNIEMI, K.; PIRINEN, T. (org.). *Rule-Based Language Technology*. Tartu: University of Tartu, 2023. (NEALT Monograph Series, v. 2), p. 112-140.
- BICK, E. A anotação gramatical do C-ORAL-BRASIL. In: RASO, T.; MELLO, H. (org.). *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora da UFMG, 2012. p. 223-254.
- BICK, E. *The parsing system "PALAVRAS": automatic grammatical analysis of Portuguese in a constraint grammar framework*. Aarhus: Aarhus University Press, 2000. 412 p. Disponível em: <https://edu.visl.dk/~eckhard/pdf/PLP20-amilo.ps.pdf>. Acesso em: 29 mai. 2025.
- BRESSEM, J. A linguistic perspective on the notation of form features in gestures. In: MÜLLER, C.; CIENKI, A.; FRICKE, E.; LADEWIG, S.; MCNEILL, D.; TESSENDORF, S. (org.). *Body - Language - Communication*, v. 1. Berlin/Amsterdam/New York: De Gruyter Mouton, 2013. p. 1079-1098.
- CORCORAN, R.; MERCER, G.; FRITH, C. D. Schizophrenia, symptomatology and social inference: Investigating "theory of mind" in people with schizophrenia. *Schizophrenia Research*, v. 17, n. 1, p. 5-13, 1995.
- COSTA-Jr, J. C. *Padrão informacional de stanzas de pacientes com esquizofrenia*. Belo Horizonte: UFMG, 2022. Tese (Doutorado em Estudos Linguísticos) do Programa de Pós-Graduação em Estudos Linguísticos, Faculdade de Letras, UFMG, Belo Horizonte, 2022.
- CRESTI, E. *Corpus di Italiano parlato*. Firenze: Accademia della Crusca, 2000a. 2 v.
- CRESTI, E.; MONEGLIA, M. *C-ORAL-ROM: integrated reference corpora for spoken romance languages*. Amsterdam/Philadelphia: John Benjamins, 2005.
- EKMAN, P.; FRIESEN, W. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto: Consulting Psychologists Press, 1978.
- FIORILLO, I.; ROCHA, B. *As proeminências de tópico em indivíduos com e sem esquizofrenia*. (em preparação).

- FIORILLO, I.; VITAL, A.; ROCHA, B.; RASO, T. *Pausas silenciosas em indivíduos com e sem esquizofrenia*. (em preparação).
- FLEISS, J. L. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, v. 76, n. 5, p. 378-382, 1971. DOI: <https://doi.org/10.1037/h0031619>.
- IZRE'EL, S.; MELLO, H.; PANUNZI, A.; RASO, T. Introduction. In search of a basic unit of spoken language: Segmenting speech. In: IZRE'EL, S.; MELLO, H.; PANUNZI, A.; RASO, T. (org.). *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins Publishing Company, 2020, p. 1-32. Disponível em: <https://doi.org/10.1075/scl.94.int>. Acesso em: 15 de maio 2025.
- KAY, S. R.; FISZBEIN, A.; OPLER, L. A. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, v. 13, n. 2, p. 261-276, 1987.
- KEEFE, R. S. E.; GOLDBERG, T. E.; HARVEY, P. D.; GOLD, J. M.; POE, M. P.; CUDEMAN, W. L. The Brief Assessment of Cognition in Schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. *Schizophrenia Research*, v. 68, n. 2-3, p. 283-297, 2004.
- KENDON, A. *Gesture: Visible Action as Utterance*. Cambridge: University Press, 2004.
- LOPES, L. M. C. *A produção da linguagem na esquizofrenia e a sua associação com os sintomas negativos e cognitivos da doença: um estudo transversal e observacional baseado em um corpus linguístico*. 2025. Dissertação (Mestrado em Neurociências) do Programa de Pós-Graduação em Neurociências, Instituto de Ciências Biológicas, UFMG, Belo Horizonte, 2025.
- MACWHINNEY, B. J. *The CHILDES Project: tools for analyzing talk*. 3rd ed. Mahwah: Lawrence Erlbaum Associates, 2000. 2 v.
- MANN, H. B.; WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, v. 18, p. 50-60, 1947. DOI: <https://doi.org/10.1214/aoms/1177730491>.
- MARTÍNEZ-SÁNCHEZ, F.; MUELA-MARTÍNEZ, J. A.; CORTÉS-SOTO, P.; GARCÍA MEILÁN, J. J.; VERA FERRÁNDIZ, J. A.; EGEA CAPARRÓS, A.; PUJANTE VALVERDE, I. M. Can the Acoustic Analysis of Expressive Prosody Discriminate Schizophrenia? *Spanish Journal of Psychology*, v. 18, E86, 2015. DOI: 10.1017/sjp.2015.85. PMID: 26522128.
- MELLO, H. R. Methodological issues for spontaneous speech corpora compilation: The case of C-ORAL-BRASIL. In: RASO, T.; MELLO, H. R. (org.). *Spoken corpora and linguistic studies*. Amsterdam: John Benjamins Publishing Company, 2014. p. 27-68. (Studies in Corpus Linguistics, v. 61).
- MELLO, H.; RASO, T.; MITTMANN, M.; VALE, H.; CÔRTEZ, P. Transcrição e segmentação prosódica do corpus C-ORAL-BRASIL: critérios de implementação e validação. In: RASO, T.; MELLO, H. (org.). *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora da UFMG, 2012. p. 125-145.
- MONEGLIA, M.; CRESTI, E. L'intonazione e i criteri di trascrizione del parlato adulto e infantile. In: BORTOLINI, U.; PIZZUTO, E. (org.). *Il Progetto CHILDES Italia*. Pisa: Del Cerro, 1997. p. 57-90.

- MONEGLIA, M.; RASO, T. Notes on Language into Act Theory (L-AcT). In: RASO, T.; MELLO, H. (org.). *Spoken Corpora and Linguistic Studies*. Amsterdam/Philadelphia: John Benjamins, 2014. p. 468-494.
- PATTERSON, T. L.; GOLDMAN, S.; MCEVOY, J. P.; KAPLAN, R. M.; JUKES, S.; DAVIDSON, M. Development and validation of a performance-based measure of functional capacity in schizophrenia: the UCSD Performance-Based Skills Assessment (UPSA). *Schizophrenia Research*, v. 48, n. 1, p. 35-44, 2001.
- RASO, T. O corpus C-ORAL-BRASIL. In: RASO, T.; MELLO, H. (org.). *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal*. 1. ed. Belo Horizonte: Editora UFMG, 2012a, p. 55-90.
- RASO, T. O C-ORAL-BRASIL e a Teoria da Língua em Ato. In: RASO, T.; MELLO, H. (org.). *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora da UFMG, 2012b. p. 91-124.
- RASO, T.; MELLO, H. *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora da UFMG, 2012.
- RASO, T.; ROCHA, B.; SALGADO, J. V.; CRUZ, B. F.; MANTOVANI, L. M. The C-ORAL-ESQ project: a corpus for the study of spontaneous speech of individuals with schizophrenia. *Language Resources & Evaluation*, v. 58, p. 903-923, 2024. DOI: <https://doi.org/10.1007/s10579-023-09675-y>.
- ROCHA, B.; RASO, T.; MELLO, H.; FERRARI, L. Information structure in the speech of individuals with schizophrenia: Methodology and first analyses from complex structure of corpus based data. *CHIMERA: Revista De Corpus De Linguas Romances Y Estudios Lingüísticos*, v. 9, p. 217-242, 2022. DOI: <https://doi.org/10.15366/chimera2022.9.010>.
- SACCONI, V.; TRILLOCCO, S. Segmentation of the Speech Flow for the Evaluation of Spontaneous Productions in Pathologies Affecting the Language Capacity. 4 Case Studies of Schizophrenia. In: KOKKINAKIS, D.; THEMISTOCLEOUS, C. K.; LUNDHOLM FORS, K.; TSANAS, A.; FRASER, K. C. (org.). *Proceedings [...]*. Marseille, France: European Language Resources Association, 2022. p. 94-99. Disponível em: <https://aclanthology.org/2022.rapid-1.12/>. Acesso em : 17 de setembro de 2025.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, v. 52, n. 3-4, p. 591-611, 1965.
- SIMPSON, G. M.; ANGUS, J. W. A rating scale for extrapyramidal side effects. *Acta Psychiatrica Scandinavica Supplementum*, v. 212, p. 11-19, 1970.
- WILCOXON, F. Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, v. 1, n. 6, p. 80-83, 1945.
- WITTERBURG, P.; BRUGMAN, H.; RUSSEL, A.; KLASSMANN, A.; SLOETJES, H. ELAN: A Professional Framework for Multimodality Research. In: LANGUAGE RESOURCES AND EVALUATION CONFERENCE (LREC), 2006. *Proceedings [...]*. Paris: European Language Resources Association (ELRA), 2006. p. 1556-1559.
- YATES, F. Contingency table involving small numbers and the χ^2 test. *Supplement to the Journal of the Royal Statistical Society*, v. 1, n. 2, p. 217-235, 1934.

Declaração de disponibilidade de acesso a dados

O corpus C-ORAL-ESQ encontra-se integralmente disponível gratuitamente em <<http://www.c-oral-brasil.org/c-oral-esq.php>>.

Contribuições dos autores

Os autores elaboraram em conjunto as seções *Introdução*, *As gravações*, *Etapas de compilação do corpus*, *Os pacientes* e *Medidas gerais do corpus*. Contudo, Bruno Rocha foi responsável pelas seções *As gravações*, *Os pacientes*, *Medidas textuais e sintomas negativos*. Tommaso Raso foi responsável por *Etapas de compilação do corpus*, *Estrutura informacional* e *Conclusão*. Átila Vital foi responsável por *Medidas gerais do corpus* e *Pausas silenciosas*. Gabriela Toledo foi responsável pela seção *Expressões faciais*. Ísis Fiorilo foi responsável por *Proeminências de tópico*.

Conflito de interesse

Os autores declaram a inexistência de conflitos de interesse relacionados ao artigo.

Financiamento

Não aplicável.

Agradecimentos

Os autores agradecem ao Instituto de Estudos Avançados e Transdisciplinares (IEAT) da UFMG.

Anexo I – Arquivo de metadados de uma gravação do C-ORAL-ESQ

@Title: Consultation_ROM

@File: besqmm33

@Participants: P33, RON (male, B, 2, unemployed, patient, Belo Horizonte/MG)
M19, GAL (male, C, 3, resident psychiatrist, psychiatrist, Belo Horizonte/MG)
A13, MAO (female, C, 1, unemployed, companion_mother, Novo Cruzeiro/MG)

@Date: 25/10/2020

@Place: Hospital das Clínicas, Belo Horizonte/MG

@Situation: Psychiatric consultation, second recording of this patient

@Topic: About voices, study, medication, sleep, and videogame

@Source: C-ORAL-ESQ

@Class: Medical_schizophrenia

@Length: 28'54"

@Words: 2379

@Patient_words: 548

@Physician_words: 995

@Companion_words: 836

@Intervenant_words: Not applicable

@Acoustic_quality: AB

@Transcriber: Átila Vital

@Revisor: Thaís Viana, Geisibel Alves Silva, Gabriela Toledo

@Video_cuts: (1) video/audio cut: 00'30", the doctor types on the computer; (2) video/audio cut: 11'51", the doctor left the room; (3) video/audio cut: 00'22", moment without dialogue; (4) video/audio cut: 00'11", moment without dialogue

@Comments:

1) Conventionalized forms: tô (estou), xá (deixa), des (deles), e' (ele), tem (tenho), tá (está), d' (de), pa (para), ea (ela), eas (elas), siora (senhora), de' (dele), No' (Nossa Senhora), pr' (para), c' (com), xá (deixa), vão (vamos)

2) Apheretic forms: cê (você), tá (está), tar (estar), tão (estão), tão (então), tendeu (entendeu), tô (estou), tendi (entendi), pregnado (impregnado), gora (agora), taria (estaria), tava (estava), xá (deixa), brigado (obrigado)

3) Reduced diminutives: diretim (diretinho), minutim (minutinho)

4) Acronyms: emeele (mL), PEG

5) Foreign words: NoNe

6) Onomatopoeias: NoNe

7) Production errors: At 06'04", RON pronounces "omeprazol" as "omprazá"; at 13'04", RON pronounces "Haldol" as "odol"; at 13'11", MAO pronounces "risperidona" as "rispiridona"

8) Other: At 07'27", RON yawns speaking something like "foi muito bom"

Anexo II – Medidas de palavras e duração de cada consulta

Gravação	Paciente	Médico	Acompanhante	Interveniente	Total	Proporção paciente-médico	Duração (mm:ss)
<i>besqau01</i>	2.469	1.063	212	0	3.744	70%	18:36
<i>besqau02</i>	2.470	1.833	0	0	4.303	57%	27:21
<i>besqau03</i>	2.047	840	0	0	2.887	71%	16:12
<i>besqau04</i>	543	268	41	0	852	67%	06:21
<i>besqau05</i>	1.544	261	0	0	1.805	86%	11:10
<i>besqau06</i>	190	365	297	168	1.020	34%	05:24
<i>besqau07</i>	1.925	1.255	544	0	3.724	61%	19:38
<i>besqau08</i>	1.109	1.723	0	0	2.832	39%	18:33
<i>besqau09</i>	1.778	662	0	0	2.440	73%	12:06
<i>besqau10</i>	265	1.158	10	2	1.435	19%	11:34
<i>besqau11</i>	865	2.322	978	0	4.165	27%	26:53
<i>besqau12</i>	267	575	189	0	1.031	32%	06:24
<i>besqau13</i>	1.736	718	0	0	2.454	71%	14:52
<i>besqau14</i>	1.485	910	0	0	2.395	62%	22:54
<i>besqau15</i>	775	1.152	0	13	1.940	40%	12:18
<i>besqau16</i>	319	1.481	376	0	2.176	18%	11:27
<i>besqau17</i>	1.296	813	0	0	2.109	61%	16:17
<i>besqau18</i>	257	1.054	556	0	1.867	20%	12:50
<i>besqau19</i>	648	1.119	281	0	2.048	37%	14:30
<i>besqau20</i>	2.489	1.119	710	0	4.318	69%	23:51
<i>besqau21</i>	2.252	633	0	0	2.885	78%	16:59
<i>besqau22</i>	1.909	2.147	0	0	4.056	47%	22:10
<i>besqau23</i>	2.599	1.218	0	0	3.817	68%	18:00
<i>besqau24</i>	509	963	93	0	1.565	35%	12:36
<i>besqmm25</i>	306	489	0	110	905	38%	18:04
<i>besqmm26</i>	1.569	2.022	0	53	3.644	44%	21:52
<i>besqmm27</i>	642	994	0	120	1.756	39%	15:17
<i>besqmm28</i>	394	2.052	0	72	2.518	16%	25:25
<i>besqmm29</i>	1.510	1.546	0	0	3.056	49%	22:29
<i>besqmm30</i>	272	619	836	2	1.729	31%	13:02
<i>besqmm31</i>	348	561	0	0	909	38%	08:15
<i>besqmm32</i>	661	1.160	0	0	1.821	36%	14:59
<i>besqmm33</i>	502	953	804	0	2.259	35%	28:54
<i>besqmm34</i>	1.621	970	0	0	2.591	63%	17:54
<i>besqmm35</i>	284	577	0	0	861	33%	05:30
<i>besqmm36</i>	1.811	1.870	0	0	3.681	49%	17:28
<i>besqmm37</i>	990	949	255	0	2.194	51%	17:58
<i>besqmm38</i>	462	982	0	7	1.451	32%	15:04
<i>besqmm39</i>	429	2.196	645	21	3.291	16%	37:31
<i>besqmm40</i>	270	529	0	0	799	34%	05:25
<i>besqmm41</i>	332	634	52	0	1.018	34%	08:48
<i>besqmm42</i>	2.460	1.803	176	46	4.485	58%	28:16
<i>besqmm43</i>	1.113	1.422	0	32	2.567	44%	16:55

Fonte: Elaborada pelos próprios autores