

UNIVERSIDADE FEDERAL DE MINAS GERAIS

LÚCIA DE ALMEIDA FERRARI

**Aspectos prosódicos e sintáticos dos pronomes clíticos em
português do Brasil e no vernáculo florentino**

BELO HORIZONTE

2015

Lúcia de Almeida Ferrari

**Aspectos prosódicos e sintáticos dos pronomes clíticos em
Português do Brasil e no vernáculo florentino.**

Tese apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do título de Doutor em Língua Teórica e Descritiva.

Área de Concentração: Linguística Teórica e Descritiva

Linha de Pesquisa: Estudos Linguísticos Baseados em Corpora

Orientador: Prof. Dr. Tommaso Raso

Belo Horizonte
Faculdade de Letras da UFMG
2015

A Lucílio, sempre.

Agradecimentos

Quero agradecer todos aqueles que direta ou indiretamente, contribuíram para a realização desta Tese.

Gostaria de agradecer aos professores da UFMG que tanto me ensinaram, em especial à Profa. Heliana Mello, pelo seu profissionalismo e seriedade.

Aos membros desta Banca, que prontamente aceitaram o convite, em especial à Profa. Lúcia Fulgêncio e ao Prof. Plínio Barbosa pelos preciosos conselhos já na fase de qualificação deste trabalho, e também antes.

Aos professores do LABLITA, Profa. Emanuela Cresti e Prof. Massimo Moneglia, juntamente com os colegas do laboratório, Alessando e Lorenzo, por ter me ajudado durante minha estadia em Florença.

À Capes, pelo suporte financeiro durante os três meses de estágio de doutoramento “sanduíche” em Florença, Itália.

A Horácio e ao Prof. César Reis, que me ensinaram como utilizar o Praat e como segmentar.

Aos colegas do LEEL, especialmente pela ajuda final: Bárbara, Marcelo e Luís, que teve que ler tudo!

A Maryualê, pela ajuda, apoio e amizade.

A Heloisa, por sempre se lembrar.

A Chiara, por ter me considerado, imerecidamente, seu modelo. Que eu não te decepcione nunca.

Marco, queria te agradecer pela paciência e compreensão, nos altos e baixos, nos feriados, férias e finais de semana em que fiquei trabalhando. E pelo suporte técnico.

Um agradecimento especial a minha mãe, que sempre acreditou e torceu para que eu chegasse aqui.

E mais do que todos quero agradecer meu orientador, Prof. Tommaso Raso, por ter acreditado que eu merecia seu tempo, por tanto ter se dedicado à minha formação, ter me apoiado e incentivado. Um obrigado especial por ter guiado meus passos e ter tido uma paciência infinita comigo. Por ter sido muito mais que um mestre justo e exigente, mas também um amigo compreensivo e disponível.

*“[...] Ho sceso milioni di scale dandoti il braccio
non già perché con quattr’occhi forse si vede di più.
Con te le ho scese perché sapevo che di noi due
le sole vere pupille, sebbene tanto offuscate,
erano le tue.”*

Eugenio Montale (*Xenia*)

RESUMO

Este trabalho trata das formas pronominais no PB e no vernáculo florentino. A análise adota uma metodologia inovadora através do estudo de corpora de fala espontânea: o C-ORAL-BRASIL para o PB, e o *Corpus Stammerjohann* e o *Corpus per il Confronto Diacronico* para o florentino. O estudo tem por objetivos: (I) no PB, quantificar e analisar a distribuição das formas pronominais *cê(s)/ocê(s)/ocê(s)*; (II) no vernáculo florentino, investigar o uso dos pronomes clíticos de terceira pessoa singular sujeito *gl'/e/la/l'* em conjunto com os verbos *essere* (ser), *fare* (fazer), *andare* (ir) e *venire* (vir) e fazer um estudo diacrônico comparando os resultados dos dois corpora consultados, gravados a uma distância de cerca de trinta-quarenta anos um do outro.

A novidade da pesquisa reside na metodologia adotada: ao invés de inferir o valor clítico ou tônico das formas estudadas a partir de uma análise morfossintática, sua tonicidade ou atonicidade foi determinada a partir de dados prosódicos. Os pronomes foram analisados acusticamente em sua duração e, a partir da forma suprasegmental, foi possível estabelecer sua função. O estudo quis observar processos diferentes em duas línguas: no florentino as formas clíticas estão plenamente gramaticalizadas, portanto, a partir de sua forma segmental é possível estabelecer com certeza sua função. Já no PB a gramaticalização ainda não foi completada: uma mesma forma pode ser tônica ou átona e ocupar posições e funções diferentes. A comparação entre as duas línguas permite, então, observar o comportamento de formas já estabelecidas - aquelas em florentino - e as diferentes distribuições a que estão sujeitas as formas *cê(s)/ocê(s)/ocê(s)* no PB. A análise acústica efetuada é, portanto, determinante para se entender melhor o processo de mudança pelo qual estão passando os pronomes de segunda pessoa no PB. No florentino, a análise diacrônica permitiu verificar a redução do uso das formas clíticas sujeito no arco de trinta a quarenta anos da qual se fala na literatura mas que, até este momento, não foi comprovada com estudos quantitativos.

Os resultados obtidos indicam: (I) no PB, a confirmação da hipótese inicial de que a forma suprasegmental evidencia durações, e portanto funções, diferentes para as mesmas formas; (II) no florentino, uma redução significativa do sujeito clítico, anterior à época investigada.

Palavras-chave: linguística de corpus; fala espontânea; análise acústica; cliticização

ABSTRACT

This study investigates pronominal forms in BP and in vernacular Florentine. Our analysis adopts an innovative methodology by using spontaneous speech corpora: for BP, the C-ORAL-BRASIL, and for Florentine, the *Corpus Stammerjohann* and the *Corpus per il Confronto Diacronico*. Our study aims to: (I) in BP, quantify and analyze the distribution of pronominal forms *cê(s)/ocê(s)/você(s)*; (II) in vernacular Florentine, investigate the usage of 3rd person singular clitic pronouns *gl'/e/la/l'* when used together with the verbs *essere* (to be), *fare* (to do), *andare* (to go) and *venire* (to come) and compare diachronically the results of the two analyzed corpora, recorded thirty to forty years apart. This research's innovative element lies in the methodology we adopted: instead of inferring the clitic or tonic value of the studied forms by using morphosyntactic analysis, their tonicity or atonicity was determined based on prosodic data. Pronouns' duration was acoustically analyzed and, based on their suprasegmental form, it was possible to establish their function. Our study aimed to observe different processes in two languages: in Florentine, clitic forms are fully grammaticalized, so it's possible to establish their function based on their segmental form. In BP, grammaticalization process is not complete yet: the same form can be tonic or atonic and can occupy different functions and positions. Comparison between the two languages, therefore, allows us to observe the behavior of well-established forms – those in Florentine – and the different distributions of the *cê(s)/ocê(s)/você(s)* forms in BP. Acoustical analysis is, therefore, crucial to a better understanding of the changes that are affecting 2nd person pronouns in BP. In Florentine, diachronic analysis allows us to verify the reduction in usage of subject clitic forms through a thirty to forty years arc that is mentioned in literature but that had not been previously proved by quantitative research.

Results indicate: (I) in BP, the confirmation of our hypothesis that suprasegmental form presents different durations and, therefore, different functions for the same forms; (II) in Florentine, a significant reduction of the clitic subject, preceding the investigated period.

Key-words: Corpus linguistics; spontaneous speech; acoustic analysis; cliticization

Sumário

1. INTRODUÇÃO	16
1.1. Justificativa	16
1.2. Objetivos	17
1.3. Hipótese de trabalho	20
2. REVISÃO DA LITERATURA.....	21
2.1. História do pronome <i>você</i>	21
2.2. Questões abordadas nas propostas	22
2.3. Gramaticalização e cliticização	23
2.3.1. A forma <i>cê</i> : gramaticalização ou forma fraca?.....	33
2.3.2. Outras abordagens	41
2.3.3. Síntese das propostas para as formas <i>você(s)/ocê(s)/cê(s)</i>	60
2.4. Análise do vernáculo florentino.....	61
2.4.1. Breve história da língua italiana e do vernáculo florentino.....	62
2.4.2. Pronomes pessoais sujeito e terceira pessoa singular no vernáculo florentino ..	67
2.4.3. A escolha do pronome de terceira pessoa singular.....	73
2.4.4. O verbo <i>essere</i> (ser) e suas variações semânticas no italiano: motivos da escolha	74
2.4.5. Outros verbos.....	77
3. METODOLOGIA: ANÁLISE PARALELA <i>CORPUS-BASED</i> SOBRE DUAS LÍNGUAS.....	79
3.1. Análise do PB	79
3.1.1. Escolha do <i>corpus</i> : descrição do C-ORAL-BRASIL.....	81
3.1.1.1. Análise do <i>corpus</i>	89
3.1.1.2. Revisão da transcrição.....	90
3.1.1.3. Critérios adotados para a seleção das formas objeto de pesquisa	90
3.1.1.4. Taxonomia das anotações em tabelas do tipo <i>Excel</i>	96
3.1.2. Análise acústica: descrição do método utilizado.....	98
3.1.2.1. Definição dos critérios de escolha dos enunciados para análise	102
3.1.2.2. Normalização dos dados.....	104
3.1.2.3. O software <i>SGdetector</i> : funcionamento e princípios básicos.....	104
3.1.3. Validação estatística	106
3.2. Análise do vernáculo florentino.....	107

3.2.1.	O <i>Corpus Stammerjohann</i> : a coleta original	107
3.2.2.	O trabalho de remasterização, transcrição e etiquetagem.....	108
3.2.3.	O <i>Corpus per il Confronto Diacronico</i> : critérios e seleção dos textos	111
4.	RESULTADOS	117
4.1.	Resultados do PB	117
4.1.1.	Primeira varredura do corpus com o <i>software AntConc</i>	117
4.1.2.	Resultados parciais das formas e ocorrências encontradas	117
4.1.3.	Resultados da forma <i>você(s)</i>	127
4.1.4.	Resultados da forma <i>ocê(s)</i>	129
4.1.5.	Resultados da forma <i>cê(s)</i>	132
4.1.6.	Resultados da análise acústica.....	136
4.1.6.1.	Função sintática e função informacional.....	141
4.1.6.1.1.	O valor ilocucionário.....	142
4.1.6.1.2.	Parentético.....	150
4.1.6.1.3.	As unidades escansionadas.....	152
4.1.6.1.4.	Tópico.....	157
4.1.6.1.5.	Motivação pragmática	159
4.1.7.	A marcação da função sintática das formas <i>cê</i> , <i>ocê</i> e <i>você</i>	164
4.1.7.1.	Forma <i>cê</i> em posição pré-verbal	164
4.1.7.2.	Forma <i>cê</i> em posição pós-verbal	170
4.1.7.3.	Forma <i>ocê</i> em posição pré-verbal	183
4.1.7.4.	Forma <i>ocê</i> em posição pós-verbal	188
4.1.7.5.	Forma <i>você</i> em posição pré-verbal.....	196
4.1.7.6.	Forma <i>você</i> em posição pós-verbal	204
4.1.8.	Validação estatística	209
4.2.	Resultados do vernáculo florentino	213
4.2.1.	Verbo <i>essere</i>	213
4.2.2.	Discussão	220
4.2.3.	Verbos <i>fare</i> , <i>andare</i> e <i>venire</i>	224
4.2.3.1.	Análise de <i>fa</i>	224
4.2.3.2.	Análise de <i>va</i>	227
4.2.3.3.	Análise de <i>viene</i>	229
4.2.4.	Discussão	231

5. CONCLUSÕES.....	232
5.1. Conclusões sobre as formas do pronome de segunda pessoa no PB	232
5.2. Conclusões sobre os pronomes sujeito no vernáculo florentino.....	234
5.3. Conclusões gerais	236
APÊNDICE A: Conteúdos do CD	239
REFERÊNCIAS	240

Lista de Tabelas

Tabela 1	Ocorrências das formas pronominais de segunda pessoa singular submetidas a análise acústica.....	103
Tabela 2	Ocorrências dos textos analisados em base à sua qualidade acústica.....	103
Tabela 3	Apresentação das ocorrências e distribuição das formas pronominais de segunda pessoa, subdivididas por posições e funções, encontradas no C-ORAL-BRASIL.....	118
Tabela 4	Total das ocorrências das formas pronominais de segunda pessoa excluindo-se os casos de interrupções do enunciado.....	120
Tabela 5	Total das ocorrências das formas pronominais de segunda pessoa excluindo-se os casos de <i>retracting</i>	120
Tabela 6	Total das ocorrências das formas pronominais de segunda pessoa excluindo-se os casos de interrupção do enunciado e de <i>retracting</i>	121
Tabela 7	Apresentação das ocorrências e distribuição da forma <i>você(s)</i> , encontradas no C-ORAL-BRASIL.....	127
Tabela 8	Apresentação das ocorrências e distribuição da forma <i>ocê(s)</i> , encontradas no C-ORAL-BRASIL.....	129
Tabela 9	Apresentação das ocorrências e distribuição da forma <i>cê(s)</i> , encontradas no C-ORAL-BRASIL.....	132
Tabela 10	Ocorrências das formas pronominais de segunda pessoa singular submetidas a análise acústica.....	137
Tabela 11	Resultados das análises acústicas da forma <i>cê</i> em posição pré-verbal no <i>corpus</i> C-ORAL-BRASIL.....	164
Tabela 12	Resultados das análises acústicas da forma <i>cê</i> em posição pós-verbal no <i>corpus</i> C-ORAL-BRASIL.....	170
Tabela 13	Resultados das análises acústicas da forma <i>ocê</i> em posição pré-verbal no <i>corpus</i> C-ORAL-BRASIL.....	183
Tabela 14	Resultados das análises acústicas da forma <i>ocê</i> em posição pós-verbal no <i>corpus</i> C-ORAL-BRASIL.....	187
Tabela 15	Resultado das análises acústicas da forma <i>you</i> em posição pré-verbal no <i>corpus</i> C-ORAL-BRASIL.....	195
Tabela 16	Resultado das análises acústicas da forma <i>you</i> em posição pós-verbal no <i>corpus</i> C-ORAL-BRASIL.....	203
Tabela 17	Formas de expressão do sujeito encontradas nos dois <i>corpora</i> comparados em valores absolutos e normalizados.....	213
Tabela 18	Formas de sujeito duplicado encontradas nos dois <i>corpora</i> comparados em termos absolutos e normalizados.....	219
Tabela 19	Total dos sujeitos clíticos encontrados nos dois <i>corpora</i> comparados, em porcentagem sobre os sujeitos expressos, divididos por faixas etárias.....	221
Tabela 20	Dados absolutos de <i>fa</i> nos dois <i>corpora</i>	224
Tabela 21	Dados normalizados por 100.000.....	225
Tabela 22	Análise das formas de expressão do sujeito dos enunciados com a forma verbal <i>fa</i>	226
Tabela 23	Ocorrências dos sujeitos duplicados com a forma verbal <i>fa</i>	227
Tabela 24	Dados dos dois <i>corpora</i> normalizados por 100.000.....	228
Tabela 25	Análise das formas de expressão do sujeito dos enunciados com a forma verbal <i>va</i>	228
Tabela 26	Ocorrências dos sujeitos duplicados com a forma verbal <i>va</i>	229
Tabela 27	Dados absolutos de <i>viene</i> nos dois <i>corpora</i>	230
Tabela 28	Dados dos dois <i>corpora</i> normalizados por 100.000.....	230
Tabela 29	Análise das formas de expressão do sujeito dos enunciados com a forma verbal <i>viene</i>	230
Tabela 30	Total dos sujeitos clíticos encontrados nos dois <i>corpora</i> comparados, em porcentagem sobre os sujeitos expressos, divididos por faixas etárias.....	231

Lista de Figuras

Figura 1	Imagem da tabela <i>Excel</i> preenchida com os dados sobre a forma <i>cê</i> e sua análise.....	97
Figura 2	Espectro do pronome “você” segmentado no programa Praat: é possível visualizar os formantes, em específico as vogais [o] e [e] e a sibilante não vozeada [s] descritas acima.....	100
Figura 3	Tela do Praat com segmentação em código ASCII (primeira linha).....	105
Figura 4	Abertura do script <i>SGDetector</i> e inserção dos dados para normalização.....	105
Figura 5	Saída de dados do <i>SGDetector</i> : gráfico e arquivos de texto.....	106
Figura 6	Tela do <i>Praat</i> com o texto (bfamcv18-187): é mostrado o espectrograma e nas camadas de anotação: o texto, sua segmentação em palavras e em código ASCII. É evidenciada a ilocução.....	144
Figura 7	Tela do <i>Praat</i> com o texto (bpubdl01-90-91): é evidenciado o último enunciado, ilocutivo.....	145
Figura 8	Tela do <i>Praat</i> com o texto (bfamcv13-33): é evidenciado o ilocutivo.....	146
Figura 9	Tela do <i>Praat</i> com o texto (bfamcv03-247): é evidenciado o ilocutivo “é ocê”.....	147
Figura 10	Tela do <i>Praat</i> com o texto (bpubdl11-241): é evidenciado o ilocutivo “pr’ocê”.....	148
Figura 11	Tela do <i>Praat</i> com o texto (bpubcv04-61): é assinalada a forma <i>você</i> ; o ilocutivo é “e você Héliida”.....	149
Figura 12	Tela do <i>Praat</i> com o texto (bfamcv13-261): é evidenciado o parentético “com cê”.....	151
Figura 13	Tela do <i>Praat</i> com o texto (bfamcv13-261): note-se que foi retirado o parentético “com cê”.....	152
Figura 14	Tela do <i>Praat</i> com o texto (bfamcv31-187): é evidenciada a escansão “cê acertar direitim”.....	153
Figura 15	Tela do <i>Praat</i> com o texto (bfamcv32-105).....	154
Figura 16	Tela do <i>Praat</i> com o texto (bfammn27-64).....	155
Figura 17	Tela do <i>Praat</i> com o texto (bpubmn11-24).....	156
Figura 18	Tela do <i>Praat</i> com o texto (bfamd105-185).....	157
Figura 19	Tela do <i>Praat</i> com o texto (bfammn02-47): é evidenciado o tópico, <i>cê</i>	158
Figura 20	Tela do <i>Praat</i> com o texto (bfammn03-115).....	159
Figura 21	Tela do <i>Praat</i> com o texto (bfamcv19-27).....	160
Figura 22	Tela do <i>Praat</i> com o texto (bfamd105-282).....	161
Figura 23	Tela do <i>Praat</i> com o texto (bfamcv20-190).....	162
Figura 24	Tela do <i>Praat</i> com o texto (bfamcv02-335-RUT): é posta em evidência a forma <i>cê</i> pré-verbal.....	165
Figura 25	Tela do <i>Praat</i> com o texto (bfamcv18-208-HER): é posta em evidência a forma <i>cê</i> pré-verbal.....	167
Figura 26	Tela do <i>Praat</i> com o texto (bfamd102-45-BAL): é posta em evidência a forma <i>cê</i> pré-verbal.....	168
Figura 27	Tela do <i>Praat</i> com o texto (bfamcv05-240-CEL): é posto em evidência o <i>cê</i> pós-verbal.....	171
Figura 28	Tela do <i>Praat</i> com o texto (bfamcv04-453-BRU): é posto em evidência o <i>cê</i> sujeito posposto.....	173
Figura 29	Tela do <i>Praat</i> com o texto (bpubdl04-260-ELI): é posto em evidência o <i>cê</i> preposicionado.....	174
Figura 30	Tela do <i>Praat</i> com o texto (bfamcv23-28-DIN): é posto em evidência o <i>cê</i> preposicionado.....	175
Figura 31	Tela do <i>Praat</i> com o texto (bfamd102-247-BAL): é posto em evidência o <i>cê</i> preposicionado.....	176
Figura 32	Tela do <i>Praat</i> com o texto (bfammn29-6-JDL): é posto em evidência o <i>cê</i> preposicionado.....	177
Figura 33	Tela do <i>Praat</i> com o texto (bfamcv03-215-TON): é posto em evidência o <i>cê</i> preposicionado.....	178
Figura 34	Tela do <i>Praat</i> com o texto (bfamcv23-43-DIN): para melhor visualização foi inserido somente o trecho da unidade tonal analisada.....	180
Figura 35	Tela do <i>Praat</i> com o texto (bfamcv23-43-DIN) segmentado com a forma <i>ocê</i> : para melhor visualização foi inserido somente o trecho da unidade tonal	181

	analisada.....	
Figura 36	Tela do <i>Praat</i> com o texto (bfamcv23-43-DIN): foi evidenciada a sílaba “cê”.....	182
Figura 37	Tela do <i>Praat</i> com o texto (bfamcv21-43-FLA): foi evidenciada a sílaba fonética [‘se].....	184
Figura 38	Tela do <i>Praat</i> com o texto (bfamcv23-7-DIN): foi evidenciada a forma “ocê”.....	185
Figura 39	Tela do <i>Praat</i> com o texto (bfamdl03-35-LUZ): foi evidenciada a forma “ocê”.....	186
Figura 40	Tela do <i>Praat</i> com o texto (bfamdl30-68-REN): foi evidenciada a forma “ocê”.....	188
Figura 41	Tela do <i>Praat</i> com o texto (bfamdl19-73-AVI): foi evidenciada a forma “ocê”.....	189
Figura 42	Tela do <i>Praat</i> com o texto (bfamcv02-70-TER): foi evidenciada a forma “ocê”.....	190
Figura 43	Tela do <i>Praat</i> com o texto (bfamcv11-197-ONO): foi evidenciada a forma “ocê”.....	192
Figura 44	Tela do <i>Praat</i> com o texto (bfammn03-102-ALO): foi evidenciada a forma “ocê”.....	193
Figura 45	Tela do <i>Praat</i> com o texto (bfammn03-105-ALO): foi evidenciada a forma “ocê”.....	194
Figura 46	Tela do <i>Praat</i> com o texto (bfamcv19-49-MAE): foi evidenciada a forma “você”.....	196
Figura 47	Tela do <i>Praat</i> com o texto (bfamdl11-70-MAR): foi evidenciada a forma “você”.....	197
Figura 48	Tela do <i>Praat</i> com o texto (bfamdl14-25-CAR): foi evidenciada a forma “você”.....	199
Figura 49	Tela do <i>Praat</i> com o texto (bfammn05-63-CAR): foi evidenciada a sílaba [‘vse].....	200
Figura 50	Tela do <i>Praat</i> com o texto (bfammn12-104-JUN1): foi evidenciada a forma “você”.....	202
Figura 51	Tela do <i>Praat</i> com o texto (bfamdl11-126-MAR): foi evidenciada a forma “você”.....	204
Figura 52	Tela do <i>Praat</i> com o texto (bfammn05-1-CAR): foi evidenciada a forma “você”.....	205
Figura 53	Tela do <i>Praat</i> com o texto (bfammn07-146-NAN): foi evidenciada a forma “você”.....	207
Figura 54	Resultados dos testes de normalidade de [‘se] em relação ao enunciado em posição pré-verbal.....	209
Figura 55	Resultados dos testes de normalidade de [‘se] em relação ao enunciado em posição pós-verbal.....	209
Figura 56	Resultados dos testes de normalidade de [‘se] em relação à unidade tonal em posição pré-verbal.....	210
Figura 57	Resultados dos testes de normalidade de [‘se] em relação à unidade tonal em posição pós-verbal.....	210

Lista de Gráficos

Gráfico 1	Visualização das porcentagens de uso das formas pronominais de segunda pessoa encontradas no C-ORAL-BRASIL.....	119
Gráfico 2	Visualização das porcentagens de uso das formas pronominais de segunda pessoa, singulares e plurais, encontradas no C-ORAL-BRASIL.....	119
Gráfico 3	Visualização das porcentagens de uso das formas pronominais de segunda pessoa, singulares e plurais, excluídos os casos de interrupção do enunciado, encontradas no C-ORAL-BRASIL.....	120
Gráfico 4	Visualização das porcentagens de uso das formas pronominais de segunda pessoa, singulares e plurais, excluídos os casos de <i>retracting</i> , encontradas no C-ORAL-BRASIL.....	121
Gráfico 5	Visualização das porcentagens de uso das formas pronominais de segunda pessoa, singulares e plurais, excluídos os casos de interrupção e de <i>retracting</i> , encontradas no C-ORAL-BRASIL.....	122
Gráfico 6	Exposição sintética das ocorrências das formas pronominais de segunda pessoa divididas por posições, extraídas do C-ORAL-BRASIL.....	123
Gráfico 7	Visualização das porcentagens de uso das formas pronominais de segunda pessoa na posição pré-verbal.....	123
Gráfico 8	Visualização das porcentagens de uso das formas pronominais de segunda pessoa na posição pós-verbal.....	124
Gráfico 9	Subdivisão em percentual das formas pronominais de segunda pessoa na posição pós-verbal.....	124
Gráfico 10	Subdivisão das formas pronominais de segunda pessoa em isolamento.....	126
Gráfico 11	Subdivisão das formas pronominais de segunda pessoa em isolamento, em termos percentuais.....	126
Gráfico 12	Visualização das porcentagens de uso das formas pronominais de segunda pessoa <i>você(s)</i>	127
Gráfico 13	Subdivisão da forma <i>você(s)</i> na posição pós-verbal em termos percentuais.	128
Gráfico 14	Subdivisão da forma <i>você(s)</i> em isolamento, em termos percentuais.....	129
Gráfico 15	Visualização das porcentagens de uso das formas pronominais de segunda pessoa <i>ocê(s)</i>	130
Gráfico 16	Subdivisão da forma <i>ocê(s)</i> na posição pós-verbal em termos percentuais..	130
Gráfico 17	Subdivisão da forma <i>ocê(s)</i> em isolamento em unidade tonal, em termos percentuais.....	131
Gráfico 18	Visualização das porcentagens de uso das formas pronominais de segunda pessoa <i>cê(s)</i>	133
Gráfico 19	Visualização das porcentagens de uso das formas pronominais de segunda pessoa <i>cê(s)</i> em posição pré-verbal.....	133
Gráfico 20	Visualização das porcentagens de uso das formas pronominais de segunda pessoa <i>cê(s)</i> como objeto de preposição.....	134
Gráfico 21	Dados dos informantes: gênero em termos percentuais.....	137
Gráfico 22	Dados dos informantes: idade em termos absolutos.....	138
Gráfico 23	Dados dos informantes: escolaridade em termos absolutos.....	139
Gráfico 24	Dados dos informantes: origem em termos absolutos.....	139
Gráfico 25	Duração das sílabas fonéticas do texto (bfamcv18-187).....	145
Gráfico 26	Duração das sílabas fonéticas do texto (bpubdl01-90-91).....	146
Gráfico 27	Duração das sílabas fonéticas do texto (bfamcv13-33).....	147
Gráfico 28	Duração das sílabas fonéticas do texto (bfamcv03-247).....	148
Gráfico 29	Duração das sílabas fonéticas do texto (bpubdl11-241).....	149
Gráfico 30	Duração das sílabas fonéticas do texto (bpubcv04-61).....	150
Gráfico 31	Duração das sílabas fonéticas do texto (bfamcv31-187).....	153
Gráfico 32	Duração das sílabas fonéticas do texto (bfamcv32-105).....	154
Gráfico 33	Duração das sílabas fonéticas do texto (bfammn27-64).....	155
Gráfico 34	Duração das sílabas fonéticas do texto (bpubmn11-24).....	156
Gráfico 35	Duração das sílabas fonéticas do texto (bfamd105-185).....	157
Gráfico 36	Duração das sílabas fonéticas do texto (bfammn02-47).....	158
Gráfico 37	Duração das sílabas fonéticas do texto (bfammn03-115).....	159
Gráfico 38	Duração das sílabas fonéticas do texto (bfamcv19-27).....	160
Gráfico 39	Duração das sílabas fonéticas do texto (bfamd105-282).....	161
Gráfico 40	Duração das sílabas fonéticas do texto (bfamcv20-190).....	162

Gráfico 41	Porcentagem dos segmentos analisados divididos por formas e posições.....	163
Gráfico 42	Duração das sílabas fonéticas do texto (bfamcv02-335-RUT).....	166
Gráfico 43	Duração das sílabas fonéticas do texto (bfamcv18-208-HER).....	167
Gráfico 44	Duração das sílabas fonéticas do texto (bfamdl02-45-BAL).....	168
Gráfico 45	Duração das sílabas fonéticas do texto (bfamcv05-240-CEL).....	172
Gráfico 46	Duração das sílabas fonéticas do texto (bfamcv04-453-BRU).....	173
Gráfico 47	Duração das sílabas fonéticas do texto (bpubdl04-260-ELI).....	174
Gráfico 48	Duração das sílabas fonéticas do texto (bfamcv23-28-DIN).....	175
Gráfico 49	Duração das sílabas fonéticas do texto (bfamdl02-247-BAL).....	177
Gráfico 50	Duração das sílabas fonéticas do texto (bfammn29-6-JDL).....	178
Gráfico 51	Duração das sílabas fonéticas do texto (bfamcv21-43-FLA).....	184
Gráfico 52	Duração das sílabas fonéticas do texto (bfamcv23-7-DIN).....	185
Gráfico 53	Duração das sílabas fonéticas do texto (bfamdl03-35-LUZ).....	186
Gráfico 54	Duração das sílabas fonéticas do texto (bfamdl30-68-REN).....	188
Gráfico 55	Duração das sílabas fonéticas do texto (bfamdl19-73-AVI).....	189
Gráfico 56	Duração das sílabas fonéticas do texto (bfamcv02-70-TER).....	171
Gráfico 57	Duração das sílabas fonéticas do (bfammn03-102-ALO).....	193
Gráfico 58	Duração das sílabas fonéticas do (bfammn03-105-ALO).....	194
Gráfico 59	Duração das sílabas fonéticas do texto (bfamcv19-49-MAE).....	196
Gráfico 60	Duração das sílabas fonéticas do texto (bfamdl11-70-MAR).....	198
Gráfico 61	Duração das sílabas fonéticas do texto (bfamdl14-25-CAR).....	199
Gráfico 62	Duração das sílabas fonéticas do texto (bfammn05-63-CAR).....	201
Gráfico 63	Duração das sílabas fonéticas do texto (bfammn12-104-JUN1).....	203
Gráfico 64	Duração das sílabas fonéticas do texto (bfamdl11-126-MAR).....	205
Gráfico 65	Duração das sílabas fonéticas do texto (bfammn05-1-CAR).....	206
Gráfico 66	Duração das sílabas fonéticas do texto (bfammn07-146-NAN).....	208
Gráfico 67	Variação no uso das formas de sujeito clítico, simples e reduplicado, nos dois <i>corpora</i> , em valores normalizados.....	216
Gráfico 68	Variação no uso das formas de expressão do sujeito, nos dois <i>corpora</i> , em valores normalizados.....	217

Lista de Quadros

Quadro 1	Tabela das médias de intensidade e duração em Ciríaco <i>et al.</i> , 2004, p. 149.....	42
Quadro 2	Tabela de ocorrências do <i>corpus</i> de 2002 em PERES, 2006, p. 131.....	53
Quadro 3	Tabela de ocorrências do <i>corpus</i> de 1982 em PERES, 2006, p. 170.....	56
Quadro 4	Os pronomes pessoais sujeito em italiano.....	67
Quadro 5	Os pronomes pessoais sujeito no vernáculo florentino.....	68
Quadro 6	Estrutura do <i>corpus Stammerjohann</i> , com número de palavras e textos para cada campo da estrutura.....	110
Quadro 7	Estrutura do <i>corpus Corpus per il Confronto Diacronico</i> , com número de palavras e textos para cada campo da estrutura.....	112
Quadro 8	Estrutura do <i>Corpus de Florentinos</i> , com número de palavras e textos para cada campo da estrutura.....	113
Quadro 9	Comparação em termos percentuais entre o <i>Corpus per il Confronto Diacrônico</i> e o <i>Corpus de Florentinos</i> , para cada campo da estrutura.....	115

1. INTRODUÇÃO

1.1. Justificativa

Este trabalho quer apresentar respostas a questões ainda não solucionadas de maneira satisfatória na literatura, ou seja: 1) qual é o real estatuto das formas reduzidas do pronome *cê* no Português do Brasil (doravante PB); 2) quais são os efeitos de quarenta anos de pressão do italiano padrão sobre o uso, próprio do vernáculo florentino, das formas de sujeito pronominal *gl' / e / la / l' .*

Enquanto as respostas à primeira pergunta são muito debatidas na literatura, aquelas para a segunda foram rediscutidas somente em tempos recentes, e sem estudos sobre dados empíricos. O que se diz é que as formas pronominais reduzidas diminuíram ou até desapareceram (GALLI DE' PARATESI, 1984; SOBRERO, 1997; BERRUTO, 1997, 2005; BINAZZI, 1997, 1998, 2007; CARDINALETTI e MUNARO, 2009). No entanto, essas afirmações não foram verificadas em dados quantitativos confiáveis. A ausência deste tipo de estudos justifica, portanto, um trabalho diacrônico sobre os pronomes sujeito no florentino.

Ao contrário, muitas são as hipóteses sobre o estatuto (clítico, fraco ou de outra natureza) do pronome *cê* em PB. Contudo, não foram examinados dados do tipo daqueles considerados neste trabalho, ou seja, dados extraídos de um *corpus* de fala espontânea (publicado e verificável) que permita também uma análise fonética. Além disso, não é de nosso conhecimento um trabalho sobre as formas *cê, ocê e você* que tenha partido, assim como foi feito aqui, da análise prosódica (ou seja, da averiguação da tonicidade ou atonicidade da forma) para depois entender quais posições e funções são compatíveis com o estatuto átono ou tônico desta forma. A metodologia original, como será explicada adiante, justifica, portanto, um novo trabalho sobre esse tema.

A presença, em um único trabalho, de dois objetos de pesquisa aparentemente tão diferentes se justifica pelas seguintes razões:

1. o estatuto categorial, ou seja, a qual categoria gramatical pertence a forma *cê* no PB, é objeto de grandes dúvidas por parte dos estudiosos. Frequentemente o sistema pronominal de segunda pessoa do singular/plural do PB é descrito como em curso de gramaticalização, mas com várias ressalvas. Por outro lado, nesta pesquisa, os pronomes sujeito do florentino são apresentados como exemplo de um caso de plena gramaticalização, no qual, a partir da forma segmental é possível determinar qual o estatuto, tônico ou clítico, e o comportamento, sujeito ou objeto, de tais pronomes.

2. Poder comparar dois casos em que uma forma pronominal reduzida apresenta características muito diferentes (e talvez opostas) quanto à relação entre forma segmental e estatuto funcional pode esclarecer, a nosso ver, tanto metodologicamente quanto do ponto de vista teórico, a análise de ambas as formas, e principalmente aquela do pronome *cê*;

3. os dois idiomas (PB e florentino) dispõem agora de *corpora* comparáveis (também diacronicamente no caso do florentino) de fala espontânea. Essa coincidência, ainda rara nos estudos linguísticos, abre novas possibilidades sob o aspecto metodológico, tanto para a comparação interlínguística quanto para a comparação diacrônica na mesma língua.

1.2. Objetivos

Os objetivos do trabalho são:

- a) para o PB, o estudo das diferentes formas pronominais (plenas e reduzidas) de segunda pessoa *você(s)/ocê(s)/cê(s)* em seus diferentes usos e contextos, a partir de uma definição de seu estatuto de tonicidade ou atonicidade com base em parâmetros prosódicos;
- b) para o vernáculo florentino, o exame da situação de alguns pronomes clíticos sujeito nos anos de 1960 e de 2000, ou seja, com um intervalo de cerca de quarenta anos, que, como mostraremos, são fundamentais na história linguística italiana¹. Em específico, trataremos dos clíticos de terceira pessoa do singular, *gl' / e/ la/ l'* em concomitância com alguns dos verbos de mais alta frequência da língua italiana, como: *essere* (ser), *fare* (fazer), *andare* (ir) e *venire* (vir) no presente do indicativo.

Como já dito, a situação dessas formas pronominais nas duas línguas é muito diferente, permitindo comparar o funcionamento daquelas completamente gramaticalizadas, como é o caso do florentino, com aquelas frequentemente descritas (e com muitas conclusões diferentes) como em curso de gramaticalização, como é o caso do PB.

Essa possibilidade é oferecida pela presença de *corpora* comparáveis: o C-ORAL-BRASIL (RASO & MELLO, 2012) para o PB, que apresenta a variedade mineira; o C-ORAL-ROM (CRESTI & MONEGLIA, 2005) para o italiano, na variedade do florentino contemporâneo; o *Corpus Stammerjohan* (1970) remasterizado pelo laboratório LABLITA e

¹ Tais anos, como será explicado adiante, são fundamentais para a difusão do italiano padrão, especialmente através da escolarização de massa e da mídia, sobretudo aquela televisiva, em oposição aos dialetos (ou no caso do florentino, ao vernáculo local).

o *Corpus per il Confronto Diacronico* (SCARANO, 2005), que cria dois *subcorpora* comparáveis a partir do *corpus Stammerjohan* e do C-ORAL-ROM².

O ponto de vista do qual se olham os fenômenos também é original. Isso é especialmente evidente para a análise do PB. De fato, no passado, as formas *você/ocê/cê* foram analisadas com base em uma das perspectivas seguintes:

a) partindo da análise das funções e posições que, segundo métodos introspectivos, pareciam possíveis para cada forma e, a partir disso, concluir sobre o estatuto átono ou tônico das formas;

b) analisando *corpora* não verificáveis (porque não publicados ou não descritos) e sempre caracterizados por textos de duas únicas tipologias: narrativas e entrevistas;

c) analisando sentença preparadas em laboratório.

O presente trabalho, ao contrário, parte da análise suprasegmental das formas. Em primeiro lugar, busca-se definir quando estas são átonas e quando são tônicas; somente depois são observadas as posições e as funções em que as formas átonas e as formas tônicas podem se encontrar. Além disso, o trabalho se baseia em *corpora* de fala espontânea com ampla variação diafásica, capazes, portanto, de documentar o uso real da fala como um todo, e não somente de uma fala muito específica como aquela monológica ou de entrevistas. Por fim, todos os dados analisados são publicados e disponíveis para acesso da comunidade científica.

Especificando melhor as características dos *corpora* usados, no momento é importante ressaltar que:

- todos os *corpora* são balanceados, ou seja, construídos de maneira a serem comparáveis entre si, e representativos das respectivas diatopias, seguindo a mesma arquitetura baseada também na comparabilidade diafásica;

- os *corpora* possuem a transcrição semiortográfica das gravações e a transcrição alinhada ao áudio, o que permite a análise acústica dos enunciados de interesse para a pesquisa;

- a unicidade de dois *corpora* do vernáculo florentino gravados em períodos diversos também possibilita um trabalho de tipo diacrônico, do qual não possuímos outros exemplos.

Dissemos que os *corpora* são representativos da fala espontânea. Por fala espontânea entendemos a fala que é planejada ao mesmo tempo em que é realizada (NENCIONI, 1983). A “fala falada”, aquela realmente espontânea, é caracterizada por uma grande acionalidade e

² Sobre as características específicas dos *corpora* consultados e sobre a arquitetura dos *corpora* da família C-ORAL-ROM será discutido amplamente adiante.

por valores ilocutivos diversificados, em que a tipologia ilocucionária muda continuamente. Isso acontece porque no monólogo, ou mesmo em um diálogo de tipo narrativo, o falante desenvolve uma linha discursiva, que pode até ser eventualmente interrompida, mas continua em sua macroconstrução semântica. O diálogo do tipo entrevista, portanto, não pode ser considerado um texto verdadeiramente dialógico. Na fala acional, quando os interlocutores estão fazendo algo enquanto falam, a construção do texto é totalmente imprevisível: o planejamento da fala é feito no momento da interação, portanto depende não somente do macropiano de um falante, como também do que está acontecendo no ambiente. Por isso sua sintaxe é peculiar e as estruturas linguísticas encontradas nas gravações são muito diferentes daquelas às quais estamos habituados nos exemplos fictícios e de *corpora* sem variação diafásica geralmente utilizados como exemplificações. Acreditamos que este tipo de fala seja mais reveladora do comportamento de formas como, por exemplo, o uso do *cê(s)* e do *ocê(s)*, porque a diafasia diversificada e a espontaneidade propiciam uma maior chance de produção de uma linguagem informal.

Dos *corpora* utilizados, o único que não persegue em sua totalidade a linha de plena acionalidade é o *Corpus Stammerjohann*, pelo fato de ter sido gravado em uma época em que, seja pela tecnologia, seja pelo tipo de arquitetura de compilação seguida, essa não era a preocupação primária³. Mas o *Corpus per il Confronto Diacronico* tenta recuperar o máximo de acionalidade presente no *Corpus Stammerjohann*.

Os objetivos específicos desta pesquisa podem ser divididos em dois grupos. No primeiro:

- a. quanto ao PB:
 - i. verificar a frequência de ocorrência das formas *você(s)/ocê(s)/cê(s)* utilizadas no *corpus* e seu contexto de uso;
 - ii. avaliar, através da duração e de sua normalização⁴, a tonicidade e atonicidade das várias formas;
 - iii. verificar a distribuição das várias formas distinguindo sua tonicidade ou atonicidade, tendo como fundamento para essa distinção sua forma suprasegmental. Somente a partir desta distinção serão atribuídas funções às várias formas;

³ No parágrafo 3.2.1 detalharemos as especificações do *corpus* e as características de sua coleta.

⁴ A duração bruta de um segmento acústico não é um bom parâmetro de análise, em específico quando se deseja compará-la com outros segmentos. O processo de normalização elimina justamente os condicionamentos intrínsecos dos segmentos que impediriam uma comparação confiável.

- iv. investigar as diferenças de uso entre falantes de diastratias diferentes que possam ser relevantes.
- b. no vernáculo florentino:
- i. verificar a frequência de ocorrência das formas utilizadas como expressão de sujeito de terceira pessoa do singular, masculino e feminino, *gl' e' la/ l'*, em conjunto com os verbos *essere* (ser), *fare* (fazer), *andare* (ir) e *venire* (vir) no presente do indicativo, nos dois *corpora* comparáveis;
 - ii. implementar, nos dois *corpora*, uma análise percentual sobre o uso das formas reduzidas;
 - iii. comparar, nos dois *corpora*, a frequência do uso das formas clíticas, e indagar sobre a utilização de formas plenas, uso de outros pronomes, de sujeitos nominais ou forma zero;
 - iv. investigar as diferenças de uso entre falantes de diastratias diferentes que possam ser relevantes.

1.3. Hipótese de trabalho

As hipóteses que são a base desta dupla pesquisa são:

- I. Para o PB:
 - a. não é possível afirmar que para uma determinada forma segmental corresponda necessariamente uma única forma prosódica (tônica ou clítica);
 - b. tonicidade ou cliticidade são independentes da forma segmental, como demonstram os dados recolhidos em análise acústica. Portanto, somente identificando o estatuto prosódico das ocorrências das diferentes formas segmentais é possível entender o comportamento dos pronomes.
- II. Para o vernáculo florentino:
 - a. os trinta a quarenta anos que separam as gravações dos dois *corpora* representam anos de forte pressão do italiano padrão sobre os dialetos. É, portanto, plausível que seja detectável uma diminuição do fenômeno de expressão do sujeito, mais especificamente do sujeito clítico. Deve-se, contudo, verificar se e em que medida isso realmente aconteceu;
 - b. espera-se encontrar uma redução das ocorrências de sujeito clítico e uma reconfiguração da expressão do sujeito com formas plenas ou nulas.

2. REVISÃO DA LITERATURA

2.1. História do pronome *você*

Nesta seção será traçado brevemente o percurso histórico das formas *você(s)/ocê(s)/cê(s)* desde sua origem até os dias atuais.

Em Portugal, no início da monarquia, não havia uma grande distinção na forma de tratamento entre o rei e os outros nobres. Somente a partir do século XIII há uma certa diferenciação entre o monarca e as outras classes, sendo ele tratado pela forma *vós*. Outras formas surgiram e *vós* foi, aos poucos, sendo substituído por *Vossa + Nome*. Não há consenso sobre o período exato em que surge essa estrutura, como é o caso de *Vossa Mercê*⁵. De qualquer forma, é de se esclarecer que o tratamento ao rei era indireto, sendo feita referência à sua generosidade e poder de concessão de favores, graças, justiça e proteção.

Inicialmente a forma *Vossa Mercê* era a mais utilizada para se dirigir ao monarca, mas com o passar do tempo surgiram outras formas como *Vossa Senhoria*, *Vossa Alteza* e enfim *Vossa Majestade* (LUZ, 1956; SAID ALI, 1965; COOK, 1997). Por este motivo, o uso de *Vossa Mercê* para tratamento do monarca começou a diminuir, sendo utilizado para outras figuras da aristocracia. Em seguida, essa forma passou a ser usada como tratamento para a burguesia, baixando cada vez mais na escala social, até ser utilizada como tratamento de respeito para qualquer cidadão que não pudesse ser tratado por formas mais nobres, como *Vossa Senhoria*, ou mais íntimas, como *tu* ou *vós*.

A forma *Vossa Mercê* foi, portanto, suplantando *vós*, que cai em desuso como referência a um único interlocutor até praticamente desaparecer no fim do século XIX (LOPES, 2003; LOPES e DUARTE, 2003). Devido à sua frequência de uso, *Vossa Mercê* começou a sofrer uma série de reduções de segmentos e sílabas átonas, em um processo geral de redução fonética. Isso deu origem a diferentes variantes como: *cê*, *mecê*, *mincê*, *ocê*, *oncê*, *sucê*, *suncê*, *vacê*, *vainicê*, *vancê*, *vansmincê*, *vassuncê*, *voncê*, *vosmecê*, *vossemecê*, *vosmincê*, *vossuncê*, *ucê* (NASCENTES, 1956). A forma que se fixou na língua portuguesa foi *você*, transformada em pronome de tratamento. Algumas das outras ainda se encontram nas regiões rurais brasileiras e portuguesas⁶.

Dentro desse processo de mudança de *Vossa Mercê*>*você*, houve um rearranjo da estrutura da língua, em específico no sistema pronominal⁷. A forma *vós* cai em desuso e *você*

⁵ Vejam-se, a esse respeito, Luz (1956), Cintra (1972), Faraco (1996), Said Ali (1965), Nascentes (1956), entre outros.

⁶ A esse respeito: Amaral (1976), Said Ali (1969), Coelho (1999), Oliveira (1996), Basto (1931) e Cook (1997).

⁷ Veja-se, a esse respeito, Lopes (2003).

é utilizada em português na sua forma plural. Contudo, tendo sido originada de um pronome de tratamento, *você* também faz concordância com a terceira pessoa, havendo uma combinação em um só pronome de 2ª e 3ª pessoas. Isso acabou por alterar também o uso dos possessivos, determinando a utilização cada vez mais frequente da forma *deles* para evitar ambiguidade. Outro efeito da introdução da forma *você* no paradigma pronominal foi a perda da distinção de pessoa através das desinências verbais e, por consequência, uma necessidade cada vez maior do uso do sujeito.

A partir do início do século XX a forma *você(s)* apresenta uma distribuição mais homogênea, ocupando a posição de sujeito pré-verbal, e expressa além de referência definida, também referência indefinida (LOPES, 2002; 2003). Outro aspecto a ser evidenciado é a difusão da forma entre as camadas mais populares e seu uso nas relações assimétricas de superior para inferior. É de se esclarecer que em Portugal essa forma é, hoje em dia, bastante estigmatizada.

A situação linguística descrita acima foi trazida para o Brasil à época da colonização, no século XVI. Naquele tempo, o uso de *Vossa Mercê* e *vós* já estava em declínio e a simplificação fonética de *Vossa Mercê*, nas variantes vistas, estava avançada. Os colonizadores vinham de todas as partes de Portugal, mas a diferenciação linguística foi anulada em solo brasileiro. Como não havia nem escolas, nem imprensa, e a educação ficava somente a cargo dos jesuítas, a mudança linguística ocorreu de forma muito mais rápida que em Portugal⁸. No início do século XX o uso de *você* é quase predominante.

Nas últimas décadas há um amplo uso da forma *você*, alternada com o *tu* que, segundo alguns autores, está em expansão em algumas regiões do Brasil⁹.

As formas *ocê* e *cê* no Brasil são pouco atestadas nos textos escritos. Contudo, já aparecem nas primeiras décadas do século XX (AMARAL, 1976; SALLES, 2001; NASCENTES, 1956), sendo registradas pela primeira vez em 1920 como variantes de *Vossa Mercê* (AMARAL, 1976).

Na próxima seção serão discutidas as abordagens sobre os usos e distribuições atuais destas formas.

2.2. Questões abordadas nas propostas

As propostas que serão discutidas abaixo apresentam afirmações muitas vezes divergentes. Como as hipóteses e conclusões de cada autor serão apresentadas em detalhe,

⁸ Estudos sobre este assunto encontram-se em: Menón (2000), Salles (2001), Lopes e Duarte (2003; 2004).

⁹ Veja-se a discussão de Oliveira (2000) e Paredes Silva (2003).

introduziremos aqui algumas contradições que o leitor encontrará, e que são relevantes para as finalidades de nossa pesquisa.

Em primeiro lugar, há a questão da definição da forma *cê*. Como se verá adiante, a discussão é ampla, pois muitos autores buscam analisar a forma *cê* tentando, antes de mais nada, defini-la. Vitral (1996), Ramos (1997), Vitral e Ramos (2006; 2008), Vitral, Ramos e Souza (2007) afirmam que essa partícula seria um clítico ou se encontraria em processo de cliticização, estando na etapa de formação de expletivos¹⁰. Ciríaco, Vitral e Reis (2004) afirmam terem confirmado a cliticidade de *cê* através de testes acústicos em laboratório. Já segundo Petersen (2008), a forma *cê* seria um pronome fraco, segundo a teoria da tripartição de Cardinaletti e Starke (1994)¹¹. Por outro lado, Barbosa L. P. (2005) diz demonstrar que *cê* seria uma palavra plena. Othero (2013) acredita, por sua parte, que *cê* seja uma nova forma pronominal, nem clítica, nem fraca. A proposta de Peres (2006) é aquela que mais se aproxima da nossa, pois ao invés de tentar formular definições, observa os dados de que dispõe sem tentar encaixá-los em uma teoria.

2.3. Gramaticalização e cliticização

Nos próximos parágrafos serão revistas as contribuições que obtiveram maior impacto e foram alvo de debates sob o ponto de vista da hipótese da gramaticalização.

Vitral (1996) correlaciona a forma *cê* com a noção de gramaticalização¹². Em uma análise sintática sobre a distribuição dessa forma nas diferentes posições, o estudioso afirma que a forma *cê* não pode ocorrer em vários dos ambientes em que as formas *você* e *ocê* são encontradas, isto é:

- a) como sujeito em posição pós-verbal;
- b) como objeto de verbo ou de preposição;
- c) em posição de tópico;
- d) e focalizado.

Segundo Vitral (1996), a única posição de *cê* relacionada no *corpus* por ele consultado é aquela pré-verbal.

As afirmações de Vitral (1996) baseiam-se em dados de Dutra (1991). Tal estudo não é acessível à comunidade científica, assim como seus dados, pois trata-se de um manuscrito de

¹⁰ Adiante será discutida esta questão.

¹¹ Adiante será discutida a proposta.

¹² O conceito de gramaticalização a que se refere Vitral é aquele apresentado em Hopper & Traugott (1993).

“Folha de exemplos”¹³, dos quais não foi possível obter maiores informações como: tipo de coleta, informantes, data e local da coleta, se se trata de gravações transcritas ou outro tipo de material. É importante ressaltar que, não somente em nosso estudo aqui apresentado, mas em outros que serão discutidos adiante, as ocorrências nas diferentes posições enumeradas acima foram verificadas em vários *corpora*, ou admitidas como possíveis por diferentes estudiosos.

Segundo Vitral (1996), a forma *cê* representa uma etapa, aquela da cliticização, do processo de gramaticalização da forma *Vossa Mercê*. O estudioso propõe, portanto, o seguinte processo, que ainda estaria em curso (VITRAL, 1996, p. 119):

Item com significado lexical: Vossa Mercê > item gramatical: você > clítico: cê > afixo flexional.

A forma *Vosmecê* é considerada por Vitral (1996) como um estágio intermediário entre a primeira e a segunda etapas, enquanto a forma *ocê* poderia ser um estágio intermediário entre a segunda e a terceira etapas.

Outro estudo basilar para a análise das formas pronominais de segunda pessoa é aquele de Ramos (1997). Com o duplo intuito de testar o processo de cliticização de *cê* e de descrever em uma análise sociolinguística o uso de *cê*, *ocê* e *você*, a estudiosa verifica um conjunto de 342 dados extraídos de entrevistas com falantes de Belo Horizonte. Os critérios de investigação das formas dividem-se entre internos (função sintática; focalização; tipo de referência; tipo de oração em que aparecem; posição em relação ao verbo e topicalização) e externos (gênero e idade). Os resultados obtidos são:

- a) a posição de sujeito é a mais recorrente nas três formas, com 94,4% das ocorrências;
- b) há uma diminuição no uso da forma *você*, segundo a faixa etária; a forma *ocê* teve um pequeno aumento; a forma *cê* é a favorita em todas as faixas etárias mas sem aumento;
- c) a forma *cê* ocorre em 88% dos casos nas construções com topicalização;
- d) a forma *cê* aparece em posição não-contígua ao verbo somente com advérbios curtos como *não*, *já*, *só*;

¹³ Segundo quanto consta na bibliografia do autor.

- e) há preferência pela forma *você* quando a oração é interrogativa simples; no caso de orações declarativas e interrogativas com estrutura “que que” a forma mais utilizada é *cê*;
- f) não há ocorrências de *cê* focalizado;
- g) há um aumento na utilização de referência indefinida, pois ela ocorre mais entre os jovens, seguidos dos medianos, e menos nos velhos¹⁴;
- h) no caso de referência definida, jovens e medianos utilizam mais a forma *você* (67% das ocorrências); no caso de referência indefinida a forma por eles preferida é *cê* (56% das ocorrências). Já *você* é a forma mais utilizada pelos velhos no caso de referência indefinida (53% das ocorrências) e *cê* quando é definida (65% das ocorrências);
- i) segundo o arcabouço teórico seguido, a hipótese da cliticização da forma *cê* é confirmada.

Em um trabalho conjunto, Vitral e Ramos (1999) rediscutem a gramaticalização da forma *você*. Baseando-se nas teorias de Hopper e Traugott (1991), segundo os quais o processo de gramaticalização passa de um item com significado lexical para um item gramatical, seguido de um clítico e terminando em um afixo, os estudiosos afirmam examinar as mudanças da forma *você* pelas primeiras três etapas. Citando Hopper e Traugott (1991), os autores alegam perda de conteúdo semântico no processo de gramaticalização. Para explicar tal argumento são revistos por eles os trabalhos anteriores de Vitral (1996) e especialmente aquele de Ramos (1997), dos quais são retirados os dados que indicariam haver tendência à especialização das formas *cê* e *você*. Como visto nesses estudos, os jovens utilizam *cê* com referência menos definida enquanto os velhos com referência definida. O processo de especialização indicaria, segundo os estudiosos, etapas da gramaticalização. A perda de autonomia lexical é por eles correlacionada à redução fonológica do item cliticizado, passando de duas sílabas no caso de *você* para uma no caso de *cê*. A perda de conteúdo semântico, por outro lado, é correlacionada à referência não definida como no exemplo abaixo (VITRAL e RAMOS, 1999, p. 58):

(1) *Cê* pode visitar sempre esse local. (= Pode-se visitar sempre esse local.)

¹⁴ A terminologia referente às faixas etárias é aquela adotada por Ramos.

Segundo os estudiosos, perda de substância fonética e perda de conteúdo semântico induzem à falta de referencialidade de uma das formas: no caso, aquela clítica. Isso levaria necessariamente o processo de gramaticalização à sua última etapa, a formação de expletivos¹⁵. Trata-se de itens não-referenciais, ou seja, desprovidos de traços semânticos em sua matriz lexical. Os autores explicam que um nome é associável a um determinado “ser”¹⁶ fora do enunciado. O mesmo não acontece com os pronomes. Para analisar seu comportamento semântico, se faz necessário conhecer o ato de fala de que ele participa. Os pronomes pessoais são termos carentes de referência virtual, pois não é possível a eles associar algo fora do enunciado, mas caracterizados por referência real no enunciado em que aparecem. Os estudiosos afirmam, portanto, que, quando se fala de perda de conteúdo semântico, se está falando da perda de traços semânticos que caracterizam a referência virtual. Termos que são usados não referencialmente são ditos expletivos. Segundo Vitral e Ramos o português não possui itens expletivos. Contudo, seguindo Duarte (1997), há usos expletivos de formas pronominais: é o caso do *você* em posição não temática, como nos exemplos abaixo de Vitral e Ramos (1999, p. 61):

- (2) Em Kioto *você* tem aquela confusão nas ruas.
- (3) Em Buenos Aires *você* tem confeitarias.

A conclusão dos autores é de que, contrariamente a quanto suposto, não é a forma *cê*, a mais reduzida e clítica, que está sendo usada como expletivo, mas *você*. Isso configura um problema na corrente teoria da gramaticalização, que supõe uma correlação entre perda de substância fonética e perda de conteúdo semântico.

Em uma obra mais recente, Vitral e Ramos (2006) retomam os estudos anteriores em torno do conceito de gramaticalização, que envolvem, além das formas pronominais de segunda pessoa, também aquelas de terceira pessoa e itens negativos como *não* > *num*. O intuito da obra é analisar os fenômenos de cliticização no PB, inclusive através de uma discussão sobre a interpolação do pronome *se*. Para os objetivos da presente pesquisa, a parte mais interessante refere-se à abordagem do conceito de gramaticalização aplicado aos pronomes de segunda pessoa.

Os autores novamente propõem que as formas *você*, *ocê* e *cê* se encontrem em um processo de gramaticalização: o pronome *cê* estaria na fase final do processo de cliticização

¹⁵ Para esclarecer: no inglês, termos expletivos são *it* e *there*, no francês, *il*.

¹⁶ Termo adotado pelos autores.

iniciado com a forma *Vossa Mercê*. Vitral e Ramos (2006) apontam para alguns problemas na afirmação de que a forma *cê* seja um novo clítico como:

- a) a diminuição da frequência de clíticos em geral no PB;
- b) a possibilidade de aparecimento da forma *cê* em contextos vedados aos clíticos¹⁷ como em Vitral e Ramos (2006, p. 35):

(4) *Cê* foi o culpado.

Contudo, solucionam a questão com as seguintes observações:

- a) apesar de estudos quantitativos mostrarem que a frequência dos clíticos está diminuindo no PB falado, a inserção do clítico *se* indeterminador (*se diz isso, se pensa aquilo*) estaria em expansão. Isto, segundo os autores, corresponderia ao uso também de *cê* como fator de indeterminação¹⁸. Portanto, concluem, não é o total dos clíticos que está diminuindo, mas a frequência dos clíticos acusativos que estaria em processo de queda muito forte no PB falado;
- b) no PB parece haver uma tendência ao preenchimento da posição de sujeito;
- c) a forma *cê* é agramatical nos mesmos contextos e posições em que os outros clíticos não são aceitos. Segundo os autores esta agramaticalidade é justificada pela natureza clítica da partícula. Vitral e Ramos (2006), utilizando os exemplos de Dutra (1991), afirmam que no PB falado não é encontrada a forma clítica nas seguintes situações:

I. topicalizado como em Dutra, 1991, *apud*¹⁹ Vitral e Ramos, 2006, p. 32:

(5) **Cê* ele não viu.

II. modificado por advérbio, como em Dutra, 1991, *apud* Vitral e Ramos, 2006, p. 32:

(6) **Só cê* tava mentindo.

¹⁷ No caso abaixo, por exemplo, o clítico está em início de sentença, fenômeno que segundo os autores é restrito a certos contextos.

¹⁸ O exemplo (1) citado anteriormente é característico do fenômeno do uso de *cê/se* como fator de indeterminação. Contudo os autores não citam outros exemplos deste uso.

¹⁹ Como foi dito anteriormente, não foi possível obter os dados originais de Dutra (1991).

III. sozinho, como resposta a uma pergunta, como em Dutra, 1991, *apud* Vitral e Ramos, 2006, p. 32:

(7) – Quem vai sair?
– *Cê.

IV. como complemento de preposição, como em Dutra, 1991, *apud* Vitral e Ramos, 2006, p. 32:

(8) Eu falei **pra cê*...

V. coordenado com uma forma tônica, como em Dutra, 1991, *apud* Vitral e Ramos, 2006, p. 33:

(9) *Ele e *cê* podem votar contra.

Gostaríamos de sublinhar que nos próximos parágrafos serão citados exemplos encontrados não somente em nosso estudo, mas também em outros que utilizam *corpora* de língua falada, em que a forma *cê* ocorre justamente como complemento de preposição ou em coordenação, o que os autores negam categoricamente em sua obra.

No capítulo 3 os autores retomam uma análise quantitativa anterior das formas *você*, *ocê* e *cê*. Os dados são os mesmos utilizados em Ramos (1997), compostos por entrevistas realizadas com 12 informantes de Belo Horizonte de três faixas etárias, para um total de 342 ocorrências das formas estudadas. Os critérios adotados são: fatores sociais (faixa etária) e distribuição sintática. A análise visa verificar se e em quais casos as três formas coexistem em diacronia, e se podem ocorrer nos mesmos contextos, ou se há especialização de um ou outro item. O objetivo é verificar, portanto, quais funções sintáticas as formas assumem em um determinado estágio da gramaticalização. Os resultados indicariam que *cê* estaria se especializando na função de sujeito sintático. Retomando estudos prévios, os autores articulam uma série de argumentações que atestariam a cliticidade de *cê* como clítico fonológico, pois segue a sequência [especificador + núcleo lexicalizado + clítico fonológico], e não sintático. De fato:

I. haveria uma tendência no uso de *cê* como item preferido para indicar entidade com referência [-específica], como abaixo, Vitral e Ramos, 2006, p. 42:

(10) *Cê* pode visitar sempre esse local.

- II. a ocorrência da forma *cê* como adjacente ao complementizador, lexicalmente realizado, em interrogativas, evidenciaria sua falta de acento no nível fonológico, como em Vitral e Ramos, 2006, p. 48:

(11) Como que *cê* acha que era o namoro?

- III. o fato de que, na forma *cê*, em 84% dos casos analisados pelos autores, a posição favorecida seria contígua ao verbo²⁰. Nos casos de posição não contígua os itens favoritos são: *já* (7,4%), *só* (3,8%), *sempre* (3,8%) e *não* (85%), como em Vitral e Ramos, 2006, p. 46:

(12) *Cê* não paga nada.

O capítulo 4 concentra-se sobre as evidências fonéticas da realização da forma *cê*. É descrita a análise acústica de orações que apresentam as formas *se*, *cê* e *Zé*, em três contextos distintos²¹:

- I. entre dois segmentos átonos:

- A. (1) Como *cê* falou que a gente devia fazer?
 (2) Como *Zé* falou que era mesmo?
 (3) Como *se* falou que a gente devia pronunciar?

- II. em início de sentença:

- B. (4) *Cê* sabe de toda a história.
 (5) *Zé* sabe de toda a história.
 (6) *Se* sabe de toda a história.

²⁰ Devemos ressaltar que estas porcentagens referem-se somente à forma *cê*. Faltando aquelas sobre a distribuição das formas *ocê* e *você* em relação ao verbo, o leitor não possui os meios necessários para saber se esses números são relevantes.

²¹ Todos os exemplos são extraídos de Vitral e Ramos, 2006, p. 55.

III. entre dois segmentos tônicos:

- C. (7) Quem *cê* disse que viajou junto?
 (8) Quem *Zé* disse que viajou junto?
 (9) Quem *se* disse que foi comigo?

Foram observados os parâmetros de intensidade (I, em decibéis, dB) e duração (D, em milésimos de segundo, *ms*).

Cinco informantes efetuaram duas leituras das sentenças. Foi extraída a média de D e I das leituras, como segue:

Exemplos de Vitral e Ramos, 2006, p. 56-57:

(13) *Como cê falou que a gente devia fazer?*

I: 36 dB D: 137 ms²²

(14) *Como Zé falou que era mesmo?*

I: 41 dB D: 170 ms

(15) *Como se falou que a gente devia pronunciar?*

I: 33 dB D: 137 ms

(16) *Cê sabe de toda a história.*

I: 20dB D: 122 ms

(17) *Zé sabe de toda a história.*

I: 27 dB D: 217 ms

(18) *Se sabe de toda a história.*

I: 18 dB D: 117 ms

(19) *Quem cê disse que viajou junto?*

²² Aqui, como nos exemplos seguintes apresentados, estamos adequando a medida exata em *ms* pois no texto original está marcada após vírgula (0,137ms).

I: 23 dB D: 156ms

(20) *Quem Zé disse que viajou junto?*

I: 26 dB D: 196 ms

(21) *Quem se disse que foi comigo?*

I: 20 dB D: 134 ms

A média dos valores encontrados nos três ambientes é a seguinte:

Vitral e Ramos, 2006, p. 57

(22) *SE: I: 23,6 dB*

D: 129 ms

(23) *CÊ: I: 26,3 dB*

D: 138 ms

(24) *ZÉ: I: 31,3 dB*

D: 194 ms

Os autores concluem que *cê* assemelha-se à forma *se*, confirmando a hipótese de sua cliticidade.

É reconsiderado, enfim, pelos autores o estudo de Cardinaletti e Starke (1994)²³. Vitral e Ramos (2006) afirmam que a perspectiva de Cardinaletti e Starke (1994), de uma distinção em pronomes fracos, fortes e clíticos, faz parte de um processo maior de gramaticalização, sugerindo a inclusão de um novo item, o *simple clitic*, segundo a terminologia de Zwicky (1977). Segundo Vitral e Ramos (2006) a forma *cê* poderia ser inserida na classificação apontada por Cardinaletti e Starke (1994) como um clítico, estruturalmente em projeção máxima²⁴, ou como uma forma fraca. Mas seu parecer é que esta terminologia “não tem

²³ Segundo Cardinaletti e Starke, há três tipos de pronomes: fortes, fracos e clíticos, com domínios sintáticos distintos. Segundo essa perspectiva, a classificação se deveria a uma complexa interseção de um conjunto de propriedades sintáticas, morfológicas, semânticas e fonéticas. Clíticos e fracos se distinguiriam pelo fato de os clíticos poderem ser adjacentes ao verbo, o que não é possível aos pronomes fracos.

²⁴ A Teoria X-barras revê as regras de movimento, tentando identificar os princípios comuns a todas as línguas. Sem entrar em detalhes, se supõe que todo sintagma seja a projeção máxima de um núcleo, sendo X o sintagma, que pode ser nominal, verbal, preposicional e adjetival. A esse respeito: Chomsky (1970); Jackendoff (1977).

nenhum estatuto teórico explicativo” (Vital e Ramos, 2006, p. 69). pois seria mais proveitoso incluir a análise na compreensão geral do processo de cliticização das formas pronominais²⁵.

O capítulo 6 do livro faz uma revisão das análises até aquele momento descritas para propor uma nova etapa no processo de gramaticalização que leva do item com significado lexical *Vossa Mercê* até a forma *cê*. Como visto em estudos anteriores, a hipótese dos autores aponta para o reconhecimento de uma nova etapa final: aquela da formação dos expletivos. O processo de explicação segue aquele visto anteriormente e as conclusões são de que é necessário aceitar essa nova fase do processo de gramaticalização.

Em outro trabalho, Ramos, Vital e Souza (2007) analisam a preferência pela posição de sujeito por parte dos pronomes reduzidos. Em diacronia, essa posição era ocupada pelas formas de tratamento *Vossa Mercê* assim como *Vossa Senhoria* e *O senhor*. A questão que os autores debatem é o motivo pelo qual a posição de sujeito seria sede dos processos de cliticização. A proposta deles está baseada em uma análise de tipo sintática, com orientação teórica gerativista: haveria uma dependência sintática entre a posição de sujeito e o núcleo de uma categoria, seja funcional, seja lexical. Os autores conjecturam que o fator responsável seja a dissociação entre Número e Caso.

Exemplificando, quando o pronome é forte, o verbo aparece no plural, enquanto com pronome não-forte²⁶ o verbo encontra-se na terceira pessoa do singular, sem concordância de número, conforme os exemplos extraídos de Ramos, Vital e Souza, 2007, p. 3²⁷:

- (1) a. *es inventa* um bocado de coisa (E42)
 b. os próprios industriais, *eles num ampliam* (E21)

As etapas da análise de dados dos autores levariam a estabelecer que os contextos que favorecem os pronomes fracos caracterizam-se por:

- I. estar numa sentença com o verbo no tempo presente;
- II. possuir aspecto inconcluso;
- III. possuir traços *phi* de 3ª pessoa, plural e [- específico].

Abaixo os exemplos apresentados por Ramos, Vital e Souza, 2007, p. 4:

²⁵ É interessante observar que esta menção à teoria de Cardinaletti e Starke, por parte de Vital e Ramos, originará um debate acalorado a partir de um estudo de Petersen (2008). Os dois artigos serão comentados adiante.

²⁶ Termo adotado no artigo. Não foi explicitado se há uma real diferença entre não-forte e fraco.

²⁷ Os autores não esclarecem de onde foram retirados os exemplos abaixo, se de algum *corpus*, ou de dados de introspecção.

- (2) a. Eu o fiz [= o que lhe havia prometido]
 b. Eu os fiz [- genérico]
 c. Homens são assim. Contentá-los é difícil.
 d. *Homem é sempre assim. Contentá-lo é difícil.

Parece-nos, neste ponto, que não ficou clara a origem dos dados: o leitor não tem certeza se se trata de exemplos provindos de *corpus* e se pergunta se sejam de um corpus ou de introspecção. Além disso, se a exemplificação das características I e III soa-nos convincente, não ocorre o mesmo com as características II. Como não são apresentados dados para tal, é complexo para o leitor conseguir entender a afirmação de que os contextos favorecedores dos pronomes fracos apresentam aspecto inconcluso.

Fazendo referência a vários estudos sobre o PB e o francês, os autores questionam o motivo pelo qual justamente a 3ª pessoa é a concordância *default*. Por um lado, afirmam eles, a maioria dos estudos sobre o PB sugeriria que é o próprio enfraquecimento da concordância verbal o responsável por tal situação. Mas a explicação residiria no fato de o pronome reduzido estar mais próximo dos expletivos, por não ser referencial. Sua sugestão é que a categoria vazia licenciada pelo verbo na terceira pessoa seria um expletivo nulo. A conclusão dos autores, portanto, é que a gênese do pronome expletivo lexicalmente realizado no PB reside na ausência de concordância entre pronome reduzido de terceira pessoa e V²⁸.

2.3.1. A forma *cê*: gramaticalização ou forma fraca?

Os estudos apresentados no parágrafo anterior sustentam a cliticidade, mesmo que *in fieri*, da forma *cê*. Nesta seção será exposto um debate recente sobre essa questão, a partir do artigo de Petersen (2008).

A autora, retomando os trabalhos de Vitral (1996) e Ramos (1997), questiona as análises feitas anteriormente e, com base na teoria da tripartição entre formas fortes, fracas e clíticas de Cardinaletti e Starke (1999), argumenta que *cê* é um pronome fraco.

Petersen (2008) reanalisa um exemplo de Vitral (1996, p. 120):

Eu vou fazer cê feliz.

²⁸ A argumentação dos autores vale-se de amplas análises em árvore, além de inúmeras citações de vários autores. O que, contudo, nos perguntamos mais uma vez é a origem dos exemplos citados, cuja fonte nem sempre é clara.

Segundo o estudo precedente de Vitral, a forma *cê* está em uma posição de predicação, o que possibilita sua ocorrência pós-verbal, sem a violação da obrigatoriedade da próclise. Petersen (2008) afirma que essa sentença não é aceitável para todos os falantes e pode ser vista mais como uma exceção do que como uma sentença comum. Após testar outros clíticos em contextos similares, a autora afirma que:

se cê fosse um clítico, seria esperado que cê se comportasse como os demais clíticos, e que se mantivesse em próclise com os verbos temáticos das sentenças. Mas não é o que ocorre. [...] parece que analisar cê como um clítico implica admitir um comportamento bastante idiossincrático desta proforma.(PETERSEN, 2008, p. 287)

Continuando a análise dos artigos de Vitral (1996) e Ramos (1997), Petersen (2008) critica a aproximação por eles efetuada entre a forma *cê* e o nominativo *se*, considerados ambos marcadores de indeterminação do sujeito do enunciado. Segundo a pesquisadora, mesmo as formas *ocê* ou *você* poderiam ser utilizadas para tal. E acrescenta que em construções negativas a posição de *cê* é distinta em relação àquela de *se*.

As conclusões de Ramos (1997) também não convencem Petersen (2008) quanto ao fato que, seja no caso das construções interrogativas do tipo “que que”, seja nas configurações sintáticas em que *cê* é contíguo ao verbo, os dados não comprovam o estatuto clítico da forma *cê*. Para a estudiosa, *cê* possui uma distribuição muito diferente da dos demais clíticos e, ao mesmo tempo, suas restrições sintáticas fazem pensar em uma natureza diferente dos outros pronomes de segunda pessoa. Sua proposta é uma reanálise dos dados de Vitral (1996) e Ramos (1997) à luz da teoria da Tripartição do Sistema Pronominal de Cardinaletti e Starke (1999).

Abaixo segue uma explicação da teoria, que propõe que o sistema pronominal seja dividido entre proformas fortes, fracas e clíticas. Para melhor compreensão, tentaremos resumi-la no parágrafo a seguir.

Cardinaletti e Starke (1999) analisam as posições do DP, ou seja, do *Determinant Phrase*. A projeção mais alta do mesmo coincide com a realização dos traços de concordância (*Agreement*) e dos traços referenciais. Há inúmeras posições funcionais às quais está associada a cabeça nominal que divide o DP em diferentes níveis funcionais. O nível mais alto, realizando os traços referenciais, é o nível do *Complementizer Phrase* (CP) Nominal, ou seja, do complementizador nominal. O pronome forte possui todos os níveis funcionais, mesmo não os realizando foneticamente. Os pronomes fracos e os clíticos carecem

do nível do CP, aquele do complementizador, ou seja, dos traços referenciais e portanto não podem ser nem coordenados nem modificados. No nível da estrutura superficial, os clíticos necessitam ocorrer em uma projeção funcional que possa atribuir-lhes os traços de caso. Por este motivo, os pronomes fracos e aqueles clíticos necessitam de uma relação estrutural com *Agreement* (projeção de concordância). O pronome fraco, sendo uma projeção máxima, posiciona-se no especificador (Spec) do sintagma de concordância em uma configuração local com nó terminal (X°). Por outro lado, os clíticos ocorrem em um núcleo funcional da Forma Fonológica, que Cardinaletti e Starke (1999) acreditam estar associada aos traços prosódicos que faltam aos clíticos, mas estão presentes nas outras tipologias de pronomes. Os pronomes fracos, portanto, são referenciais, mas somente quando associados a um antecessor, sintagma nominal ou pronome forte, sozinhos não podem ser interpretados.

Voltando à análise de Petersen (2008), haveria, portanto, dois níveis de deficiências:

Os pronomes fracos são deficientes em relação aos pronomes fortes; os clíticos são deficientes em relação aos pronomes fracos. Cada classe compartilha a propriedade de deficiência da sua classe superior e acrescenta novas deficiências.(PETERSEN, 2008, p. 290)

A reanálise dos exemplos citados por Vitral (1996) leva Petersen (2008) a afirmar que os ambientes em que *cê* ocorre são na realidade exclusivos dos pronomes fortes. Segundo a autora, dentro do modelo por ela aplicado da tripartição pronominal, os argumentos que levaram Vitral (1996) a afirmar a cliticidade de *cê*, mostram simplesmente que a proforma não é um pronome forte, mas um pronome deficiente. Através da análise de exemplos de vários autores, Petersen (2008) conclui que *cê* diferencia-se do comportamento dos clíticos, mostrando ao contrário um comportamento compatível com aquele dos pronomes fracos. De fato, afirma a pesquisadora:

Somente outros clíticos podem intervir entre verbo e clítico sujeito. A combinação de clíticos de diferentes funções cria uma ordem rígida entre eles, os clitic clusters.(PETERSEN, 2008, p. 296)

Prosseguindo em sua análise, a autora afirma que também no PB atual existe a impossibilidade de separar o clítico do verbo. Contudo, a forma *cê* não está sujeita a esse tipo de restrição: isso diferencia *cê* dos outros clíticos, e sendo tipicamente uma proforma deficiente, sua classificação corresponde àquela dos pronomes fracos.

Outro teste efetuado pela autora compara o comportamento de *cê* àquele dos clíticos sujeito dos dialetos do norte da Itália e dos pronomes fracos do francês em construções de topicalização. As evidências por ela encontradas mostram que *cê* comporta-se como uma proforma fraca.

A análise continua com outras estruturas do PB que, segundo Petersen (2008), evidenciam a diferença entre pronomes fracos e clíticos: o hiperalçamento do sujeito²⁹ é compatível com os pronomes fracos, mas vedado aos clíticos. O mesmo acontece com as estruturas de tópico marcado, como no exemplo retirado de Ferreira (2000, 2004, p. 299):

(A) **Cê*, o João me contou que não quer sair.

Polemizando com as afirmações de Vitral (1996), e em clara contradição com o que é afirmado naquele estudo, Petersen (2008) afirma que:

- a) *cê* não possui o mesmo comportamento do clítico nominativo *se*, como nos exemplos abaixo, Nunes (2007) *apud* Petersen, 2008, p. 301³⁰:

(B) a. *Cê* que fez isso?

b. **Se* que fez isso?

- b) *cê* pode ser acusativo³¹, Nunes (2007) *apud* Petersen, 2008, p. 301:

(C) a. *Cê* que a Maria foi visitar?

b. **Te* que a Maria foi visitar?

²⁹ O hiperalçamento do sujeito ocorre quando o sujeito de uma oração subordinada, finita ou com infinitivo flexionado, é alçado para a oração matriz, com concordância verbal em ambas. O exemplo citado no texto, retirado de Ferreira (2000, 2004), é o seguinte: *A Maria parece que está cansada*, em que haveria hiperalçamento do sintagma nominal *a Maria*.

³⁰ Este exemplo aponta para a aceitabilidade do pronome fraco *cê* em posição de “tópico não marcado”, enquanto um pronome clítico como *se* não pode estar associado a esta posição. Essa é uma das evidências de que a forma *cê* seria uma forma fraca e não um clítico.

³¹ Esta afirmação é muito importante, pois em nossa pesquisa baseada em *corpus* foram encontradas duas ocorrências da forma nessa função, fato que outros estudos citados adiante também afirmam ser possível, contradizendo grande parte das afirmações de Cunho gerativista.

- c) mesmo quando possui função de acusativo *cê* não se comporta como um clítico, como pode ser notado no exemplo acima, em que a mesma construção não funciona com o clítico *te*.

Portanto a posição de Petersen (2008) é que a proforma *cê* seria um pronome fraco. Sua análise continua, sempre a partir dos dados apresentados por Ramos (1997) e Vitral (1996). O questionamento refere-se à impossibilidade de *cê* ocorrer na posição de complemento de preposição. A posição de Ramos (1997) de que há um processo de redução causado pelo condicionamento fonológico do contexto em formas como *com cê*, interpretada como [kõ'se], é rejeitada por Petersen (2008). A estudiosa reexamina a questão com os exemplos abaixo, Petersen, 2008, p. 303³²:

- a. *Eu disse isso p'cê ontem.*

[p'se]

- b. *Quer que eu vá c'cê lá.*

[k'se]

Segundo Petersen (2008) esses dados, observados empiricamente, refletem o uso da forma *cê* após a preposição, e pertencem mesmo a comunidades em que a forma *ocê* não é frequente, portanto não pode ser sugerida uma pronúncia com realização fonética da vogal [o]. A autora afirma que esse tipo de enunciado somente é encontrado quando há uma reestruturação prosódica na fala. Isso significaria que, quando há uma pausa entre os dois elementos fonéticos da preposição contraída e da proforma *cê*, a sequência é rejeitada pelos falantes do PB, mas quando a realização corresponde à forma [ˈpse] e [ˈkse], é plenamente aceita. Retomando as teorias de Cardinaletti e Starke (1999), Petersen (2008) afirma que somente pronomes “deficientes” estão licenciados para formar uma unidade prosódica única com o elemento adjacente. Quando a proforma *cê* é complemento de preposição recebe uma acentuação lexical. Como somente pronomes fracos, e não clíticos, recebem acento lexical, esse dado confirmaria que:

- a) *cê* não poderia ser clítico pois os clíticos são átonos;
- b) portanto *cê* seria uma proforma fraca;
- c) e proformas fracas podem ser complemento de preposição no PB.

³² Observamos aqui que não é claro se os dados apresentados por Petersen pertencem a algum tipo de *corpus* não melhor especificado ou se são fruto de introspecção, pois, além das citações dos exemplos de Vitral e Ramos que são reanalisados, não se faz referência à origem dos outros dados do PB. Quanto à simbologia fonética adotada, mantivemos aquela da autora.

Concluindo suas análises, Petersen (2008) confirma, no arcabouço teórico da Teoria da Tripartição de Cardinaletti e Starke (1999), que as formas *ocê* e *você* são formas fortes. Mas, pelo princípio *Minimizar α* ³³, deveria haver algum tipo de restrição que evitasse o uso dessas proformas quando o uso de *cê* é possível. A saída apontada pela estudiosa é que *ocê* e *você* são formas “ambíguas: possuem formas subjacentes homófonas, uma fraca e outra forte” (PETERSEN, 2008, p. 306).

O esquema final apresentado por Petersen (2008) é, portanto, o seguinte:

Pronomes fortes: ocê, você

*Pronomes fracos: cê, ocê, você*³⁴

Em resposta direta aos comentários e críticas de Petersen, Vitral e Ramos (2008) rebatem que a leitura feita pela estudiosa de seus dados não seria isenta. A principal objeção dos autores se dá quanto à utilização feita dos dados por eles apresentados. Segundo Vitral e Ramos (2008) as exemplificações propostas por Petersen (2008), como os novos ambientes sintáticos, já faziam parte do conjunto de dados por eles proposto em Vitral (1996) e Vitral e Ramos (2006). Também a análise da interpolação por eles proposta é citada como não compreendida em sua importância por Petersen (2008). As mais duras críticas referem-se ao tratamento e compreensão de seu conceito de “dado”. Vitral e Ramos (2008, p. 392) afirmam que:

*a quase totalidade dos dados que utilizamos já foi atestada em trabalhos que se valeram de análises quantitativas baseadas em dados reais, extraídos de corpus*³⁵.

Os estudiosos repudiam a atitude de Petersen (2008) naquilo que, segundo eles, é negar a gramaticalidade de um dado e por consequência recusar uma inteira análise: segundo Vitral e Ramos (2008), Petersen (2008) simplesmente julgou parte dos dados agramatical, mas não se preocupou em verificá-los com base na teoria subjacente.

³³ No Modelo *Princípios e Parâmetros* (cf. Chomsky, 1998) os diferentes tipos de transformações são substituídos por um único movimento: o Movimento α , que tenta dar conta da interpretação de constituintes que aparecem fora do lugar onde deveriam ser interpretados.

³⁴ Acreditamos que o que foi entendido por Petersen é que as formas *ocê* e *você* pertençam seja às formas fracas, seja às fortes, dependendo do contexto. Contudo, a conclusão, talvez com alguns exemplos, poderia ser um pouco mais claro para o leitor.

³⁵ Deve-se esclarecer que os estudiosos não citam em detalhe quais dados foram utilizados, mas referem-se simplesmente a Andrade (2004, p. 96), que afirma que em sua dissertação de mestrado os dados de fala analisados confirmam aqueles introspectivos de Vitral, e que nenhuma das formas agramaticais citadas pelo estudioso foi encontrada.

Quanto aos exemplos citados pela estudiosa sobre a ocorrência de *cê* após preposição, Vitral e Ramos (2008) questionam o termo “empiricamente” coletados, utilizado pela autora, por não especificar seu contexto, o tipo de coleta feito, dados sobre os falantes entre outras informações. Enfim, os estudiosos contestam o fato de que não teria sido levado em conta o processo diacrônico de mudança, na análise de Petersen (2008), e que a afirmação de que o processo de cliticização está em curso, por eles defendido, foi ignorado.

Continuando em sua revisão do artigo de Petersen (2008), Vitral e Ramos (2008) reafirmam que a forma *cê* é resultado de um processo de redução fonética, e que por isso não tem o mesmo estatuto de *ocê* e *você*, que a precedem imediatamente. Referindo-se a seus trabalhos anteriores, os estudiosos citam os resultados de suas análises acústicas e concluem que a duração e intensidade de *cê* seriam bastante próximas daquela de *se*. O que é questionado em relação ao estudo de Petersen (2008), após várias exemplificações sobre a concepção de gramaticalização, é que a análise da estudiosa se resume a uma simples definição da categoria, clítica ou fraca, a que a forma *cê* pertence. Segundo eles este é:

*um empobrecimento considerável da discussão*³⁶.

Segundo Vitral e Ramos (2008), O trabalho de Petersen é considerado uma mera tentativa de rotular o *cê*, sem qualquer espécie de visão de tipo diacrônico que tente explicar como ocorreu o fenômeno de redução. Segundo os estudiosos:

*a cliticização vista como um processo permite compreender a variação entre línguas e entre períodos de uma mesma língua*³⁷.

Somente dessa forma, afirmam, seria possível entender como as formas reduzidas se tornam afixos. Tratar a cliticização como um processo de inovação linguística que faz parte do conjunto dos processos de gramaticalização permitiria a compreensão da grande variedade das formas reduzidas nas várias línguas. E isso é fundamental para perceber como

*as formas reduzidas podem se encontrar, numa mesma sincronia, em estágios diferentes de seu processo de cliticização, e que, mesmo se a escolha for de apenas encontrar um rótulo para cada etapa, a terminologia forma forte, forma fraca e clítico é insuficiente.*³⁸

³⁶ Vitral e Ramos, 2008, p. 396.

³⁷ *Ibidem*, p. 397.

³⁸ *Ibidem*, p. 398.

Continuando na análise do artigo de Petersen (2008), Vitral e Ramos (2008) revisitam os exemplos por ela propostos para a forma *cê* como complemento de preposição. O parecer dos estudiosos é que

*os dados parecem muito forçados.*³⁹

Segundo eles é possível que haja uma vogal epentética entre o “p” e o “c”, pois foi efetuada uma análise acústica destes enunciados e não houve um resultado definitivo que pudesse confirmar as afirmações de Petersen (2008).

Um outro contra-argumento de Vitral e Ramos (2008) em relação à análise da estudiosa refere-se à possibilidade de ocorrência da forma *cê* em construções de alçamento. Discutindo a possibilidade do fenômeno em outras línguas, os estudiosos chegam à conclusão da impossibilidade de alçamento do *cê*. O fato de esta forma não poder ser topicalizada faz pensar que os dados a esse respeito propostos por Petersen (2008) pareçam “forçados” e necessitem de uma análise “mais satisfatória”⁴⁰.

Finalizando sua reanálise do artigo de Petersen (2008), Vitral e Ramos (2008) fazem objeção a respeito dos critérios utilizados para demonstrar que *cê* é uma forma fraca. São discutidos o comportamento sintático, a semântica do item e a possibilidade de apenas os pronomes fortes serem c-modificados. De acordo com os estudiosos, seguir o raciocínio de Petersen (2008) leva à conclusão de que *cê* é uma forma forte. Isso é indício das dificuldades enfrentadas na tarefa de tentar definir *cê* segundo uma simples terminologia. Ao mesmo tempo, esse raciocínio serve a Petersen (2008) para buscar uma explicação às coocorrências das formas *cê*, *ocê* e *você* na posição de sujeito.

Para responder a essa questão, Vitral e Ramos (2008) fazem referência a Engerland (2002), afirmando que o ponto de partida é aquele da mudança linguística e o ponto que desencadeia o processo é a redução morfofonológica. Como pronomes fortes e fracos são morfológicamente idênticos, e os clíticos, por outro lado, reduzidos do ponto de vista fonológico, Vitral e Ramos (2008) recriminam Petersen (2008) por ter ignorado tal fato.

³⁹ *Ibidem*, p. 399.

⁴⁰ *Ibidem*, p. 400.

Concluindo, os estudiosos optam por manter a denominação de forma clítica no que se refere ao item *cê*, mesmo admitindo que possa não ser a mais adequada, e sublinhando que a interpretação da cliticização como um processo é a mais oportuna⁴¹.

2.3.2. Outras abordagens

Nas seções anteriores foram revistos estudos que trataram a análise das formas *cê*, *ocê* e *você* sob um ponto de vista teórico, a fim de estabelecer critérios taxonômicos e processos de redução. Nesta seção, serão expostos trabalhos que propõem metodologias e testes experimentais para estabelecer o estatuto das formas em análise.

Ciríaco, Vitral e Reis (2004) realizam um estudo sobre formas reduzidas, investigando o comportamento acústico das formas *cê* (redução de *você*), *ez* (redução de *eles*) e *num* (redução de *não*). Foi desenvolvido um experimento em laboratório a fim de controlar todas as variáveis de contexto fonético para obter dados sobre a duração e a intensidade das formas pesquisadas. Foram memorizadas e repetidas, por quatro alunos do curso de Letras da UFMG, dois de cada sexo, nascidos e residentes em Belo Horizonte, treze sentenças. Destas, seis foram pronunciadas com três velocidades diferentes⁴²: normal, rápida e rapidíssima. Desta forma buscou-se reproduzir na fala a redução das formas *cê*, *ez* e *num*, já que os textos apresentavam as formas plenas, ou seja, estava escrito *você* e *eles*. O *corpus* é composto por 300 sentenças, 75 produzidas por cada informante. As sentenças foram gravadas, editadas e analisadas em seu espectrograma⁴³. As frases utilizadas são as seguintes (CIRÍACO *et al.*, 2004, p. 147):

1. *Diga Cê disse que Zé vai viajar. (Cê = Cecília)*
2. *Diga eles falou que Zé vai viajar.*
3. *Diga Jô não disse nada não.*
4. *Diga cê disse que Zé vai viajar. (cê = pronome)*

⁴¹ Acreditamos que a inserção na revisão bibliográfica de nosso trabalho da discussão sobre a oportuna definição da forma *cê* seja importante para que o leitor perceba quanto muitos se preocupam com tal problemática e, muitas vezes, acabem por deixar de lado os dados. Sem querer de alguma forma desmerecer o trabalho de outrem, e sabendo dos limites do que aqui é apresentado, sublinhamos a importância de se apresentarem os dados que foram analisados, sua forma de coleta, sua possibilidade de acesso à comunidade acadêmica, de modo a compreender melhor a evolução da língua.

⁴² No texto é usada essa terminologia apesar do correto ser “taxa de elocução”, entendida como o número de unidades linguísticas (sejam elas sílabas ou segmentos) que constam em um intervalo de fala pausa-incluída dividido pela respectiva duração. A esse respeito veja-se Trovain (2004).

⁴³ Gostaríamos de esclarecer que o artigo não menciona se, e qual tipo de instrução específica foi dada aos informantes sobre como deveria ser feita a leitura, que tipo de edição foi realizada, e de qual forma foi feita a análise.

5. *Diga **cê** falou que Zé vai viajar. (cê = pronome)*
6. *Diga o estilista vestiu **Cê** bem. (Cê = Cecília)*
7. *Diga **eles** disse que Zé vai viajar.*
8. *Diga o estilista vestiu-se bem.*
9. *Diga **Cê** falou que Zé vai viajar. (Cê = Cecília)*
10. *Diga **Jô não** falou nada não.*
11. *Diga o estilista vestiu **cê** bem. (cê = pronome)*
12. *Diga **ex** diz muita coisa⁴⁴.*
13. *Diga **eles** diz muita coisa.*

Os pesquisadores consideraram os resultados satisfatórios em demonstrar o processo de cliticização, embora admitam que o estudo esteja em uma fase preliminar. Contudo, os dados não foram submetidos a nenhum tipo de normalização ou análise estatística: suas conclusões são fruto de uma análise comparativa das formas plenas e das reduzidas dos dados apresentados acima.

A tabela (1) abaixo reporta a média da intensidade (dB) e da duração (ms) para as partículas estudadas no artigo, Ciríaco *et al.*, 2004, p. 149:

⁴⁴ Os autores esclarecem que para obter as formas reduzidas foi utilizado o recurso da velocidade de fala, apesar de não detalharem como isso foi medido, ou seja, qual taxa nominal foi utilizada. No conjunto das frases pronunciadas com o emprego da velocidade de fala fazia parte também a forma *ex*, como em *ex-marido* e *ex-mulher*. Contudo, não é explicado o que se queria obter com a introdução desse segmento.

Médias				
Grupos	Nº Sentença	Realização	Duração (ms)	Intensidade (db)
1	02	...eles falou...	285.6	54.6
	02	...ez falou...	153.8	56.4
	7 e 13	...eles diz/disse...	287.1	56.7
	7 e 13	...ez diz/disse...	165.7	59.3
	12	...ex diz...	185.2	58.6
2	10	...não falou...	179.8	51.7
	10	...num falou...	101.7	55.9
	3	...não disse...	171.7	54.9
	3	...num disse...	128.9	56.5
3	9	...CE (Cecília) falou...	222.6	57.9
	5	...ce (pronome) falou...	152.5	55.9
	1	...CE (Cecília) disse...	271.7	56.1
	4	...ce (pronome) disse...	224.9	56.5
4	6	vestiu CE (Cecília)..	233.4	56.6
	11	vestiuCe (pronome)	204.9	54.9
	8	...vestiu-se....	132.9	54.9

Tabela 1: Médias da intensidade(dB) e da duração(ms) para as partículas estudadas

Quadro 1: Tabela das médias de intensidade e duração em Ciríaco *et al.*, 2004, p. 149.

Após a análise dos dados acima, as conclusões a que eles chegaram são as seguintes:

- I. foi por eles atestada foneticamente a redução de *você*, na forma *cê*;
- II. na posição de sujeito *cê* apresentaria uma duração menor em relação à forma *CÊ*, contração de Cecília;
- III. na posição de objeto de verbo *cê* apresentaria uma duração menor em relação a *CÊ* e maior que o clítico *se*;
- IV. quanto ao parâmetro intensidade, *cê* apresentaria intensidade menor que *CÊ* e igual à de *se*;
- V. portanto, a cliticidade de *cê* seria confirmada seja pelo parâmetro duração, seja pelo parâmetro intensidade.

Um outro trabalho a ser mencionado é o de Barbosa L. P. (2005). Em sua dissertação de Mestrado a estudiosa pretende estabelecer, através de testes de vários tipos, se a forma *cê* é um clítico ou uma palavra plena. A autora utiliza em sua análise dados introspectivos e reais, os últimos coletados em São Francisco/MG, e previamente usados por Coelho (1999). Os testes obedecem a duas diferentes visões: aquela da cliticização, e são de tipo fonológico, acentual, morfológico e sintático; e aquela da fonologia prosódica, com testes prosódicos.

Os testes fonológicos seguem Zwicky (1985) com o objetivo de identificar a cliticidade de *cê*, caso apareça um elemento hospedeiro na palavra fonológica que domine *cê*.

A autora efetua também uma série de testes sintáticos, respaldada por Zwicky (1985) e Kayne (1975). Os parâmetros objeto de análise são:

- I. “Apagamento sob identidade”: quando um dos dois elementos $x + y$ pode ser apagado, ambos são palavras plenas (BARBOSA L. P., 2005, p. 107):

(28a) *Eu vou à festa e **cê vai** também.*

(28b) *Eu vou à festa e **cê**, também.⁴⁶*

Como foi possível realizar o apagamento sob identidade do verbo, as duas palavras são plenas.

- II. “Substituição”: quando um elemento pode ser substituído por outro, ambos são palavras plenas (BARBOSA L. P., 2005, p. 108):

(29) *Duas horas e **cê** só falou isso agora!*

(30) *Duas horas e **tu** só falou isso agora!*

(31) *Duas horas e **Gisele** só falou isso agora!*

O *cê* pôde ser substituído seja por *tu* ou por um NP lexical, provando que é uma palavra plena.

- III. “Movimento”: quando um de dois elementos que estejam combinados pode ser movimentado, ambos são forma plena (BARBOSA L. P., 2005, p. 108):

(32a) *Eu acho que **cês** não passaram de ano, não.*

(32b) ***Cês**, eu acho que não passaram de ano, não.*

Nos dois enunciados, *cê* mostrou poder se mover isoladamente, ou seja, sem depender de outro elemento. Portanto, é uma forma plena.

Outra série de exemplos busca comprovar a não cliticidade da forma *cê*. São eles:

- I. É possível intercalar vários elementos entre *cê* e o verbo (BARBOSA L. P., 2005, p. 108):

⁴⁶ A presença da vírgula após o *cê* respeita o exemplo da autora.

(33) *Cê nunca mais me verá.*

II. É possível a modificação de *cê* (BARBOSA L. P., 2005, p. 109):

(34a) *Cê é o mais sem-vergonha.*

(34b) *De todos, cê só é o mais sem-vergonha.*

III. É possível a coordenação de *cê* a um NP lexical e a um pronome tônico (BARBOSA L. P., 2005, p. 109):

(35a) *Cê e Joaquina* vão viajar juntas.

(35a) *Cê e ele* serão felizes.

IV. É possível a topicalização de *cê* (BARBOSA L. P., 2005, p. 109):

(36) *Cês, eu acho que cês não passaram de ano, não.*⁴⁷

Enfim, Barbosa L. P. (2005) produziu uma série de testes prosódicos para verificar a hipótese de que *cê* é uma palavra fonológica. A autora associou a forma *cê* a outros monossílabos tônicos do português, como *de*, *nu* e *dá* que foram comparados com suas respectivas formas átonas (BARBOSA L. P., 2005, p. 112):

Palavras monossílabas acentuadas - Palavras monossílabas não-acentuadas

dê [de]

de [dɜɪ]

nu [nu]

no [nɔ]

dá [da]

da [də]

Dado que estes pares de monossílabos se diferenciam, segundo Barbosa L. P. (2005), somente pela tonicidade, a pesquisadora sustenta que o mesmo acontece com a forma *cê* (BARBOSA L. P., 2005, p. 112):

⁴⁷ Esse exemplo merece destaque porque confirma, com dados de *corpus*, o que muitos autores (como Othero, 2013, adiante) negam, isto é, a possibilidade de topicalização de *cê* em posição de sujeito. Em nossa análise foi encontrada somente uma ocorrência deste tipo, mas o fato de que outros estudiosos tenham se deparado com tal situação reforça a hipótese de que isso possa ocorrer.

cê [se]

se [sɨ]

A conclusão da estudiosa é que, portanto, *cê* não seria um monossílabo clítico, mas tônico.

Outros argumentos que comprovariam tal fato são que *cê*, contrariamente a *se*:

- I. não sofre redução vocálica (BARBOSA L. P., 2005, p. 113):

dê [e]/ *[i]

de [e]/ [i]

cê [e]/ *[i]

se [e]/ [i]

- II. não é afetado por semivocalização ou apagamento (BARBOSA L. P., 2005, p. 113):

cê ia [e]/ *[i]/ *0

se una [e]/ [i]

De acordo com Barbosa L. P. (2005), a fraca intensidade da forma *cê* em alguns ambientes pode ser explicada com os argumentos de Major (1985). Dependendo do contexto em que esteja inserido, *cê* poderá sofrer a perda de seu acento, especificamente quando se trata do nível da sentença. Isto, contudo, não significa para a autora que a forma *cê* perdeu seu acento de palavra, mas que não recebeu acento frasal.

O estudo de Barbosa L. P. (2005), através de toda a série de testes aplicados, sugere que *cê* não seria um clítico, mas uma palavra plena.

Outro estudo que discute o estatuto dos pronomes *você* e *cê* no PB é o de Othero (2013). O autor propõe, através de testes de julgamento no uso das formas acima, uma análise distribucional que prove que *você* e *cê* são dois pronomes pessoais distintos, que apresentariam características sintáticas e prosódicas próprias. Segundo Othero (2013), uma análise sincrônica minuciosa permite afirmar que *cê* não é uma mera evolução de *você*, mas um pronome de uso especializado. Os testes de julgamento por ele implementados dividem-se entre testes de caráter sintático e testes de tipo prosódico.

No teste de função sintática de sujeito da frase, seja *cê* seja *você* podem desempenhar este papel, o que prova, segundo Othero (2013), que *cê* não é um clítico, pois eles não são capazes de ocupar esta função (OTHERO, 2013, p. 4):

- 1) *Aí **você** chega pra mim e diz que tá tudo errado.*
- 2) *Aí **cê** chega pra mim e diz que tá tudo errado.*
- 3) **Aí **te** chega pra mim e diz que tá tudo errado.*

Contudo, o autor faz notar que somente o pronome *você* pode aparecer na posição de sujeito posposto ao verbo (OTHERO, 2013, p. 4):

- 4) *Aí chega **você** pra mim e diz que tá tudo errado.*
- 5) **Aí chega **cê** pra mim e diz que tá tudo errado.*
- 6) **Aí chega **te** pra mim e diz que tá tudo errado.*

O autor afirma que, no caso acima, *cê* “apresenta a mesma restrição que um pronome clítico em PB”.

O teste seguinte considera a interpolação de elementos (OTHERO, 2013, p. 5):

- 7) ***Você** vive bem no RS.*
- 8) ***Você não** vive bem no RS.*
- 9) ***Cê** vive bem no RS.*
- 10) ***Cê não** vive bem no RS.*
- 11) ****Não cê** vive bem no RS.*
- 12) ***Se** vive bem no RS.*
- 13) ***Não se** vive bem no RS.*
- 14) ****Se não** vive bem no RS.*

Othero (2013) afirma que *cê* e *você* aceitam elementos interpolados, mas o mesmo não acontece com os clíticos. Este é, portanto, um argumento contrário à classificação de *cê* como clítico.

Outros testes mostram a distribuição de *cê* e *você* na posição de complemento verbal: segundo Othero (2013) apenas *você* pode aparecer nesta posição, enquanto os clíticos figuram

somente em linguagem monitorada⁴⁸. Já na posição de objeto pré-verbal, há somente clíticos, e não *cê* ou *você*.

No caso de complemento de preposição, o autor afirma que *cê* não pode comparecer como tal⁴⁹, como mostram os exemplos abaixo (OTHERO, 2013, p. 6):

21) *Eu entreguei pra você o livro da Maria.*

22) **Eu entreguei pra cê o livro da Maria.*

23) **Eu entreguei pra te o livro da Maria.*

Como, segundo o autor, somente os pronomes tônicos ocorrem nessa posição, esse seria um argumento favorável à cliticidade de *cê*.

Outro teste no qual *cê* se comporta como clítico é, para Othero (2013), aquele da topicalização⁵⁰ (OTHERO, 2013, p. 6):

24) *Você, a Maria não viu ontem na festa.*

25) **Cê, a Maria não viu ontem na festa.*

26) **Te, a Maria não viu ontem na festa.*

Como somente o pronome tônico pode ser deslocado à esquerda, o comportamento de *cê* assemelha-se àquele dos clíticos.

A conclusão parcial de Othero (2013) é que tônicos e clíticos estão em distribuição complementar. O que ele observa é que *cê* assemelha-se em alguns pontos aos pronomes tônicos, e em outros aos pronomes clíticos. Segundo o autor:

“Morfofossintaticamente, o pronome cê parece ser um pronome exclusivamente de Caso nominativo, uma vez que não pode aparecer na posição de complemento verbal nem de complemento de preposição. [...] Entretanto, cê pode figurar apenas como sujeito preposto e não posposto. Isso nos leva a crer que exista aí uma restrição de natureza prosódica, pois ele não tem força sintática⁵¹ para

⁴⁸ O autor não especifica o que quer dizer com isso, se se refere a um linguajar mais formal ou a outra situação, como a entrevista, o discurso, ou testes controlados em laboratório.

⁴⁹ Como já foi dito acima, e como será visto adiante, na exposição de nossos dados, o teste não parece ser a solução mais apropriada, já que dados de corpus mostram diferenças. O que se pode notar na argumentação de Othero é que basear-se somente em dados de outros autores ou em testes de julgamento diz muito pouco sobre o uso real da língua. Parece-nos também que o a lógica de tal argumentação soe pouco convincente.

⁵⁰ Como visto acima, no exemplo (36) de Barbosa (2005), para a autora existe a possibilidade de topicalização de objeto gramatical. Mais uma vez, há grandes discordâncias entre os autores e faltam verificações empíricas.

⁵¹ O autor não esclarece o que significaria a expressão “força sintática”, que neste contexto ficou pouco clara.

ficar à direita de um agrupamento prosódico, onde se espera que apareçam elementos prosodicamente mais fortes, já que ele pode ser sujeito de um verbo apenas em posição pré-verbal e não pós-verbal.” (OTHERO, 2013, p. 7-8)

São apresentados outros testes. No primeiro, observa-se o comportamento do uso de um pronome em posição de objeto de um verbo e sujeito de outro, em lexias verbais complexas (OTHERO, 2013, p. 8):

27) *Eu vi **você** chegando tarde em casa ontem.*

28) *Eu vi **cê** chegando tarde em casa ontem.*

29) *Eu **te** vi chegando tarde em casa ontem.*

Othero (2013) afirma que os três pronomes podem aparecer em uma estrutura aparentemente idêntica, contudo, há três explicações distintas para cada caso:

- I. para o pronome tônico esta seria uma posição prototípica tanto para objeto como para sujeito;
- II. o pronome *cê* encontra-se em posição de complemento do verbo, mas isso seria somente em aparência, pois, na realidade, ele pode estar em tal posição por ser sujeito do verbo que segue;
- III. o pronome clítico encontra-se em sua posição prototípica, em próclise.

Outro teste efetuado refere-se à ocorrência de *cê* em estrutura de coordenação. Os exemplos buscam restrições prosódicas a que o pronome possa estar sujeito (OTHERO, 2013, p. 9):

34) *Até agora só chegaram **você** e ela.*

35) *Até agora só chegaram [**cê** e ela].*

36) *Até agora só chegaram ela e **você**.*

37) **Até agora só chegaram [ela e **cê**].*

Segundo Othero (2013), os exemplos 35 e 37 evidenciarão as restrições prosódicas a que *cê* está sujeito, podendo atuar como sujeito, preposto ou posposto, desde que fique à esquerda de seu agrupamento prosódico.

Nos exemplos dados a seguir são evidenciadas estruturas de coordenação em que *cê* aparece na função de objeto (OTHERO, 2013, p. 9):

38) *Eu vi **você** e ela juntos.*

39) *?Eu vi [**cê** e ela] juntos.*

40) *Eu vi ela e **você** juntos.*

41) **Eu vi [ela e **cê**] juntos.*

O autor afirma que no exemplo 39 aparece uma situação de conflito: o pronome *cê* não poderia desempenhar o papel de objeto de verbo, mas a condição prosódica, ou seja, a presença de *cê* à esquerda do grupamento prosódico estaria satisfeita. A frase 39, portanto, não seria completamente bem formada. Já o exemplo 41 é, segundo Othero (2013), totalmente agramatical, por *cê* ser objeto de verbo, e estar à direita em seu grupo prosódico.

Othero (2013) reafirma a necessidade de elementos prosódicos que deem respaldo à análise de *cê*, pois as explicações sintáticas não conseguem relatar completamente a distribuição do pronome.

Retomando as análises dos experimentos feitos por Vitral e Ramos (2006) e os resultados de intensidade e duração, Othero (2013) reafirma que *cê* encontra-se entre o clítico e o monossílabo tônico, por isso não pode ser considerado um clítico. Outro argumento contra a cliticidade de *cê* que o autor utiliza é que esse pronome não sofre o processo de alçamento de vogal, como mostram os exemplos abaixo (OTHERO, 2013, p. 12):

51) ***Cê** vive bem aqui.*

51') *[**se**] vive bem aqui.*

51'') **[**si**] vive bem aqui.*

52) ***Se** vive bem aqui.*

52') *[**se**] vive bem aqui.*

52'') *[**si**] vive bem aqui.*

52) ***Me** contaram que **nos** viram quando **te** chamei para sair ontem.*

53') *[**me**] contaram que [**nos**] viram quando [**te**] chamei para sair ontem.*

53''') [*mi*] contaram que [*nus*] viram quando [*tʃi*]⁵² chamei para sair ontem.

Othero (2013) conclui que o pronome *cê* possui propriedades prosódico-acentuais distintas das dos clíticos. Portanto, segundo o autor, “o pronome *cê* não pode ser considerado um pronome nem clítico nem tônico” (OTHERO, 2013, p. 13). Seria ele então um pronome novo na língua, uma “partícula, com as mesmas propriedades prosódicas que *es*, *eis*, *a’ente*, *ente*, *ta* e *num* [...] mas com suas características sintáticas particulares”⁵³.

O último trabalho que se pretende expor aqui é o de Peres (2006). Em sua tese de doutorado, a estudiosa investiga o uso das formas *você*, *ocê* e *cê* por parte de falantes de Belo Horizonte. Seu trabalho, de cunho sociolinguístico, indaga, em um estudo de tipo sincrônico e diacrônico, o comportamento dessas formas levando em conta fatores internos e externos que possam favorecer seu uso.

A autora utiliza dois *corpora*. O primeiro é de 1982, e foi coletado por Viegas (1987), sendo constituído por dezesseis entrevistas sociolinguísticas a jovens e adultos, representantes de classe baixa e de classe média de Belo Horizonte/MG. O outro *corpus*, dito *corpus de 2002*, compõe-se de duas amostras: uma constituída por dados recolhidos pela autora em entrevistas a crianças e adolescentes de Belo Horizonte/MG, e a outra coletada por Hauback (2003), com informantes jovens e adultos, sempre de Belo Horizonte/MG, para um total de 47 textos. Os dois *corpora* são totalmente constituídos por entrevistas gravadas e transcritas pelos autores das coletas. Isso, por parecer da própria Peres (2006), constituiu um problema, pois, em suas conclusões, relata a dificuldade de explicação de alguns fenômenos observados, especialmente em se tratando das entrevistas a crianças e adolescentes, e àquelas de 1982, a informantes de classe baixa. A autora cita a esse propósito o Paradoxo do Observador (cf. LABOV, 1972) afirmando que:

Um “bom” corpus, no caso dos estudos deste tipo, deveria ser constituído de gravações de fala natural, ou seja, de falantes em situação normal de interação. (PERES, 2006, p. 229)

⁵² A simbologia adotada pelo autor foi aqui corrigida com os caracteres IPA normalmente utilizados.

⁵³ *Ibidem*.

Voltando ao estudo, a autora examinou os dados segundo fatores linguísticos e extralinguísticos significativos segundo indicado pelo programa Goldvarb 2001⁵⁴. Os resultados da primeira etapa da análise, aquela sincrônica sobre o *corpus de 2002*, evidenciaram como significativos os seguintes parâmetros:

- I. linguísticos: Função Sintática, Contiguidade em relação ao verbo, Referência e Focalização contrastiva;
- II. extralinguísticos: Faixa etária e Gênero.

Os dados, consultáveis na tabela abaixo, mostram que a forma *cê* é a mais frequente entre as três, com 72,6% das ocorrências, seguida por *você* (23,5%) e *ocê* (3,9%) (PERES, 2006, p. 131):

TABELA 7
Total de ocorrências de *você*, *ocê* e *cê* – *corpus de 2002*

Formas	Ocorrências	%
Você	342	23,5
Ocê	56	3,9
Cê	1055	72,6
Total	1453	100

Quadro 2: Tabela de ocorrências do *corpus de 2002* em PERES, 2006, p. 131.

Conforme a Função Sintática, os dados revelam o seguinte:

- I. na função de Sujeito a forma mais utilizada é *cê* (75,1% das ocorrências), seguida por *você* (22,4%) e *ocê* (2,5%);
- II. na função de Objeto de Verbo a forma mais utilizada é *ocê* (41,7%), seguida por *você* (33,3%) e *cê* (25%);
- III. na função de Objeto de Preposição a forma mais utilizada é *você* (57,8%), seguida por *ocê* (35,5%) e *cê* (6,7%).

Deve ser evidenciado que houve uma distribuição não homogênea nas três formas: 96,1% das ocorrências totais refere-se à função de sujeito e somente 3,9% às outras funções. Segundo a autora, tal fato significa que a função original, não-marcada, seria aquela de

⁵⁴ Segundo Guy e Zilles (2007, p. 105) o "GoldVarb 2001" é uma versão para ambiente Windows do pacote de programas VarbRul - do inglês *Variable Rules Analysis*, "é um conjunto de programas computacionais de análise multivariada, especificamente estruturado para acomodar dados de variação sociolinguística". O programa foi idealizado por Steve Harlow com base na versão anterior, GoldVarb 2.0 de Rand & Sankoff (1990) para Macintosh. O GoldVarb 2001 foi desenvolvido na Universidade de York como um projeto colaborativo entre o Departamento de Língua e Linguística e o Departamento de Ciências da Computação. (ROBINSON, LAWRENCE & TAGLIAMONTE, 2001).

sujeito. A hipótese de Peres (2006) é de que as três formas sofram concorrência de outros itens como pronomes ou forma zero.

Quanto à distribuição, apesar de a forma *cê* ser a favorita entre os informantes, seu peso relativo é de 0.52, a medida que o de *você* é de 0.48 e o de *ocê* de 0.47⁵⁵. Portanto, segundo a autora, não é possível afirmar que a função de sujeito favoreça de maneira marcante a forma *cê*.

Quanto à função objeto de verbo, surpreende a presença de *cê*, apesar de seu peso relativo ser de 0.17, portanto altamente desfavorecedor. Na realidade, as ocorrências de *cê* nesta função são somente três (PERES, 2006, p. 136):

- (E117) *Eu vou ma:ta: cê, fdp!*
 (E142) *E hoje dinheiro suj<o> te lev<a> cê à morte*
 (E614) *(...) cê vai assistir o futebol cê ... nego te arras<ta> cê lá dentro*⁵⁶

Contudo, a própria Peres (2006) sublinha a importância de ter encontrado exemplos de uma função para a forma *cê* que até esse momento a literatura negava.

Com respeito à função objeto de preposição, a forma *cê* também é desfavorecida, com peso relativo de 0.03. Todavia, a autora admite que a ocorrência de tal fenômeno não era esperada, mas foi encontrada. Também nesse caso, as frases pronunciadas foram três (PERES, 2006, p. 136):

- (E182) *(...) guardar aquilo ali p'cê, dentro de si (...)*
 (E683) *Eu vou com [kũ] cê.*
 (E734) *Xô (Deixe eu) perguntar p'cê.*

Todos os enunciados acima, sejam aqueles em que *cê* possui função de objeto de verbo, sejam aqueles em que possui função de objeto de preposição, passaram por análise

⁵⁵ O pacote Varbrul 2001 apresenta seus resultados em porcentagens e pesos relativos encontrados para cada fator inserido na análise. O programa seleciona as variáveis, em vários ciclos de rodadas, determinando sua importância estatística por meio de um valor chamado “nível de significância”. Os pesos relativos variam em um intervalo de 0 a 1: os valores maiores de 0,50 favorecem a ocorrência do dado estudado.

⁵⁶ Como não é possível verificar à oitiva a real pronúncia dos exemplos citados pela autora, optamos por manter a grafia por ela utilizada, que, todavia, deixa margem de dúvidas quanto ao que foi dito pelos informantes, pois foi utilizada em parte uma transcrição fonética e outra ortográfica.

acústica no Laboratório de Fonética da Faculdade de Letras da UFMG. Em todos os casos foi atestado que realmente os informantes utilizaram a forma *cê*⁵⁷.

Com respeito a outro fator linguístico, “contiguidade em relação ao verbo”, em termos de porcentagem, *cê* é a forma mais usada pelos informantes, mas os pesos relativos das três formas indicam que nenhuma é favorecida. Por sinal, afirma a autora:

*O desfavorecimento de *cê* em casos de não-contigüidade confirma a hipótese de cliticização dessa forma, pois entre um clítico nominativo e o verbo não deve ocorrer qualquer elemento (Ramos, 1997, p.55). [...] Por sua vez, essa não-contigüidade levou Andrade (2004) a afirmar que *cê* ainda não seria propriamente um clítico, mas um pronome fraco (2004, p. 129). Já Barbosa (2005) afirma que a não-contigüidade comprova que *cê* não é clítico.*(PERES, 2006, p. 143)

Quanto ao fator linguístico referência – definida ou indefinida – o estudo indica que *cê* é a forma preferida pelos falantes para indicar seja referência definida (79,8% dos casos), seja indefinida (65,7%). Contudo, a observação do peso relativo mostra que a referência definida favorece a forma *cê* (peso relativo de 0.64) e desfavorece a forma *você* (peso relativo 0.36). Por outro lado, quando a referência é indefinida, a forma *você* é favorecida, com peso relativo de 0.63 e *cê* é desfavorecido, com peso relativo de 0.37. Mesmo analisando em separado os dados da função sujeito quanto à referência, e ainda observando o fator faixa etária a conclusão da autora é de que:

*a forma *cê* é usada preferencialmente para indicar referência definida, sendo a forma *você* deixada para expressar referência indefinida. Entretanto, analisando-se o comportamento de cada faixa etária, vê-se um crescimento da forma *cê* também para expressar referência indefinida. [...] Em suma, os resultados apontam para a expansão do uso da forma *cê* e sua conseqüente vitória sobre a forma *você*.*(PERES, 2006, p. 149)

Com respeito ao fator focalização, os dados coletados evidenciaram que as situações de não-focalização representam 99,2% dos dados. Neste caso, os pesos relativos da forma *cê* e da forma *você* indicam que nenhuma delas é favorecida. A forma *ocê* por sua vez não foi considerada significativa pelo Programa. Quanto aos casos de focalização, as análises indicam que a forma *você* é a forma padrão e a mais utilizada nos casos de foco enfático. Ela também é fortemente favorecida, com peso relativo 0.93.

⁵⁷ Para os detalhes da análise, veja-se Peres (2006, p. 137-138).

Há um único caso, citado pela autora, de foco com a forma *cê*, o enunciado seguinte (PERES, 2006, p. 152):

(E 117) *Eu vou matar CÊ, fdp!*

Peres (2006) se mostra surpreendida por esta ocorrência, pois a focalização implicaria: 1) a pronúncia do item com intensidade maior; 2) portanto, maior lentidão. Mas sendo a forma *cê* reduzida, deveria, ao contrário, ser pronunciada mais rapidamente. Por último, *cê*, no enunciado, é objeto direto, o que é um fator desfavorecedor.

A autora admite não saber solucionar essa questão senão supondo que, por apresentar um início de mudança, *cê* estaria ampliando gradualmente seus contextos de uso.

Enfim, são expostos os resultados das análises dos fatores extra-linguísticos: idade e gênero. Quanto ao primeiro, idade, os dados mostram que *cê* é mais presente que *você* em todas as faixas etárias. A observação das diferenças percentuais e dos pesos relativos leva a estudiosa a afirmar que há uma mudança em curso, pois a preferência pelo uso por parte dos falantes de todas as idades da forma *cê* mostraria que esta está em expansão.

Quanto ao fator gênero, as mulheres favorecem a forma *cê*, com peso relativo 0.57, enquanto os homens aquela *você*, com idêntico peso relativo. Analisando as diferenças etárias e de classe social e seus relativos pesos, a autora conclui que:

a forma cê não sofre preconceito quanto ao seu uso na linguagem informal em Belo Horizonte, haja vista que as mulheres a favorecem.(PERES, 2006, p. 160)

O estudo prossegue com a análise do *corpus de 1982*. O total de ocorrências das formas *você*, *ocê* e *cê* não diverge muito daquele de 2002, sendo 15,9% das ocorrências totais referentes a *você*, 6,6% a *ocê* e 77,5% a *cê*.

A tabela abaixo ilustra este quadro (PERES, 2006, p. 170):

TABELA 21
Total de ocorrências de *você, ocê e cê* - corpus de 1982

Formas	Ocorrencias	%
Você	93	15,9
Ocê	39	6,6
Cê	455	77,5
Total	587	

Quadro 3: Tabela de ocorrências do *corpus* de 1982 em PERES, 2006, p. 170.

O programa *Goldvarb 2001* selecionou nesse caso somente três fatores como significativos:

- I. linguístico: Função Sintática;
- II. extralinguísticos: Idade e Classe Social.

Em vista destes parâmetros, a autora procedeu à análise. Os resultados da distribuição das três formas conforme a Função Sintática são os seguintes:

- I. na função de Sujeito a forma mais utilizada é *cê* (81,2% das ocorrências), seguida por *você* (13,4%) e *ocê* (5,4%);
- II. na função de Objeto de Verbo a forma mais utilizada é *você* (66,7%), seguida por *ocê* (33,3%), enquanto *cê* não apresenta nenhuma ocorrência;
- III. na função de Objeto de Preposição a forma mais utilizada é *você* (64%), seguida por *ocê* (32%) e *cê* (4%).

Também neste *corpus* a função não-marcada é aquela de sujeito para as três formas, apresentando 95,2% do total de ocorrências. Nessa função, *cê* é a forma amplamente mais utilizada. Contudo, a análise dos pesos relativos evidencia que seu favorecimento é somente um pouco mais provável do que as outras formas, apresentando peso relativo de 0.54, enquanto na forma *ocê* o peso relativo é de 0.48 e na forma *você* de 0.46.

Na função de objeto de verbo, a forma mais utilizada pelos informantes é *ocê*, apesar de a probabilidade de ocorrência ser de 0.90, menor do que aquela de *você*, de 0.97. O que mais chama a atenção é o fato que, em números absolutos, foram encontradas somente três ocorrências das formas nesta função, e a total ausência de *cê*⁵⁸.

Na função de objeto de preposição, a forma *você* é a mais frequente, com 64% das ocorrências. Do ponto de vista da probabilidade, as formas mais favorecidas são *você* e *ocê*,

⁵⁸ Este dado nos parece um pouco estranho pois em nossa pesquisa foram encontrados dois casos de *cê* objeto direto, e as outras formas ocorreram nesta função em várias situações.

ambas com peso relativo 0.99. A probabilidade da ocorrência de *cê* é baixíssima, com peso relativo de 0.01. Contudo, uma única ocorrência dessa forma nessa função foi encontrada, o que é indicativo de que já em 1982, em um processo muito inicial, *cê* mostrava tender para outros contextos que não somente de sujeito (PERES, 2006, p. 174):

(E 161) *Eu vou lá olhar p'cê.*

Em seguida são expostos os resultados das análises dos fatores extra-linguísticos. O fator Idade tem por objetivo verificar se há uma mudança em curso. Os dados de 1982 mostram que a forma *cê* já se fazia presente em Belo Horizonte/MG, mas em geral seu uso era equilibrado em relação a *ocê* e *você*.

O fator Classe Social mostra que a forma *cê* era utilizada tanto pela classe baixa quanto pela classe média. No entanto, a probabilidade de ocorrência desta forma varia: levemente favorecida no caso da classe média, com peso relativo de 0.58 e desfavorecida no caso da classe baixa, com peso relativo de 0.38. O comportamento contrário é verificado para a forma *você*: suas ocorrências são baixas em ambas as classes sociais, mas especialmente na média, onde também é levemente desfavorecida, com peso relativo de 0.42. A forma *ocê* não foi considerada significativa pelo Programa. Segundo a pesquisadora, esses resultados, um tanto quanto estranhos, podem ser reflexo do Paradoxo do Observador, como foi citado anteriormente.

Para tentar entender melhor a distribuição das formas nas várias funções, Peres (2006) procede com a análise sob o ponto de vista do contexto fonético. Sua hipótese é que o uso de *você*, *ocê* e *cê* procure adaptar cada forma aos padrões fonéticos e fonológicos do PB em relação à estrutura silábica, acentuação e ritmo.

O teste dessa hipótese é aplicado nos dois *corpora*, utilizando a totalidade dos dados. Na função de Sujeito não foram encontradas novidades e os índices mostraram que não há variação no favorecimento das formas. A função sintática Objeto de Preposição foi a que apresentou, após a de sujeito, o maior número de ocorrências, para um total de 70 usos. A autora analisa todos os casos, dividindo-os por preposição: *para*, com variações em “pra” e [p]; *com*, inclusive nas variações [kũ] e [c]; *de*, sem variações, que, todavia, a autora admite serem possíveis; *em*, com a variação “n”; *a*, sem variações e *por*, também sem variações⁵⁹.

⁵⁹ Como os exemplos são muitos, apesar de interessantes, não vamos aqui citá-los, remetendo-nos à bibliografia: Peres (2006, p.185-188).

Foram efetuados testes acústicos⁶⁰ para verificar a efetiva pronúncia, especialmente nos caso de epêntese, assimilação da labialização, elisão das vogais e encontros consonantais.

Os resultados confirmam a possibilidade de *cê* ocorrer como objeto de preposição e, apesar das poucas ocorrências, Peres (2006) sublinha sua importância como índice da expansão dessa forma.

Quanto à função Objeto de Verbo, as ocorrências foram pouquíssimas nos dois corpora, somente quinze, das quais seis (40%) com a forma *você*, seis (40%) com *ocê* e três (20%) com *cê*⁶¹.

Os casos relacionados pela autora são reportados abaixo (PERES, 2006, p. 191-192):

a) Com a forma *você*:

(E 18/82) “Tia, gostei de ter **você** de novo.”

(E 548/82) “Eu tô querendo **você** mesmo.”

(E 97/02) “(...) e nunca larguei **vocês** por isso.”

(E 437/02) “Eles podem te... levar **vocês** lá na zona rural.”

(E 1288/02) “Eu passo e nem sou<do> ([!sowdu] = saúdo) **você**.”

(E 1418/02) “(...) matar sua mãe e **você**”.

b) Com a forma *ocê*:

(E 117/82) “Cê cê confunde muit’**ocê** com a outra pessoa.”⁶²

(E 98/02) “Nunca larguei **ocês**.”

(E 167/02) “(...) mas às ve<ze>s te jog’**ocê** pro buraco.”

(E 172/02) “(...) ela acab<a> te jogand’**ocê** pro buraco.”

(E 188/02) “(...) te jog’**ocê** pra trás”.

(E 1396/02) “(...) Ah, deix’**ocê**.”

c) Com a forma *cê*:

(E 117/02) “Eu vou ma:ta: **cê**, fdp!”

(E 142/02) “(...) dinheiro suj<o> te lev’**cê** à morte (...)”

⁶⁰ A autora não especifica os tipos de testes a que foram submetidos os dados utilizados em sua pesquisa, somente afirma terem sido realizados no Laboratório de Fonética da Faculdade de Letras da UFMG.

⁶¹ Os exemplos de *cê* como objeto de verbo são os mesmos citados anteriormente.

⁶² Neste exemplo não é claro se se trate de um caso de *retracting*, do qual discutiremos adiante. Em nosso estudo quando o falante repete a mesma palavra, esta é contada somente uma vez, como parece que foi feito também por Peres.

(E 614/02) “Cê vai assistir o futebol cê ... ne<go> te arras<ta> cê lá dentro.

Foi efetuada uma análise quanto à estrutura silábica das formas nos enunciados em relação à estrutura silábica e ao ritmo do PB.

A estudiosa conclui que:

parece não haver restrições de ordem fonética ou fonológica à ocorrência de cê como objeto de verbo ou de preposição, haja vista que essa forma apareceu exercendo essas funções sintáticas nos dados do estudo.(PERES, 2006, p. 194)

Segundo Peres (2006), as expressões que contenham preposição ou verbo, seguidas de uma das três formas, poderão ocorrer com maior facilidade se os processos fonológicos atuarem para facilitar sua produção, tornando-as mais “econômicas” e também não marcadas⁶³. De fato, segundo a autora, é o contexto fonético que acaba facilitando ou favorecendo a ocorrência de determinada forma.

A análise final, uma comparação de todos os fatores significativos nos resultados dos dois *corpora*, leva a estudiosa a afirmar que o processo linguístico que envolve as três formas estudadas reflete um processo de mudança que, enquanto tal, ocorre dentro de um contexto linguístico e social que não podem ser desprezados. Peres (2006) não quer entrar na discussão se *cê* é um clítico ou um pronome fraco, mas mostrar que *cê*, aos poucos, está aparecendo em contextos onde os clíticos não apareceriam, ou seja, isso é revelador de um início de mudança.

2.3.3. Síntese das propostas para as formas *você(s)/ocê(s)/cê(s)*

Os estudos apresentados acima são somente os mais significativos dos muitos que demonstram a complexidade em se analisar as formas *você(s)/ocê(s)/cê(s)*. Como foi visto, as opiniões são divergentes sob muitos aspectos, seja pelo tipo de abordagem, seja pelos arcabouços teóricos que as embasam, seja pela metodologia. Nesta seção, tentaremos sintetizar ao máximo quais são as principais conclusões a que os estudiosos chegaram:

- I. segundo alguns autores, *cê* pode ocorrer somente em posição de sujeito pré-verbal (VITRAL, 1996; RAMOS 1997; VITRAL E RAMOS, 1999, 2006 e 2008);

⁶³ *Ibidem*, p. 195.

- II. outros admitem que as funções em que *cê* ocorre estão se expandindo, para aquelas de objeto de verbo e objeto preposicionado (BARBOSA, 2005; PERES, 2006);
- III. *cê* recebe, de acordo com cada autor, estatutos diferentes: para alguns é clítico (VITRAL, 1996; RAMOS 1997) ou em processo de cliticização, na fase de formação de expletivos (VITRAL E RAMOS, 1999, 2006 e 2008; CIRÍACO *et al.*, 2004); para outros forma fraca (PETERSEN, 2008); para outros ainda uma palavra plena (BARBOSA, 2005); enquanto há quem prefira não tentar definir seu estatuto e analisar os dados sob vários pontos de vista, seja sociolinguísticos, seja fonético e fonológico (PERES, 2006) ou refletir sobre seu comportamento sintático e prosódico, através de uma análise distributiva, chegando à conclusão de que se trata de um novo pronome (OTHERO, 2013).

A proposta que será apresentada neste trabalho é justamente aquela de partir dos dados para concluir sobre os aspectos que essas três formas possuem e as funções que elas desempenham. Isso, como será visto no próximo capítulo, será feito através da análise de *corpora* de fala espontânea, que servirão como base para uma metodologia de análise acústica que pretendemos implementar.

2.4. Análise do vernáculo florentino

Esta pesquisa desdobra-se em duas partes: a primeira aborda a questão das formas pronominais de segunda pessoa no PB e sua redução, como foi explicado acima. A segunda parte analisa, seguindo a mesma metodologia, mas de forma sincrônica e diacrônica, a questão dos pronomes sujeito no vernáculo florentino. Será possível, através desse paralelo, observar o comportamento de uma língua em que ainda não há estabilidade na expressão das formas, encontrando-se em plena fase de mudança, o PB, e aquele de outra língua, o vernáculo florentino, em que as mudanças já estão consolidadas.

No próximo parágrafo, será apresentada brevemente a história da língua italiana, no tocante à diferenciação entre a mesma e os vernáculos, com ênfase no florentino. Em seguida, trataremos da questão da expressão do sujeito e da terceira pessoa do singular.

2.4.1. Breve história da língua italiana e do vernáculo florentino

A história linguística do italiano está diretamente ligada ao vernáculo florentino. Sem nos delongarmos nos séculos de passagem do latim aos vulgares locais do território italiano, será aqui abordado, muito sinteticamente, o momento de surgimento da língua literária, e da separação entre florentino e italiano.

O século XIII é o período em que os vernáculos se afirmam definitivamente, em relação ao latim, como língua não somente falada, mas também escrita. Dentre tantos exemplos e experiências escritas que desabrocham, aquelas que predominam são em vulgar florentino. Os motivos são vários, mas podem-se identificar dois deles como principais:

- I. o grande poder econômico (mercantil e financeiro) da Toscana e, a partir do final do século, da cidade de Florença;
- II. a enorme densidade da cultura escrita na Toscana, e em Florença (BARTOLI LANGELI, 2000).

Os dois motivos são ligados entre si devido à escola de *abaco e algorismo*, uma escola voltada para a alfabetização das classes mercantis que se preocupava em ensinar a escrita em vulgar e a matemática financeira (ULIVI, 2000; LUCCHI, 1982; DAVIS, 1988; ARRIGHI, 1973; 1985; 1987). Essa experiência, comum na Toscana da baixa idade média, é desconhecida no resto da Europa (MANACORDA, 1914), onde a alfabetização continua sendo em latim e restrita a um número muito pequeno de pessoas. Contrariamente, na Toscana medieval, grande parte da população era alfabetizada (GRENDLER, 1991). Ao ponto que, segundo a *Cronica* de Villani de 1338, muito mais da metade das crianças em idade escolar de sexo masculino frequentava a escola de *abaco* (AQUILECCHIA, 1979). Esse contexto sociocultural é responsável pela impressionante quantidade de produção escrita, principalmente em vulgar, por parte da Toscana, durante os séculos XIII e XIV, incomparável com aquela de qualquer outra região europeia. Basta pensar que 8/9 de todos os manuscritos editados até 1390 e escritos até 1375 na Itália inteira são toscanos e 50% florentinos (POGGI SALANI, 1992). Esses números mostram com clareza como a escrita na Toscana era uma atividade de massa e como isso tornou a Toscana um lugar excepcional do ponto de vista cultural na época. Para entender a força de Florença ainda melhor, é útil observar que até o começo do século XIV os lugares mais evoluídos na Toscana, tanto economicamente quanto culturalmente, são as cidades da área ocidental (Pisa, Lucca, Pistoia), e que somente a partir do começo do séc. XIV Florença se torna o centro econômico-cultural principal (CASTELLANI, 1997; GENSINI, 1988). Se nos limitarmos, portanto, somente ao séc. XIV, a produção escrita florentina adquire uma proporção ainda maior (MANNI, 1994; 2003).

As famosas *Tre Corone*, ou seja Dante, Petrarca e Boccaccio, não devem portanto ser consideradas como causas do sucesso do florentino e sim um produto natural de um movimento muito amplo e rico de escritores. Em um panorama tão vasto de dedicação à escrita, é muito mais provável que surjam também algumas personalidades de valor intelectual superior.

Mas é somente dois séculos mais tarde, no primeiro quarto do século XVI, que o florentino é reconhecido como língua de referência literária para todo o território italiano. Esse processo se deu de uma maneira muito singular e, contrariamente ao que aconteceu para o idioma de Lisboa, o idioma de Paris e aquele de Madrid (PRINCIPATO, 2008; BRUNOT, 1987), ocorreu sem a participação do poder político. Em ausência de um centro político, na Itália, a discussão sobre a unificação linguística foi sempre conduzida através de uma dialética entre o que se pode considerar o centro linguístico, ou seja Florença, e a periferia que, dependendo dos momentos históricos, varia entre os diferentes centros políticos e culturais da Itália: a Sicília de Frederico II (ALFIERI, 1992), os movimentos religiosos da Umbria e da Itália central (PETRUCCI, 1992; CASAPULLO, 1999; BRUNI, 1983; BOLOGNA, 1982; DELCORNO, 1999; LIBRANDI, 1994; RUSCONI, 1994; BALDELLI, 1983), as principais cortes da planície do rio Po (VITALE, 1953; BONGRANI, 1986; DRUSI, 1995; MATARRESE, 1988), Veneza, Milão, Roma (PALERMO, 1991; TRIFONE, 1992), Napoli (SABATINI, 1993).

O momento crucial para o reconhecimento do florentino como língua de referência, por enquanto somente para a literatura, é o começo do século XVI. Nessa fase a discussão sobre o italiano literário (VITALE, 1978; TROVATO, 1995; MARAZZINI, 1992; TESI, 2001) adquire um novo nível de maturidade e três posições se enfrentam: a posição assim chamada de *cortigiana*, liderada por Trissino e Castiglione (DRUSI, 1995); a posição do florentinismo vivo, liderada, entre outros, por Machiavelli (CASTELLANI POLLIDORI, 1978); e a posição purista, liderada pelo humanista veneziano Pietro Bembo. A posição *cortigiana* achava que o italiano literário devia emergir através do uso dos intelectuais das cortes italianas, que eram muitas e com idiomas variados. O paradigma proposto era aquele dos três vértices: latim, toscano e vulgar local; a coincidência (ou maior proximidade) de dois desses três vértices deveria decidir sobre o uso, prosseguindo caso a caso. Os florentinistas sustentavam que o italiano literário devia se pautar pelo uso vivo de Florença, que no começo do séc. XVI era muito distante do idioma de dois séculos antes, pelas razões que veremos. Finalmente, havia a posição de Bembo, expressada de maneira detalhada nas *Prose della vulgar lingua* (MORGANA; PIOTTI; PRADA, 2000). O ensaio foi publicado somente em

1525, mas circulava já há duas décadas e, portanto, pode ser considerado a primeira gramática italiana, apesar da publicação em 1516 da gramática do Fortunio, um outro vêneto. Bembo, como um bom humanista, transferia para a língua vulgar o princípio da *imitatio* válido para o latim: assim como para essa língua os modelos eram Virgílio para a poesia e Cícero para prosa, para o italiano literário os modelos deviam ser o *Canzoniere* de Petrarca para a poesia e o *Decameron* de Boccaccio para a prosa (PATOTA, 1992; BRANCA 1976; 1996). Então, o italiano literário devia se modelar em duas obras florentinas, mas de dois séculos antes, ou seja, modelos que para os florentinos representavam uma língua morta.

A proposta de Bembo de fato acaba por se impor na Itália toda, menos na Toscana e principalmente em Florença. O sucesso da proposta se deve principalmente a dois fatores:

1. o fato de Bembo fornecer uma norma clara e de fácil aquisição para todos. Não seria possível para os não florentinos adquirir o uso vivo de Florença, nem seria possível para a maioria dos literatos italianos participar do debate linguístico das cortes e negociar um processo de unificação a partir de tantos centros;

2. a aliança com um outro veneziano, mesmo que somente de adoção, Aldo Manunzio, o maior tipógrafo europeu do século. Vale lembrar que Veneza foi por todo o século XVI e por grande parte do XVII o maior centro editorial da Europa. Para a nascente indústria editorial, era vital poder contar com uma norma linguística e ortográfica única, que permitisse economia de custos e certeza de uso. Foram de Manunzio as primeiras edições de Boccaccio e Petrarca (BARTOLI LANGELI e INFELISE, 1992; LOWRY, 1979; TROVATO, 1991; VIANELLO, 1992).

É a partir desse momento que é possível falar de *lingua italiana* e de dialetos do italiano (ALINEI, 1981). De fato, é só então que se cria uma hierarquia sociolinguística entre um idioma nacional, mesmo que somente para uso escrito literário (uma língua italiana como sistema social de comunicação deverá esperar até a segunda metade do século XIX), e idiomas locais socialmente subordinados à língua e que, portanto, devem ser considerados dialetos, na definição que os linguistas da Europa continental dão a esse termo. Antes desse momento, a variedade de idiomas própria da península itálica não configurava uma hierarquia: todos os idiomas locais possuíam o mesmo *status* e eram chamados de vulgares.

O italiano, portanto, deve ser visto substancialmente como "filho" do florentino dos séculos XIII-XIV. O destino do italiano e do florentino coincidem, substancialmente, até 1350, mas a partir dessa época se separam. O que acontece com o florentino?

Pelo menos duas experiências históricas devem ser evidenciadas:

1. A peste de 1348 (aquela narrada no *Decameron*), junto com a expansão econômica e política de Florença, produzem um grande fenômeno migratório de outras áreas da Toscana para a capital. Isso traz para Florença usos linguísticos muito diferentes da língua usada por Boccaccio e Petrarca (SGRILLI, 1988; POGGI SALANI, 1992; GENSINI, 1988). São usos próprios da Toscana ocidental, da Toscana meridional (Arezzo e Siena) e principalmente das áreas rurais e de montanha, com especial referência ao Mugello, região de origem da família Medici, novos senhores da cidade. Uma comparação entre a prosa de Boccaccio e a de Machiavelli mostra com evidência a distância entre dois usos no mesmo registro alto, aquele que utiliza palavras ou expressões elegantes, mas não afetadas. A comparação entre obras de registros mais baixos é ainda mais clara.
2. Já na primeira metade do século XIV, na Toscana (único lugar na Europa), inicia-se um debate sobre a origem dos idiomas vulgares e sobre a relação recíproca entre eles. Os nomes principais desse debate são Flávio Biondo, Leonardo Bruni, Lorenzo Valla e Leon Battista Alberti (GRAYSON, 1964; PATOTA, 1999; RUGGIERO, 1994; TAVONI, 1984; 1992). Durante a segunda metade do século XV, na fase do domínio de Lorenzo *Il Magnifico*, Florença tenta uma política expansionista, que vê na língua um símbolo importante, em uma tentativa de fazer o que havia sido feito, ou estava sendo feito, nas grandes monarquias europeias. Os intelectuais da corte de Lorenzo são chamados a um esforço para organizar um verdadeiro projeto de política linguística (CARDINI, 1974; TAVONI, 1992): uma antologia crítica da literatura florentina, traduções de clássicos latinos em florentino, poemas dedicados aos maiores reinantes da Itália e formulários de textos para as chancelarias. Na carta de Lorenzo (mas na realidade de Poliziano) que abre a *Raccolta Aragonese*, o seguinte trecho bem explica as intenções do autor:

E potrebbe facilmente [la nostra lingua materna], nella juventù e adulta età sua, venire ancora in maggior perfezione, e tanto più aggiungendosi qualche prospero successo del fiorentino e aumento

*imperio: come si debbe non solamente sperare, ma con tutto l'ingegno
e forse per li buoni cittadini aiutare*⁶⁴.(POLIZIANO, 1477)

O texto é considerado por alguns a fonte do Nebrija, quando, na primeira gramática do espanhol, em 1492, diz que *la lengua siempre fué compañera del império*.

A política de Lorenzo não atingiu seu objetivo, mas fortaleceu ainda mais na Toscana a própria identidade linguística, e, conseqüentemente, a distância com o modelo de italiano literário que logo após a morte dele era destinado a se difundir pela península. Muitos são os traços linguísticos que a partir da metade do século XIV mudaram o florentino áureo (como é chamada a língua das *Tre corone*) para o florentino de prata, como é chamado o florentino do século XIV (MANNI, 1979). Entre eles, e apenas para o registro mais alto, (para o registro baixo, veja-se TROLLI, 1972), houve as seguintes mudanças (MANNI, 1979; PALERMO, 1992; MANNI, 1994):

1. na morfologia, algumas desinências verbais do indicativo, do subjuntivo e do condicional; mudanças nos particípios passados; mudança no sistema dos artigos, nos pronomes pessoais, nos numerais, nos possessivos, em vários advérbios;
2. na fonética, principalmente com o aparecimento da *gorgia*. Trata-se de uma aspiração em diferentes graus das oclusivas não vozeadas em posição fraca, que é uma posição intervocálica também morfossintática (CONTINI, 1960), e a perda do ditongo espontâneo. Esta perda permaneceu como traço característico do italiano contra todos os dialetos da península (que possuem ditongos metafônicos) e que a língua herdou do florentino áureo;
3. na sintaxe, onde, entre outros fenômenos, aparece uma novidade tipológica importante e central para esta tese: mudam-se as modalidades de expressão do pronome sujeito, que se torna obrigatório em grande parte dos contextos e em muitos deles gera a reduplicação (PALERMO, 1997);
4. no léxico há termos específicos utilizados somente no florentino.

No final do século XIX o italiano recebe uma nova florentinização, como consequência da política linguística pós-manzoniana (SERIANNI, 1990), mas a grande parte dos fenômenos do assim chamado de *florentino de prata* (a fase posterior a 1350) nunca entrará no italiano.

⁶⁴ *E a nossa língua materna poderia facilmente, sendo jovem e adulta, se tornar ainda mais perfeita, sobretudo com o acréscimo de sucessos favoráveis do florentino e o aumento do poder territorial: coisa que todos os bons cidadãos devem não só esperar, mas ajudar com toda a inteligência e a força (tradução nossa).*

2.4.2. Pronomes pessoais sujeito e terceira pessoa singular no vernáculo florentino

Um dos pontos de maior divergência entre italiano moderno e vernáculo florentino refere-se à questão dos pronomes sujeito. Retomamos aqui a discussão e os estudos que desde Boström (1972), passando por Brandi e Cordin (1981, 1983, 1989), Renzi e Vanelli (1983) e Renzi (1983, 1992), entre os mais importantes, discutiram sobre esse problema.

Os dialetos do norte da Itália e o vernáculo florentino, mas não outros vernáculos toscanos, apresentam divergências substanciais em relação ao italiano moderno⁶⁵. Este último possui uma única série pronominal sujeito, tônica, não obrigatória e motivada pragmaticamente, sendo considerada uma língua *Pro-Drop*⁶⁶.

Os pronomes sujeito em italiano são os seguintes:

PESSOA	SINGULAR	PLURAL
1 ^a	<i>io</i>	<i>noi</i>
2 ^a	<i>tu</i>	<i>voi</i>
3 ^a	<i>egli/ella, esso/essa, lui/lei</i> ⁶⁷	<i>essi/esse, loro</i> (masc. e fem.)

Quadro 4: Os pronomes pessoais sujeito em italiano.

Giannelli (2000), em um estudo sobre a Toscana e seus dialetos, indica no vernáculo florentino uma dupla série de pronomes sujeito: uma tônica e outra átona.

⁶⁵ Referimo-nos especificamente aos estudos sobre o comportamento do sujeito no trentino e no florentino nos estudos de Brandi (1981), Cordin (1981) e sobre esses dois dialetos e o italiano em Brandi e Cordin (1981); os estudos sobre vários dialetos do norte da Itália e da Toscana em Renzi e Vanelli (1983) e o estudo sincrônico e diacrônico sobre o sujeito pronominal italiano e florentino em Renzi (1983, 1992).

⁶⁶ O termo, redução de *Pronoun-dropping*, indica as línguas em que os pronomes sujeito, assim como própria expressão do sujeito, podem ser omitidos por serem inferidos pela desinência verbal.

⁶⁷ Não vamos entrar aqui na discussão sobre os usos que diferenciam as duplas *egli/ella, esso/essa* e *lui/lei*. Somente para breve efeito explicativo, o uso das duas séries (respectivamente para sujeitos humanos e não humanos) é cada vez mais limitado a contextos escritos de diafasia muito alta. A dupla *lui/lei*, inicialmente utilizada como forma oblíqua, na língua falada é a mais utilizada como sujeito animado, e seu uso é cada vez mais difundido também na forma escrita.

Giannelli (2000, p. 38-39):

PRONOME TÔNICO EM FUNÇÃO DE SUJEITO	PRONOME ÁTONO EM FUNÇÃO DE SUJEITO
<i>io</i>	<i>e</i> ⁶⁸ (pode introduzir todas as pessoas, com ressalvas na terceira do sing.)
<i>te</i>	<i>tu</i>
<i>lui</i> (masculino), <i>lei</i> (feminino)	<i>gl'</i> (masculino antes de vogal), <i>e</i> (masculino antes de consoante), <i>l'</i> (masculino antes de vogal)
<i>noi</i>	<i>la</i> (feminino), <i>l'</i> (feminino antes de vogal)
<i>voi</i>	<i>e</i> quando acompanha outro clítico
<i>loro</i>	<i>gl'</i> (masculino antes de vogal), <i>e</i> (masculino antes de consoante)
	<i>le</i> (feminino) <i>l'</i> (feminino antes de vogal)

Quadro 5: os pronomes pessoais sujeito no vernáculo florentino.

Ele afirma que os pronomes átonos podem acompanhar os tônicos e serem, por sua vez, acompanhados por *e*, que dá uma “nuance enfática” ao enunciado.

Encontramos a possibilidade de *l'* como masculino singular antes de vogal nos *corpora* consultados. Contudo, não há, na bibliografia, afirmações sobre esse caso, que, todavia, é citado em vários exemplos dos autores⁶⁹.

Portanto, em italiano há uma única série de pronomes pessoais sujeitos, todos tônicos, que são facultativos, pois sua motivação é pragmática⁷⁰. No florentino, há uma dupla série, uma clítica e outra tônica. Segundo Renzi (1983, 1992), ela é obrigatória⁷¹. A posição de tais pronomes é sempre antes do verbo, e às vezes pode haver mais de um pronome, clítico ou não, anteposto ao verbo. Esse fenômeno é chamado de “reduplicação pronominal”. A posição imediatamente antes do verbo pode ser preenchida somente com clíticos, nunca com outro item lexical, que pode, ao contrário, intervir entre dois clíticos. O esquema, portanto, é o seguinte (RENZI, 1983, p. 224):

(pron, SN) X cl V

⁶⁸ No texto utilizaremos a grafia *e*, sem apóstrofo, a mesma adotada nos *corpora* utilizados e nos escritos de muitos dos estudiosos, apesar de alguns preferirem a grafia com apóstrofo, *e'* (BINAZZI, 1997, 1998, 2007, 2009 e GIANNELLI, 2000). Nos textos mais antigos e nas citações dos mesmos é possível encontrar o *e'*, com apóstrofo, que, todavia, refere-se a uma outra origem pronominal (veja-se RENZI, 1983, p.226).

⁶⁹ Esta possibilidade nos foi, contudo, confirmada por Binazzi em comunicação pessoal.

⁷⁰ Sobre os casos em que o sujeito é obrigatório, veja-se: Palermo (1997 e 2011).

⁷¹ Segundo Renzi (1983, p. 224) as formas homográficas aos pronomes tônicos no italiano, podem, em certas condições, serem átonos do ponto de vista fonético. Os clíticos são sempre átonos mas, segundo Renzi, sua definição não é baseada em um conceito fonético, mas sintático.

Os dois termos entre parênteses são opcionais e se excluem reciprocamente; *X* indica a possibilidade de inserção de material lexical; *cl* indica o clítico e *V* o verbo.

Esse esquema será ilustrado com alguns exemplos fictícios na terceira pessoa do singular, aquela que será analisada nesta pesquisa:

(37) **Lui** oggi **l'**ha mangiato.

Ele (tônico) hoje ele (clítico) comeu.

(38) **I' bambino** oggi **l'**ha mangiato.

O menino (SN) hoje ele (clítico) comeu.

(39) **Lei** oggi **la** preparerà il pranzo.

Ela (tônico) hoje ela (clítico) preparará o almoço.

(40) **La ragazza** oggi **la** preparerà il pranzo.

A garota (SN) hoje ela (clítico) preparará o almoço.

No exemplo (37), o pronome sujeito *lui* é retomado pelo clítico sujeito *l'* (que pode ser utilizado somente diante de vogal). No exemplo (38), é o SN *i' bambino* que é retomado pelo clítico *l'*. Já os exemplos (39) e (40) ilustram a mesma situação no feminino, antes com um pronome tônico e em seguida com um SN, retomados pelo clítico sujeito *la*⁷².

Renzi (1983) afirma que este esquema inclui também o caso da primeira pessoa do plural (RENZI, 1983 p. 224):

(noi) **si** canta⁷³

(nós) clítico impessoal canta

em que o pronome entre parênteses é opcional, e o verbo é conjugado na terceira pessoa do singular.

Há, contudo, contextos em que não ocorre a reduplicação do clítico, e o florentino apresenta o mesmo comportamento do italiano. Os casos são os seguintes:

⁷² Em italiano as frases acima seriam gramaticais somente como: (37) (*Lui*) oggi ha mangiato; (38) *Il bambino* oggi ha mangiato; (39) (*Lei*) oggi preparerà il pranzo; (40) *La ragazza* oggi preparerà il pranzo. O material entre parêntese é opcional.

⁷³ Em italiano o verbo deve ser conjugado obrigatoriamente com a primeira pessoa do plural, portanto: 1) *Noi cantiamo*; ou 2) *Ø cantiamo*. Este caso específico é bem peculiar, pois *si*, em florentino, além da possibilidade de indicar um clítico impessoal, também é correferente de *noi*, indicando a reduplicação clítica, de que foi dito. Veja-se Renzi (1983 e 1992) e Giannelli (2000).

- I. na terceira pessoa singular masculina, também no caso de sujeito nominal, quando o verbo começa com consoante:

(lui, i bambino) canta

(ele, o menino) canta

- II. na terceira pessoa do singular, usada com sujeito posposto e com verbo que começa com consoante:

*viene le bambine*⁷⁴

vem (sing.) as meninas

mas, no caso de verbo iniciado por vogal, há expressão do clítico:

*l'è venuto le bambine*⁷⁵

(clítico) veio (sing, sem concordância de gênero) as meninas

O contexto com a negação, *'un*, também apresenta um comportamento peculiar: o pronome clítico é imediatamente precedido por ela. Por possuir uma posição fixa, a própria negação se torna um clítico (RENZI, 1983, p. 225). Ilustramos isso com outros exemplos fictícios:

(41) **(Lui)** *'un l'ha cantato.*

Ele (tônico) não (clítico neg.) ele (clítico sujeito) cantou.

Neste caso o pronome sujeito *lui* é duplicado pelo clítico *l'*, precedido pela negação *'un*.

Há uma exceção para a terceira pessoa do singular no feminino, cuja ordem é invertida:

⁷⁴ Em italiano o verbo deve ser conjugado obrigatoriamente com a terceira pessoa do plural e, quando há o auxiliar *essere*, com concordância de gênero e número, portanto: *Vengono le bambine*.

⁷⁵ Como acima: *Sono venute le bambine*.

(42) **(Lei) la ‘un** canta.

Ela (tônico) ela (clítico) não (clítico neg.) canta.

Neste caso o sujeito pronominal *lei* é retomado pelo clítico *la*, desta vez seguido pela negação *‘un*.⁷⁶

Para finalizar, é necessário apresentar o clítico *e*. Alguns estudiosos afirmam que essa forma é encontrada na fala atual florentina somente em caso de “ênfase” ou na terceira pessoa do plural⁷⁷. Segundo Brandi (1981) e Brandi e Cordin (1981), o clítico *e* preenche o lugar do clítico pessoal ausente, no caso da primeira pessoa do singular, de forma que a inteira série pronominal esteja completa, assim como apresentado também por Giannelli (2000). Renzi (1983) afirma que sua função como sujeito é historicamente bastante recente, mas seu uso nunca é obrigatório. Segundo o estudioso, essa forma pode ocorrer com todas as pessoas e, sempre de forma facultativa, pode ser encontrada com os verbos impessoais, meteorológicos e na construção de sujeito posposto. Uma de suas peculiaridades é que, quando na oração ocorre *e*, não aparece o pronome tônico, mas no caso do clítico pode haver reduplicação. Abaixo exemplificamos seus usos:

(44) *e* parlo; (*e*) **tu** parli; (*e*) **la** parla; (*e*) **‘un** parliamo; (*e*) **vu** parlate; *e* parlano.⁷⁸

clit. falo; (clit.) tu (clit.) falas; (clit.) ela (clit) fala; (clit.) não (clit. neg.) falamos; (clit.) vós (clit.) falais; (clit.) falam

(45) (*e*) **fa** caldo; (*e*) **bisogna** fare presto.⁷⁹

(clit.) faz calor; (clit.) precisa ser rápido

(46) (*e*) **piove**; (*e*) **tuonò**.⁸⁰

(clit.) chove; (clit.) trovejou

(47) (*e*) **la** mi fa male **la** testa.⁸¹

(clit.) ela (clit.) me doi a cabeça (SN)

⁷⁶ Em italiano as duas frases seriam: (41) *(Lui) non ha cantato*; (42) *(Lei) non canta*.

⁷⁷ Binazzi, comunicação pessoal.

⁷⁸ Em italiano: *(io) parlo*; *(tu) parli*; *(lei) parla*; *(noi) non parliamo*; *(voi) parlate*; *(loro) parlano*.

⁷⁹ Em italiano: *fa caldo*; *bisogna fare presto*.

⁸⁰ Em italiano: *piove*; *tuonò*.

⁸¹ Em italiano: *mi fa male la testa*.

(48) *E gl'è andato a casa.*⁸²

clit. ele (clit.) foi para casa

(49) *E la 'un capisce niente.*⁸³

clit ela (clit) não (clit. neg.) entende nada

Os parênteses nos exemplos acima indicam opcionalidade. No exemplo (44), a forma *e* é sujeito de todas as pessoas: no caso da primeira do singular o único clítico existente é *e*. Na segunda do singular o clítico é *tu*, pois em florentino a forma tônica é *te*. Na terceira pessoa ocorre o clítico feminino *la*. Na primeira pessoa do plural, não sendo possível utilizar a forma tônica, devido à presença do *e*, o clítico reduplicado é a negação *'un*, em caso de enunciado negativo. No caso de positivo a única opção é o pronome tônico ou um SN. A forma *vu* indica o clítico sujeito da segunda pessoa do plural; e a terceira pessoa do plural utiliza como forma clítica, diante de vogal, *e*.

Os exemplos (45) e (46), respectivamente, mostram *e* como sujeito facultativo de verbos impessoais e metereológicos.

No exemplo (47), há a construção com o sujeito posposto: o clítico *e* é facultativo, seguido pelo clítico sujeito *la*, reduplicado pelo sujeito posposto *la testa*.

Os exemplos (48) e (49) mostram a reduplicação do clítico masculino (48) com as formas *e* e *gl'*, ambos sujeito. Ocorre a mesma situação no feminino (49) com as formas *e* e *la*, sujeitos, seguidas pela negação clítica *'un*.

O panorama dos pronomes que foi apresentado, em específico dos de terceira pessoa singular, mostra-se bastante complexo. Todavia, nota-se que, à exceção do caso do *e*, há regras claras a serem seguidas. O caso do clítico *e* é bastante peculiar por se tratar de um pronome que substitui várias pessoas e pelo seu uso mais recente, motivo pelo qual parece não haver ainda um consenso pleno entre os estudiosos sobre seu conjunto de regras de uso.

As formas a serem pesquisadas, cuja vitalidade será avaliada em *corpus*, inicialmente do ponto de vista sincrônico e depois diacrônico são, portanto:

- I. A presença de sujeito \emptyset ou de sujeito expresso;
- II. a presença ou ausência da expressão *e*, e em quais casos ocorre;
- III. para o masculino, as formas *gl'* e *l'*;
- IV. para o feminino, as formas *la* e *l'*;

⁸² Em italiano: *(Lui) è andato a casa.*

⁸³ Em italiano: *(Lei) non capisce niente.*

- V. a presença da reduplicação pronominal clítica;
- VI. para todas as formas, a coocorrência de SN, pronome tônico ou outra forma pronominal.

2.4.3. A escolha do pronome de terceira pessoa singular

A discussão acima serviu para mostrar como, no caso do vernáculo florentino, apesar da evolução contínua da língua, as regras dos pronomes sujeito parecem bastante claras. Nosso intuito é verificar em *corpus* se isso é verdadeiro, tanto do ponto de vista sincrônico quanto do ponto de vista diacrônico. Em uma tese como esta, que abraça dois sistemas linguísticos, é necessário efetuar escolhas sobre o que é representativo e o que possa servir como base para generalizações. Por isso, como o fizemos para o PB, optando por analisar o pronome que tradicionalmente é considerado o mais interessante pela riqueza dos problemas que põe, também no vernáculo florentino houve uma atenta seleção de qual forma pronominal é mais útil para o objetivo atual e para eventuais extensões futuras a outras formas.

A partir da literatura consultada e de discussões com nossa coorientadora na Itália⁸⁴, a escolha recaiu sobre as formas pronominais de terceira pessoa do singular, masculina e feminina, ou seja: *gl'*, *e*, *la*, *l'*.

Segue a justificativa:

- I. porque essas são as formas mais frequentes;
- II. porque representam boa parte dos problemas do sistema pronominal como um todo;
- III. portanto, porque possibilitam, mais do que outras formas, generalizações úteis que possam servir como base na investigação de outras formas.

Para melhor averiguar as ocorrências e suas distribuições nos *corpora* consultados, optou-se por fazê-lo com apenas quatro lemas verbais de alta frequência. Isso porque, como será relatado abaixo, verificou-se ser demasiadamente longo e pouco produtivo o processo de análise de todos os lemas verbais disponíveis; portanto, preferiu-se uma análise mais aprofundada, mas limitada às ocorrências relativas a um único verbo. Assim sendo, foram escolhidas a forma *è*, terceira pessoa do singular do presente do indicativo do verbo *essere* (ser), ou seja a forma mais frequente do verbo mais frequente e mais polissêmico, e as formas

⁸⁴ Queremos aqui agradecer não somente a ajuda e as indicações fornecidas por Emanuela Cresti, mas também aproveitamos para manifestar nossa gratidão a Massimo Moneglia, Alessandro Panunzi e Lorenzo Gregori, além de Neri Binazzi.

fa, *va* e *viene*, terceiras pessoas do singular do presente do indicativo, respectivamente dos verbos *fare* (fazer), *andare* (ir) e *venire* (vir), os verbos mais frequentes após *essere*.

2.4.4. O verbo *essere* (ser) e suas variações semânticas no italiano: motivos da escolha

Na definição da melhor metodologia de análise das formas pronominais de terceira pessoa no vernáculo florentino, foram efetuados alguns testes preliminares. Como a pesquisa baseia-se em *corpora*, não havia como prever:

- I. a expressão ou não do sujeito, isto é, não se podia ter certeza de antemão se o sujeito estaria explicitado para todas as formas verbais, pois os falantes poderiam seguir as regras do italiano de não obrigatoriedade;
- II. no caso de expressão do sujeito: se ele poderia ser representado por um pronome tônico, um SN, um sintagma nominal que contém uma oração, um pronome relativo, sujeito posposto, entre outras possibilidades.

Foi efetuada uma varredura inicial de teste no *Corpus Stammerjohann*, que será apresentado adiante: selecionaram-se 940 formas verbais (lemas). Como a etiquetagem morfossintática não discrimina a pessoa, foi verificada manualmente uma amostragem de 480 ocorrências verbais na terceira pessoa do singular. Esse teste avaliou:

- I. se era possível analisar todas as formas verbais;
- II. quanto seria proveitoso este trabalho para efeito de se ter uma noção do uso efetivo dos pronomes;
- III. como averiguar a explicitação ou não do sujeito, e em que casos poderiam ocorrer;
- IV. qual poderia ser a melhor alternativa de pesquisa.

O resultado mostrou que uma análise efetuada dessa maneira seria extremamente complexa: para cada lema aparecia todo o quadro de formas verbais em números variados; portanto, o estudo se delongaria de maneira desmedida. Além disso, para vários lexemas foi constatado que não havia ocorrência de uso das formas buscadas, enquanto outros mostraram-se mais significativos.

Decidiu-se então inicialmente optar por um único verbo e uma única forma verbal que pudesse dar conta de todos os possíveis casos de uso. Assim como optou-se, na escolha do pronome, por aquele mais produtivo, o de terceira pessoa do singular, na escolha do verbo fez-se o mesmo.

O verbo *essere* (ser) é o segundo lema de maior frequência encontrado no C-ORAL-ROM italiano e o primeiro entre os lemas verbais. O estudo *corpus-driven* de Panunzi (2010) sobre a variação semântica do verbo na fala detectou cerca de 14.000 ocorrências do mesmo nesse *corpus*. Esse número equivale a 4,6% do total de *tokens* do C-ORAL-ROM italiano; *essere* é o verbo mais frequente da língua italiana.

Não é somente em termos quantitativos que o verbo *essere* é interessante para as finalidades desta pesquisa; do estudo de Panunzi (2010) emergiram também importantes questões sobre o uso do verbo.

A primeira distinção a se fazer é que *essere* pode ser utilizado como auxiliar e como verbo propriamente dito. A segunda distinção refere-se ao uso específico do verbo *essere* dentro de um lexema autônomo, o verbo *esserci*. Há, portanto, de se diferenciar entre o uso de dois lemas distintos⁸⁵.

Os principais usos verbais de *essere* são: copulares, identificativos e predicativos com base locativa. Segundo Panunzi (2010), tais usos mostram características semânticas qualitativamente diferentes, previsíveis com base em traços positivos equivalentes do complemento que o acompanha. Tais interpretações são as seguintes (PANUNZI, 2010, p. 47-62):

- I. *essere* com uso copular apresenta ausência de referencialidade no complemento⁸⁶:

(16a) *CIC: 'un è mica brutto //⁸⁷ (ifamcv14)

não é nada feio

- II. *essere* com uso identificativo apresenta referencialidade, mas com ausência de um papel temático⁸⁸:

(27.c) *CUC: il mercato è questo // (imedts07)

o mercado é este

⁸⁵ Veja-se a discussão em Panunzi (2010, 95-133).

⁸⁶ Panunzi (2010, p. 110).

⁸⁷ Reportamos as transcrições dos exemplos (extraídos do C-ORAL-ROM) com a segmentação em unidades prosódicas (a barra simples indica fronteira de unidade tonal e a barra dupla fronteira de unidade tonal que coincide com fronteira de enunciado).

⁸⁸ Panunzi (2010, p. 116).

- III. *essere* com uso predicativo com base locativa apresenta o complemento cuja referencialidade está ligada ao papel temático de lugar⁸⁹:

(37.a)*LIA: *anche qui / siamo a Versailles // (ifamcv01)*

aqui também estamos em Versalhes

Para esclarecer, o verbo *essere* significa ter identidade, característica própria ou intrínseca, apresentar-se em determinada situação, estar, ter existência real, existir, ter ou ocupar lugar. O verbo, portanto, pode ser traduzido com o português “ser” ou “estar”.

Já *esserci* é um verbo pronominal, em que um clítico, *ci*, ligou-se ao verbo. A partícula *ci* é neste caso advérbio de lugar. A acepção primordial de *esserci* é “existir” e, como consequência, “estar em um lugar”.

A categoria dos verbos pronominais, termo genérico atribuído a qualquer verbo em cujo lema ocorra um clítico a ele ligado de forma indissolúvel, torna as formas verbais cristalizadas. Exemplo disso são as expressões apresentacionais como “*In casa non c’è nessuno*” (Em casa não há ninguém), onde o clítico perde sua propriedade de advérbio para formar um todo com o verbo⁹⁰.

Para completar o quadro, há uma variação marcada no uso verbal de *essere*: isso ocorre quando o verbo não desempenha uma função ligada à sua variação semântica primária, mas:

*desempenha papéis de operador em gramaticalização ou de elemento lexical no interior de uma construção fixa, de um fraseológico ou de um uso estereotipado.*⁹¹(PANUNZI, 2010, p. 90)

Exemplos de quanto dito acima são os seguintes enunciados:

Panunzi (2010, p. 152):

(18.c) *LIA: *mah / sarà // (ifamcv01)*

bom, será (verdade)

⁸⁹ Panunzi (2010, p. 122).

⁹⁰ A esse respeito, vejamos os estudos de Nocentini (1977), Lo Cascio (1970), Jezek (2004).

⁹¹ Tradução nossa.

(18.f) *SIL: *non è bello ciò che è bello* //(ifamcv12)

Não é belo o que é belo

A escolha do verbo *essere* parece-nos, portanto, adequada para esta pesquisa porque:

- I. trata-se de um verbo de ampla utilização;
- II. seu campo semântico cobre vários usos;

Em relação à escolha do tempo verbal optou-se pelo presente do indicativo na terceira pessoa do singular. A forma è:

- I. é aquela mais frequente entre os lexemas de *essere*;
- II. sendo uma forma curta, supõe-se que provoque o preenchimento fonético e, portanto, o uso dos clíticos.

2.4.5. Outros verbos

Como a forma pesquisada do verbo *essere* impossibilitaria o uso do clítico *e*, e inibiria talvez aquele de *la*, devido ao encontro vocálico, optou-se por escolher alguns outros verbos que pudessem facilitar a produção de todo o grupo de clíticos estudados. Dentre os verbos de mais alta frequência da língua italiana, foram descartados alguns, como *dire* (dizer), o segundo em frequência, mas muito utilizado na fala florentina de maneira pragmática, como uma forma fixa. O mesmo aconteceu após analisar o uso de *vedere* (ver), também utilizado pragmaticamente na fala espontânea.

A escolha recaiu sobre três verbos de alta frequência, com os quais se esperava não encontrar grandes limitações em seu uso: o transitivo *fare* (fazer) e os intransitivos *andare* (ir) e *venire* (vir).

O verbo *fare* cobre uma ampla variedade semântica: por ser um verbo de ação positiva bastante indefinido é utilizado quando se quer dar concretude à ação; por outro lado é também empregado de forma genérica em expressões que dissimulam uma ação mais precisa e concreta ou expressam valor estático como “*fa caldo*”, isto é, “está quente, faz calor”. É utilizado também de forma intransitiva acompanhado por um complemento predicativo como “*fa l’avvocato*”, isto é, “é advogado”, “pratica a advocacia.”

O verbo *andare* é o principal verbo de movimento de distanciamento. Pode indicar seja o início da ação, seja a intenção do afastamento: “*vado*”, ou seja, “estou indo” e “*e allora, andiamo?*”, isto é, “e aí, vamos indo?”. Além de uma série de usos em expressões fixas, pode ser utilizado como auxiliar na forma passiva, substituindo *essere*.

O verbo *venire* é o verbo do movimento em direção oposta a *andare*. Os dois verbos podem ser considerados complementares nas ações de deslocamento. Esse verbo também pode ser utilizado como auxiliar na forma passiva, substituindo *essere*.

A análise destes três verbos seguiu o mesmo procedimento adotado com *essere*, e o modo e tempo verbal analisados são os mesmos, a terceira pessoa singular do indicativo presente.

3. METODOLOGIA: ANÁLISE PARALELA *CORPUS-BASED* SOBRE DUAS LÍNGUAS

3.1. Análise do PB

O panorama da linguística de *corpus* no Brasil vem se ampliando e se fortalecendo nos últimos anos. Contudo, a maioria dos *corpora* aqui produzidos é constituída por *corpora* escritos, especialmente com material provindo de jornais, textos científico-acadêmicos e *sites* da *internet*. Grandes projetos, como aqueles coordenados por Berber Sardinha⁹², têm disponibilizado material de estudo; no entanto, a produção de *corpora* eletrônicos orais no Brasil ainda é limitada.

É necessário fazer algumas distinções preliminares. A primeira é entre banco de dados de fala e *corpora* eletrônicos. Os bancos de dados são compostos por gravações que não seguem uma arquitetura de coleta, sendo na prática um repositório de informações. Os *corpora* eletrônicos seguem uma série de princípios pré-estabelecidos na coleta para serem representativos de um determinado universo, visam garantir qualidade acústica e diversificação de tipologias. Mais importante, são coletados e compilados de forma que possam ser consultáveis eletronicamente através de uma série de *softwares* que leem as etiquetas que geralmente são colocadas, podendo ser de tipo morfossintático, prosódico, informacional, entre outros, além das metainformações sobre falantes e situações, dependendo da finalidade do *corpus*.

Os *corpora* podem ser constituídos por gravações padronizadas, que seguem roteiros pré-estabelecidos e são efetuadas em cabines acústicas, ou por gravações de situações em contexto natural. As primeiras garantem uma alta qualidade do som e são ideais para estudos prosódicos e de fonética acústica, entre outros. Contudo, a fala é extremamente controlada e não espontânea, o contrário do que acontece nas situações de contexto natural. Nestas, a interação e a acionalidade entre os participantes é o fator priorizado, as situações são variadas e a estruturação da fala é imprevisível. As vantagens para a análise da fala espontânea são muitas, pois, dessa forma, torna-se concreto obter um tipo de produção o mais próximo possível do que é produzido naturalmente pelos falantes. As desvantagens são que a análise acústica torna-se mais difícil, pois sempre haverá ruídos de fundo e sobreposições de fala.

⁹² Falamos aqui do *Corpus Brasileiro*, com um bilhão de palavras e disponível online (<http://corpusbrasileiro.pucsp.br/x/>) e o *Banco do Português*, parcialmente online em <http://www2.lael.pucsp.br/corpora/bp/>

Enfim, há de se diferenciar entre os *corpora* compostos por textos em que há variação diafásica, e aqueles em que esta não está presente. Nos primeiros, a preocupação na arquitetura é diversificar ao máximo os contextos e o repertório de temas sobre o que é falado. Isso permite obter, mesmo com *corpora* de dimensões médias⁹³, uma boa amostragem dos vários macrofenômenos da estruturação da fala. Os *corpora* em que a variação diafásica não está presente, geralmente, são específicos para algum tipo de estudo como, por exemplo, aqueles sobre textos acadêmicos, jornalísticos, situações em sala de aula, entre outros.

Mello (2012) detalha os *corpora* de língua falada mais importantes existentes no Brasil⁹⁴. É citado o *Projeto NURC* (Norma Linguística Urbana Culta), que a partir dos anos 1970 agregou pesquisadores e universidades nas principais capitais do Brasil, como Rio de Janeiro, São Paulo, Recife, Porto Alegre e Salvador. Os resultados e as produções bibliográficas foram, e em alguns casos continuam sendo, variados para as diferentes equipes. Todavia, os arquivos e as transcrições não são disponibilizados, exceto em alguns casos. Outros *corpora* e projetos citados são: o *Projeto Discurso e Gramática*, da UFRJ; o *VARPORT: Análise Contrastiva de Variedades do Português*, sempre da UFRJ; o *PROFALA: Variação e Processamento de Fala e do Discurso: análise e aplicações*, promovido pela Universidade Federal do Ceará; o *Projeto Variação Linguística no Estado da Paraíba – VALPB*; o *Projeto Vertentes do Português Popular do Estado da Bahia*, da UFBA; o *VARISUL: Variação Linguística na Região Sul do Brasil*, constituído por um consórcio de equipes de três universidades federais do sul do Brasil (UFRGS, UFSC e UFPR); o projeto *Amostra Linguística do Interior Paulista – ALIP: O Banco de dados IBORUNA*, financiado pela FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo).

O que caracteriza estes *corpora* é que seu desenho em geral segue a vertente da sociolinguística variacionista, às vezes com implementações de cunho gerativista. Outra característica geral é que cada um acompanha uma modalidade de transcrição, geralmente seguindo um padrão de tipo ortográfico, mas não há qualquer tipo de anotação de aspectos pragmáticos ou prosódicos. Enfim, o mais importante para a

⁹³ Sobre as dimensões dos *corpora*, veja-se Sardinha, 2004.

⁹⁴ Remetemos ao capítulo *Os corpora orais e o C-ORAL-BRASIL*, Mello (2012), o detalhamento minucioso que a autora faz de todas as informações disponíveis sobre cada *corpus*.

análise da fala, raramente a parte de áudio é disponibilizada e somente em um caso houve a preocupação de um alinhamento formal som-texto⁹⁵.

A realização de um *corpus* oral de fala espontânea que obedeça a critérios rigorosos de balanceamento e representatividade, uma alta qualidade acústica e transcrição alinhada ao áudio é algo que comporta um enorme dispêndio de tempo e recursos. Para tal, é necessária uma equipe bem treinada, a realização de testes e validações estatísticas, o estudo das técnicas de coleta de dados, a autorização por parte dos informantes para o uso de suas contribuições e a preservação de seu anonimato, equipamentos profissionais, *softwares* que permitam o alinhamento texto-som e uma interface viável de acesso ao *corpus*.

3.1.1. Escolha do *corpus*: descrição do C-ORAL-BRASIL

O único *corpus* com as características realmente adequadas ao tipo de pesquisa aqui descrito é o *corpus* C-ORAL-BRASIL (RASO & MELLO, 2012). Trata-se de um *corpus* de fala espontânea do PB, com foco na diatopia mineira, especialmente a da região metropolitana de Belo Horizonte⁹⁶. Em relação aos outros *corpora* citados na seção de revisão da literatura, ou disponíveis para consulta, este se distingue por possuir uma arquitetura que privilegia a variação diafásica e por se apresentar com o alinhamento do texto ao som e ao espectro acústico. Isto permite atingir uma “representação mais fiel do universo” (MONEGLIA, 2005, p. 11⁹⁷) linguístico e a possibilidade de realizar análises (verificáveis) da fala enquanto processo e não somente nas transcrições.

Biber (1993) afirma que, para itens frequentes, como boa parte dos pronomes, excertos de cerca de 1000 palavras fornecem uma amostragem suficientemente confiável. O C-ORAL-BRASIL apresenta textos que variam entre um mínimo de 700 até um máximo de 4500 palavras, com uma média de 1500 palavras por texto, constituindo, portanto, um total muito maior do que aquele considerado suficiente por Biber. O objetivo da variação diafásica foi, portanto, “diversificar ao máximo os contextos situacionais para se aproximar mais da representação de um universo aberto” (RASO, 2012, p. 61).

O *corpus*, publicado em sua parte informal, é composto por 208.130 palavras, divididas em 139 textos. Estes foram coletados entre 2007 e 2011, em sua maioria entre 2008

⁹⁵ O Varsul disponibiliza uma amostra digital em seu site.

([HTTP://chu.com.br/varsul/downloads/projeto_amostra_digital_varsul.pdf](http://chu.com.br/varsul/downloads/projeto_amostra_digital_varsul.pdf)) composta por gravações em formato .wav acompanhadas pelas transcrições e com opção de ligação entre o som e a transcrição através do *software Oceanaudio*.

⁹⁶ Adiante esta questão será abordada em maiores detalhes.

⁹⁷ Tradução nossa.

e 2010⁹⁸. As gravações foram efetuadas com os equipamentos mais modernos à disposição e com microfones sem fio⁹⁹, de modo a permitir uma interação altamente espontânea entre os informantes, e uma alta qualidade acústica.

O *corpus* é composto, para cada texto, por:

- I. um arquivo de som, em formato *.wav*; sua transcrição, em formato *.rtf*; um arquivo *.xml*, que serve para o alinhamento áudio-texto, utilizável através do software *WinPitch* (MARTIN, 2005); a transcrição em formato *.txt*;
- II. metadados, que contemplam desde o título, o local e data de coleta, a situação comunicativa, duração, número de palavras, qualidade acústica e todas as possíveis informações sobre os informantes, além dos dados sobre o pesquisador que fez a coleta, os transcritores e revisores;
- III. um arquivo etiquetado léxico-morfossintaticamente em versão integral e simplificada para usos diversos.

O conjunto do *corpus*, além disso, apresenta uma série de listas de frequência e medidas para pesquisa, completado por planilhas e estatísticas sobre os informantes.

A transcrição do *corpus* segue o sistema *CHAT* (MACWHINNEY, 2000), implementado com a anotação prosódica (MONEGLIA & CRESTI, 1997), sendo segmentado em unidades tonais¹⁰⁰.

As características descritas acima permitem que o pesquisador possa realizar buscas de vários tipos: em nosso caso foi possível investigar as formas *você(s)/ocê(s)/cê(s)* através de *softwares* específicos e obter em poucos segundos todas as ocorrências com seus respectivos contextos. De cada uma foi possível conseguir informações sobre os falantes, o tipo de interação, a situação e o local de gravação.

Outro aspecto que foi importante na escolha desse *corpus* é seu alto grau de informalidade, já que o foco da pesquisa é o estudo da fala espontânea. A maior parte dos textos e das situações comunicativas, cerca de 70%, ou 159.364 palavras, refere-se a interações de caráter familiar ou particular, enquanto uma parte bem menor, de 48.766 palavras, foi reservada ao contexto público. A macrodivisão do *corpus* é, portanto, entre contexto familiar/particular e contexto público. Dentro destes, há três tipologias de interação:

⁹⁸ A parte formal do *corpus* está em fase de coleta e transcrição.

⁹⁹ Foram utilizados gravadores Marantz PMD660 Professional Solid State Recorder, microfones omnidirecionais Sennheiser MD 421-II 4 e Shure PG58-XLR e microfones de lapela Sennheiser ME 4 além de um mixer Behringer XEYX 1222 FX.

¹⁰⁰ Sobre a segmentação, ver adiante.

- I. monológica: implica a prevalência de um texto falado que é desenvolvido mesmo apesar das intervenções de um eventual interlocutor. Os turnos, portanto, são muito longos e os enunciados são mais articulados informacionalmente. Há pouca referência ao contexto situacional e há necessidade de contextualização;
- II. dialógica: com interação entre dois falantes, é constituída por turnos curtos, enunciados simples, com grande referência ao contexto situacional e ampla variação ilocucionária;
- III. conversacional: possui as características da tipologia dialógica com a interação ocorrendo entre três ou mais informantes.

Quanto ao número de palavras para cada tipo de contexto/interação, os monólogos somam 52.116 palavras no contexto familiar/particular e 16.222 no contexto público; os diálogos 55.361 no primeiro contexto e 17.997 no segundo; e, por fim, as conversações 51.887 palavras no contexto familiar/particular e 14.547 palavras no contexto público.

O *corpus* é formado por 139 textos de cerca de 1500 palavras; 20 textos menores e 10 textos de mais de 2000 palavras.

Exemplos de textos monológicos presentes no *corpus* são: explicação profissional, relato de vida, conto, piada, fábula, texto argumentativo. Já os textos dialógicos e as conversações dependem muito da atividade que os interlocutores estão desempenhando e que são as mais variadas. Há, portanto, situações como: conversa entre amigos, interação de um corretor de imóveis com a cliente enquanto procuram um apartamento, diálogo entre vendedor de sapatos e cliente, jogo de futebol entre amigos, interação de um garçom servindo em uma festa, instruções de um professor de ginástica ao aluno ou interação entre mãe e filho recém-nascido que chora, entre muitas outras.

O aspecto interessante de tais situações é que:

Cada situação estimula a emergência de atos de fala diferentes, tamanho diferente dos turnos, tamanho e estruturas diferentes dos enunciados, presença maior ou menor de silêncios etc. (RASO, 2012, p. 63)

Contudo, os aspectos fundamentais e decisivos para a escolha deste *corpus* em relação a outros são:

- I. a alta qualidade acústica por ele apresentado;
- II. a alta confiabilidade na transcrição;

- III. o alinhamento áudio-texto, que permite uma real análise da fala, tendo-se a possibilidade de ouvir a transcrição de qualquer trecho de interesse do pesquisador.

Quanto ao ponto (I) é importante ressaltar que os contextos naturais em que foram efetuadas as gravações são aqueles em que é mais difícil se obter uma boa qualidade acústica. Contudo, esta é fundamental para um *corpus* de fala espontânea, pois deve ser possível não somente realizar transcrições dos aspectos morfológicos e sintáticos, mas também poder estudar as características fonéticas, seja por si, seja pela importância que elas têm na estruturação pragmática. Para poder obter uma alta qualidade acústica, além dos equipamentos de última geração dos quais nos referimos anteriormente, foram planejadas cuidadosamente as estratégias de gravação de situações específicas. As próprias gravações duraram um tempo maior em relação àquele utilizado no *corpus*, de modo a poder escolher o material que possuía a melhor qualidade, e muitas das gravações não entraram no *corpus* por não apresentarem a qualidade acústica exigida.

Os textos presentes no *corpus* são divididos com base na qualidade acústica com as seguintes etiquetas, que mostram a relação com os parâmetros de clareza ao fim da transcrição, quantidade de sobreposição, ganho, ruído de fundo e, principalmente, computabilidade da curva de f_0 :

- I. A: qualidade muito alta;
- II. AB: qualidade alta;
- III. B: qualidade média;
- IV. BC: qualidade baixa;
- V. C: qualidade baixa, em que pelo menos 60% da curva de f_0 é computável de maneira confiável.

Do total de 139 textos, 40 possuem qualidade A e 43 AB: isso significa que cerca de 60% das gravações caracterizam-se como ótimas do ponto de vista acústico. Quanto aos textos com qualidade baixa, somam um total de 32: geralmente a qualidade é mais baixa nas conversações (onde as sobreposições aumentam muito) e em gravações que retratam situações possíveis somente em ambientes com ruído de fundo (como, por ex., um supermercado); esses textos foram escolhidos unicamente por seu alto interesse do ponto de vista diafásico ou diastrático.

Em relação ao ponto (II) temos que lembrar a alta qualidade da transcrição e as sucessivas validações a que foi submetida. Uma etapa de transcrição bem elaborada é fundamental para um *corpus*, sobretudo para um *corpus* de fala espontânea. No caso do C-

ORAL-BRASIL, houve uma longa fase de preparação e testes, além de treinamentos específicos com os transcritores¹⁰¹, para que fossem estabelecidos os parâmetros a serem seguidos nesta etapa. As implementações metodológicas que foram introduzidas, já que não há uma metodologia consolidada para a transcrição da fala, baseiam-se sobre dois alicerces fundamentais:

- I. a segmentação da fala segundo unidades de referência próprias, os enunciados e as unidades tonais, que fazem parte do arcabouço teórico da *Language into Act Theory, L-AcT* (CRESTI, 2000)¹⁰²;
- II. a necessidade de marcar na transcrição os fenômenos morfológicos, sintáticos e lexicais candidatos a estarem em curso de gramaticalização ou lexicalização, que uma transcrição plenamente ortográfica não seria capaz de capturar.

Tudo isso sem prejudicar a legibilidade do texto.

Os transcritores foram orientados a seguirem o mais fielmente possível a informação acústica, seja no nível segmental, seja naquele suprasegmental da fala. Os critérios adotados seguem aqueles do C-ORAL-ROM com o acréscimo de formas para casos que não estavam previstos. Isto acarretou uma dúplici convenção: para algumas formas, que foram elencadas, seguem-se os acordos ortográficos vigentes, para outras foram estabelecidas convenções que marcam:

- I. ruídos paralinguísticos;
- II. hesitações e palavras interrompidas;
- III. onomatopeias;
- IV. interjeições e exclamações;
- V. siglas e acrônimos;
- VI. numerais;
- VII. palavras estrangeiras e erros de pronúncia;

¹⁰¹ Os detalhes sobre a formação dos transcritores e os treinamentos aos quais foram submetidos são descritos em Mello *et al.* (2012).

¹⁰² Não vamos aqui nos aprofundar sobre a *Language into Act Theory (L-AcT)*, mas para melhor compreensão do que será dito adiante, esclarecemos que nela a unidade de referência da fala é o *enunciado*. Este corresponde ao que Austin (1962) definiu um ato de fala: um ato locutivo que seja a contraparte linguística da ilocução, ou da ação, feita quando se pronuncia um enunciado. A prosódia é, portanto, a interface entre o ato locutivo e o ato ilocutivo. Os dois, concomitantemente, realizam um ato de fala. A prosódia é fundamental para veicular a complexidade interna dos enunciados. Estes podem ser formados por uma única unidade tonal, os *enunciados simples*, ou por duas ou mais unidades, os *enunciados complexos*. Segundo a *L-AcT* as diferentes unidades tonais transmitem diferentes unidades informacionais. Veja-se Cresti (2000); Raso (2012).

Algumas palavras não foram transcritas por serem incompreensíveis, sendo marcadas com xxx. Outras palavras foram censuradas para preservar o anonimato dos informantes.

Em relação a fenômenos em via de gramaticalização ou lexicalização, estes foram devidamente assinalados no *corpus*, segundo critérios não ortográficos. Os itens registrados de maneira não ortográfica são:

- I. fenômenos de aférese;
- II. fenômenos relativos à conjugação verbal;
- III. palavras sem marca de plural;
- IV. fenômenos relativos a redução pronominal;
- V. contração e redução de preposições articuladas;
- VI. contrações de preposições e pronomes;
- VII. cliticização ou enfraquecimento da forma de negação;
- VIII. ausência de verbo nas construções interrogativas, supressão do pronome relativo e pseudo-relativo;
- IX. contrações diversas das formas *Senhor* e *Senhora*;
- X. diminutivos apocopados;
- XI. intensificador *mó*;
- XII. casos de rotacismo.

É importante, enfim, lembrar o sistema de validações adotado no *corpus*. Após uma primeira revisão das transcrições, houve uma validação de 5% de cada texto por parte de dois transcritores, que verificaram se havia qualquer erro. Esta primeira validação procurou os seguintes tipos de erro;

- I. elisão de palavras;
- II. inserção de palavras;
- III. palavras transcritas de forma errônea.

Foram então realizados controles à parte para algumas categorias de fenômenos. A primeira validação apresentou uma margem de erro médio de 1,87%, bem abaixo daquela preestabelecida de 5%, geralmente considerada estatisticamente aceitável. Em seguida à revisão final, a porcentagem de erros totais caiu para 0,81%. A validação foi feita para cada um dos fenômenos não ortográficos, sendo que o fenômeno que mostrou a porcentagem de erro mais alta alcançou apenas 3,5% das ocorrências.

Para as finalidades de nossa pesquisa, portanto, o *corpus* se mostrou altamente confiável. Em específico, a possibilidade que o C-ORAL-BRASIL oferece de busca de várias formas pronominais foi valiosa.

Quanto aos pronomes pessoais, os critérios adotados para a transcrição levaram a diferenciar (com base na percepção acústica) uma dupla série pronominal para as segundas e terceiras pessoas: uma série plena, que segue a ortografia tradicional, e uma reduzida (*cê/cês* e *ocê/ocês* para a segunda pessoa; *e'/es* e *ea/eas* para a terceira pessoa). Quanto à primeira pessoa, a opção foi por não prever nenhuma forma reduzida, dada a dificuldade em distinguir de forma perceptível e segura as formas plenas daquelas reduzidas, pela pequena consistência acústica das formas.

No que se refere à validação das formas relativas aos pronomes pessoais, após a segunda revisão, o percentual de erro foi de 0,43%. O *corpus* mostra-se deste modo altamente confiável para a pesquisa.

Graças a tal recurso, identificar as ocorrências das formas indagadas neste estudo através de um programa de pesquisa automaticamente tornou-se uma tarefa viável, e possibilitou despendar mais tempo na parte de análise acústica e morfossintática das mesmas.

O ponto (III), o alinhamento, ou seja, a perfeita correspondência entre o som e o texto, e, neste caso, também ao espectrograma, é o fator que permite verdadeiramente o estudo dos principais fenômenos da fala e é uma das características distintivas deste *corpus* em específico. Quando se possui um *corpus*, mesmo que acompanhado pelo som, mas não alinhado, o que ocorre é que na realidade acaba-se por analisar um texto escrito (que tenha a oralidade apenas como fonte), perdendo-se, assim, todas as informações do sinal acústico fundamentais no estudo da língua falada. Segundo Raso (2012):

não é possível estudar a fala sem se recorrer constantemente às informações acústicas, sem se verificar sobre elas qualquer interpretação e sem categorias linguísticas formadas a partir da observação da modalidade que se quer estudar, e não a partir de outra modalidade. (RASO, 2012, p. 58)

Para finalizar esta apresentação do *corpus*, vamos discorrer algumas palavras descrevendo a cuidadosa escolha dos informantes, outro fator que diferencia este *corpus* daqueles utilizados em trabalhos anteriores.

É importante esclarecer que todos os falantes assinaram um *Termo de Compromisso* aprovado pelo COEP (Comitê de Ética) da UFMG que preserva sua

identidade sob forma de siglas. Somente o pesquisador responsável conhece sua verdadeira identidade.

O total de informantes é de 362. Destes:

- I. 138 são de Belo Horizonte;
- II. 89 provêm de outras cidades do Estado de Minas Gerais, especialmente da região metropolitana de Belo Horizonte;
- III. 19 são de outros estados brasileiros;
- IV. 2 são estrangeiros;
- V. de 114 não foi possível estabelecer a proveniência, que todavia pronunciam menos de 3% do total de palavras.

Os informantes foram subdivididos também por gênero, idade, grau de educação, e profissão. O balanceamento por gênero é praticamente perfeito, pois 203 informantes são do sexo feminino e 159 do sexo masculino, sendo apenas um falante, que pronuncia uma única palavra, de sexo não identificado. Do total de palavras do *corpus*, 50,36% são pronunciadas por mulheres e 49,64% por homens, confirmando um balanceamento ideal.

Quanto ao critério idade, os informantes foram separados por grupos:

- I. A: idade entre 18 e 25 anos;
- II. B: idade entre 26 e 40 anos;
- III. C: idade entre 41 e 60 anos;
- IV. D: idade acima de 60 anos;
- V. M: menor de idade;
- VI. X: idade não conhecida.

O grupo A representa 27,13% do número de palavras pronunciadas no *corpus*; o grupo B pronuncia 30,28% das palavras; o grupo C, 31,01% das palavras; o grupo D, 8,05% das palavras; os menores de idade pronunciam 1,61% das palavras e aqueles de idade desconhecida 1,91% das palavras.

Quanto ao critério educação, os parâmetros seguidos são aqueles expressos nas faixas abaixo:

- I. 1: escola primária não completada ou analfabeto, 46 informantes;
- II. 2: até o nível de graduação que não utilize o diploma para seu trabalho, 101 informantes;
- III. 3: graduado que utilize o título para seu trabalho, 104 informantes;
- IV. X: nível de educação desconhecido.

Os falantes da faixa 1 pronunciam 15,78% das palavras do *corpus*; aqueles da faixa 2, 40,76%; os da faixa 3, 40,66% e os falantes dos quais não é conhecido o nível de escolaridade pronunciam 2,79% das palavras do *corpus*.

A descrição das profissões utilizou um termo para cada tipo de ocupação.

O balanceamento do *corpus* é muito bem equilibrado quanto aos fatores sexo e idade, um pouco menos quanto ao nível de escolaridade. De fato, pode-se afirmar que o C-ORAL-BRASIL é mais representativo do nível escolar médio e alto, apesar de o nível baixo também estar presente.

Quanto à ocupação profissional, assinalada como categoria aberta, é de mais difícil generalização. Todavia, a observação dos metadados mostra que muitos dos informantes são pessoas ligadas ao ambiente da educação, como professores e estudantes, isto devido seja ao método de coleta, seja ao fato que entre os jovens a principal ocupação é o estudo. Contudo, a variedade de outros tipos de ocupação é muito grande.

3.1.1.1. Análise do *corpus*

A grande novidade que é proposta neste trabalho é avaliar as formas *cê(s)/ocê(s)/você(s)* no uso real, através de um *corpus* de fala espontânea e de uma análise prosódica das ocorrências, o que até este momento não foi feito, ou quando foi feito, não partiu de ocorrências presentes em diferentes diafásias de contexto natural.

Isto significa ter a possibilidade de estudar a variação linguística dos falantes assim como é produzida e, especificamente, poder:

- I. encontrar eventuais formas de uso que a gramática tradicional desconhece ou não prevê;
- II. encontrar ocasionais valores funcionais novos e não conhecidos;
- III. colher dados fundamentais para futuras generalizações.

Como foi apresentado acima, o C-ORAL-BRASIL mostra-se confiável e suficientemente representativo para uma pesquisa sobre a mudança do sistema pronominal. A segunda pessoa, conforme a literatura, está sofrendo um forte processo de redução, passando da forma plena *você(s)*, àquela *ocê(s)* e à *cê(s)*. Posto que o interesse desta primeira etapa é:

- a) verificar as ocorrências dessas formas;
- b) verificar seus contextos de ocorrência;
- c) observar a coocorrência das três formas e suas restrições;

É importante antes de passar aos dados, esclarecer alguns pontos.

3.1.1.2. Revisão da transcrição

Trabalhar com um *corpus* de fala espontânea implica ter que analisar a fala, isto é, não se limitar à leitura das transcrições, mas verificar o sinal sonoro para poder interpretar de maneira correta o que está sendo buscado. Uma das etapas iniciais do trabalho foi justamente aquela de conferir todos os enunciados analisados, ouvindo-os através de seu alinhamento ao áudio:

- I. para averiguar eventuais erros de transcrição;
- II. a fim de cotejar as segmentações e, portanto, chegar a uma interpretação correta dos enunciados;
- III. para desambiguar casos pouco claros.

Em relação a erros de transcrição, foram encontrados pouquíssimos casos:

- I. em *bpubdl07-639*¹⁰³, a forma *ocê* foi transcrita como *cê*;
- II. em *bfamcv18-66*, a transcrição de *ocês* na realidade é *cês*;
- III. *bfamcv11-5*: a forma *cê* foi transcrita como *ocê*;
- IV. *bfamdl22-175*: na transcrição da parte final do enunciado não havia sido compreendido o que fora dito: através da sucessiva análise acústica por nós feita foi possível transcrevê-la como “de vinte e cinco”;
- V. em *bfammn14-169*, a parte final do enunciado foi transcrita incorretamente e não inclui nenhuma das formas pronominais pesquisadas.

Em todos os casos, a qualidade acústica era muito baixa naquele ponto, havia muito barulho de fundo e sobreposições. Dado o volume do *corpus*, e mesmo sendo esta a revisão apenas de uma parte, parece-nos que esses poucos casos acabem por ser uma validação adicional, e ao mesmo tempo mostrem a confiabilidade do *corpus*, visto que esses casos representam 0,18% das ocorrências conferidas.

Por nossa parte, gostaríamos de lembrar que a revisão acústica é parte integrante e distintiva deste trabalho, especialmente em relação ao que foi feito até hoje, mas comporta grande dispêndio de tempo, viável, contudo, somente graças às características específicas do C-ORAL-BRASIL.

3.1.1.3. Critérios adotados para a seleção das formas objeto de pesquisa

A classificação das formas vai além de uma simples contagem de suas ocorrências, que muito pouco acrescenta à literatura e aos conhecimentos sobre o uso da língua. Deve-se,

¹⁰³ A sigla refere-se ao arquivo áudio-texto: contexto público/familiar, diálogo/monólogo/conversação, número do arquivo e número do enunciado.

portanto, definir uma taxonomia de classificação do que é significativo para a compreensão do fenômeno e estabelecer um método de aplicação de tal taxonomia. Diferentemente de quanto feito até este momento nas pesquisas sobre pronomes sujeito do PB, o ponto central desta proposta é que o critério de seleção e classificação reside na duração silábica das várias formas, conferida na segunda etapa do trabalho. No primeiro momento, foi efetuada uma separação com base em critérios distribucionais, incluindo a posição com relação às quebras prosódicas, e com base no critério semântico de contrastividade. Abaixo serão exemplificadas, com enunciados extraídos do C-ORAL-BRASIL, as classificações das várias formas que foram encontradas no *corpus* nessa primeira etapa ¹⁰⁴:

I. pré-verbal, sujeito do verbo:

- (1) *LUZ: [90] *nũ é importante nũ passar em primeiro lugar* // *LUZ: [91] *é importante passar / isso / pruma pessoa nova igual a ela / que tá começando a* + *LUZ: [92] *é legal ter isso no currículo* // *LUZ: [93] *igual ele falou com ela* // *LUZ: [94] *falou não uai* // *LUZ: [95] *se ocê passar em segundo lugar /cê tá muito bem / ué* // *LAU: [96] *é //*

(bfamdl03-95)



- (2) *LEO: [105] <beleza> // [106] <então a gente já sabe em quem que a gente vai colocar a> culpa // *EVN: [107] *é <hhh>* // *LEO: [108] <hhh> Gilberto // [109] *hhh* // [110] *você está condenado* // *EVN: [111] *hhh você está demitido //*

(bfamcv01-110-111)



- (3) *FLA: [1] *cê divide comigo empurrar isso* // *FLA: [2] <essa joça> // *REN: [3] <hhh não> / *empurra aí / que depois nós empurramo* // *REN: [4] *tem que olhar a lista aí / <das> + *FLA: [5] <eu lembro do> dia da festa / que a gente foi comprar bebida ali* // *REN: [6] *hhh que que tem //*

(bfamdl01-1)



¹⁰⁴ Os exemplos são citados como aparecem no *corpus* em sua versão anotada prosodicamente. Para efeito de melhor compreensão, esclarecemos aqui os símbolos utilizados. O asterisco seguido de três letras maiúsculas indica o início de turno e a identificação do informante. O número entre colchetes “[2]” é aquele do enunciado. A barra simples “/” significa fronteira prosódica que delimita uma unidade tonal, e a barra dupla “//” indica fronteira prosódica de final de enunciado. O símbolo “+” significa que o enunciado foi interrompido. Os colchetes “<>” indicam que naquele trecho há sobreposição de fala. A barra simples entre colchetes “</1>”, seguida de um número, refere-se à quantidade de palavras anteriores que são idealmente canceladas pelo falante em uma retomada. O símbolo “&” antecede um termo que está sendo reformulado, mesmo que de maneira incompleta. A sigla entre parênteses refere-se ao arquivo áudio-texto: contexto público/familiar, diálogo/monólogo/conversaão, número do arquivo e, eventualmente, número do enunciado.

II. pós-verbal:

a. quando objeto de verbo:

- (4) *CAR: [236] *tem alguém chamando lá / o' // *JOS: [237] xá eu ir lá ver / quem que é // *CAR: [238] quem que tá lá / será // *JOS: [239] vou lá ver // *CEL: [240] chamou cê não / sô // *CAR: [241] Tonim // *JOS: [242] chamou // *CAR: [243] ahn //*

(bfamcv05-240)



- (5) *mas / eu / uma coisa eu falo pra minha filha // [93] pra Mislaine // [94] eu falo / Mislaine // [95] a mamãe ama sua mãe // [96] por quê // [97] falei / porque / ela / gerou **você** pra mim // [98] e você é muito especial pra mim //*

(bfammn05-97)



b. quando objeto de preposição:

- (6) *FLA: [315] *<a gente tem que ensinar esse povo a jogar ainda> // *JUL: <não tem jeito / de saber quem que ganhou na verdade> // *EME: [317] <parceiro / confio **n' ocê**> // [318] <confio **n' ocê**> // *GUS: <subi no coqueiro / caí> de costas / <truco / seu bosta> //*

(bfamcv21-317)



- (7) *CAR: [187] *cê pode vir aqui // *HER: [188] só um minutinho // *HER: [189] o' / Márcia / <abrir eu já abri> // *MAJ: [190] <dona Hermínia tá> assassinando a garrafa aqui primeiro // *HER: [191] agora é **com você** // *MAC: [192] pode fazer o suco já //*

(bfamcv18-191)



- (8) *KAR: [4] *então tá // *KAR: [5] a mamãe conversa **com cê** // *KAR: [6] que que cê quer contar pa mamãe // *EDU: [7] porquim // *KAR: [8] ahn // *EDU: [9] dos porquinho //*

(bfamcv17-5)



- (9) *HER: [165] dá aqui / Maria José // *HER: [166] o' // *HER: [167] quer que eu quebro / p' cê // *MAC: [168] eu quero // *MAC: [169] esse aí coloca os pedaço / né / na jarra // *HER: [170] uhn // *HER: [171] filha // *HER: [172] quando tá / assim grosso / precisa fazer um caldinho / viu //

(bfamcv18-167)



c. quando sujeito posposto:

- (10) *TUT: [120] <eas deram um abraço> nela / ea foi / ei // *TUT: [121] quem é **você** hhh // *CLA: [122] Nossa / que <mico> // *RAQ: [123] <xxx / essa yyyy> // *CLA: [124] <e aí as mulheres> responderam // *CLA: [125] as gordinha <lá responderam> // *TUT: [126] <elas / ah> / adivinha // *TUT: [127] aí ela ficou olhando uma data / <tipo / cê é a xis> //

(bfamcv30-121)



- (11) *GUS: [339] <eu canguei> // *EME: [340] foi ela <que cangou> // *GUS: [341] <foi eu que canguei> // *EME: [342] <não> // *JUL: [343] <então porque que que só a Lud que xxx> // *EME: [344] foi **ocê** não // [345] <foi ela> // [346] ela mandou um dois // *GUS: [347] aí eu mandei um dois depois //

(bfamcv21-344)



III. isoladas em unidades tonais¹⁰⁵:

- (12) *KAR: [1] o quê // *KAR: [2] conta a história do <hhh> + *EDU: [3] <mãe> / **com você** // *KAR: [4] então tá // *KAR: [5] a mamãe conversa com cê // *KAR: [6] que que cê quer contar pa mamãe //


(bfamcv17-3)




- (13) *ELI: [3] jura // [4] ô / vai ser a hora da gente <passar / hein> // *MUR: [5] <desvira o volante> todo pra cá // [6] todo // [7] **ocê** / Renata / e a Ana Luíza // *ELI: [8] que

¹⁰⁵ As formas isoladas em unidade tonal ocorrem quando o pronome encontra-se sozinho, ou acompanhado por preposição, em uma unidade tonal. Em nenhum caso deve haver composicionalidade sintática. Já o isolamento absoluto, logo adiante, dá-se quando o pronome forma por si só um enunciado, e no máximo é acompanhado por preposição.

Ana Luiza // \$ [9] a mesma // *MUR: [10] é // [11] tudo junto // *ELI: [12] ai / Murilo // [13] fala sério // [14] cê achou que a mulher / pegou pesado comigo // *MUR: [15] <não> //


(bpubdl04-7) 

- (14) *CAB: [1] cê tá bonzinho / Dingó // *DIN: [2] uai / eu tô // [3] porque na verdade / foi o que eu falei / **com cê** // [4] que a + [5] ocê inclusive foi um argumento // *MAR: [6] eu <fui> // *DIN: [7] <quer> dizer / ocê apareceu / várias vezes lá no diálogo // *MAR: [8] ham ham //


(bfamcv23-3) 

IV. em isolamento absoluto:

- (15) *PAU: [85] é essa a medida aqui olha // *PAU: [86] do alicerce pra cima // *PAU: [87] ahn // *PAU: [88] experimenta pôr essa pedra com esse bico virado pra cá // *PAU: [89] isso / nessa posição // *PAU: [90] acho que ea fica melhor do que da outra // *PAU: [91] **p' cê** // *ROG: [92] Nossa /


(bpubdl01-91) 

- (16) *PTG: [274] ela é <mais leve> // *PTH: [275] passa o azeite por favor // [276] isso // *PTF: [277] e é com palmito também / nã é // *PTG: [278] é // *JAD: [279] portuguesa / né // *PTI: [280] é // [281] ai que bom que veio pequenininha // *PTJ: [282] opa // *JAD: [283] jovem // [284] **você** // *PTL: [285] eu // [286] não // [287] brigado // *JAD: [288] nada //


(bpubdl07-284) 

V. com valor contrastivo:

- (17) *GUI: [144] membro superior é mais difícil / mas / a gente tenta o máximo / né // *GUI: [145] pelo menos o esforço do antebraço aqui / é // *GUI: [146] força // *TOM: [147] <ah> // *GUI: [148] <isso> // *GUI: [149] beleza // *TOM: [150] No' // *GUI: [151] tranquilo // *TOM: [152] tranquilo **você** / né // *GUI: [153] eu tô tranquilo / uê // *GUI: [154] nu doeu nada / uê // *GUI: [155] tríceps sentado // *GUI: [156] não não não // *GUI: [157] aqui o' //


(bpubdl03-152) 

- (18) *CAR: [27] *chegou que hora ontem / nũ é // *ONO: [28] onze / uai // *ONO: [29] cê [1] a hora que cê &che [5] cê+ *ONO: [30] eu cheguei primeiro que ocê //*

(bfamcv11-30) 

VI. núcleo sintático do enunciado¹⁰⁶:

- (19) *LUC: [163] <valor simbólico hhh> // *LUC: [164] *contraventora / <cê fraga> // *MIC: [165] <Nossa / mas isso> <é desde nova // *FLL: [166] <eu só tenho amiga bandida> / sô / a Elisa é igualzinha / todo mundo é assim // *MIC: [167] cê é assim / Elisa // *MIC: [168] ocê com essa cara / de supernerd <sua> // *FLL: [169] <Elisa> / a &b + *EMF: [170] <não> // *FLL: [171] yyy / cê acha que a Elisa faz direito por quê // *FLL: [172] todo mundo que faz direito / pode desconfiar / filha //*

(bfamcv26-168) 


É importante ressaltar que, não raro, ocorrem casos que poderiam ser categorizados em mais de um grupo: no exemplo (12) há isolamento em unidade tonal, mas ao mesmo tempo a forma *você* é objeto de preposição. O mesmo ocorre no exemplo 15 em que *p'cê* é objeto de preposição, e ao mesmo tempo está em isolamento absoluto.

Os casos de duplicidade de função serão devidamente assinalados quando apresentarmos os dados detalhadamente.

Citamos também, como parte da classificação, os casos genericamente chamados de disfluência. Consideramos dois tipos de disfluência: a interrupção e o *retracting*.

Na interrupção o falante pode quebrar a continuidade da sua fala devido a um fator externo, ou até mesmo interno, e não há continuação do discurso, como em (20) abaixo.

- (20) *CEL: [204] *se eu fosse cê matava aqui o' // *CEL: [205] eu te avisei / Zé // *CEL: [206] p' <cê> +*

(bfamcv03-206) 

Nesta classificação, todos os tipos de interrupção serão considerados de forma única. Contudo, em um trabalho futuro talvez fosse o caso de refinar a análise das ocorrências para averiguar a oportunidade de tratá-los de maneira diferenciada. Contrariamente a (20), (21)

¹⁰⁶ Definimos núcleo sintático do enunciado o caso em que a forma pronominal ocupa uma posição de núcleo de um enunciado em que não há verbo e em que o pronome constitui a ramificação mais alta.

abaixo poderia merecer uma melhor análise e classificação, pois é possível uma interpretação do pronome em posição pré-verbal e com função de sujeito.

(21) **EME*: <*cê* & *m* [/2] *cê mandou*> +

(bfamcv21-355)



Já os casos de *retracting* ocorrem quando o falante corrige a própria fala cancelando algum elemento, como em (22) e (23):

(22) // *cê* [/1] *a gente saiu* /

(bfamcv23-43)



(23) **CAS*: *se eu andar aqui e for lá* / *ocê* & *va* [/2] *meu rei vai ficar desprotegido* /

(bfamd134-22)



No primeiro exemplo, o informante inicia o enunciado com *cê* para mudar o projeto locutivo para *a gente*. No segundo o planejamento, com *ocê* é modificado com *meu rei*. Esses casos também serão considerados de forma unitária, mas em um trabalho futuro mereceriam uma análise mais aprofundada para verificar se há ocorrências passíveis de melhor interpretação.

Esclarecidos os critérios de classificação, serão brevemente abordadas outras informações que foram colhidas para efeito de uma descrição mais detalhada e possíveis correlações na etapa *corpus-driven*¹⁰⁷ da pesquisa.

3.1.1.4. Taxonomia das anotações em tabelas do tipo *Excel*

Para uma melhor visualização dos resultados, e em vista das pesquisas cruzadas a serem feitas, foi preparada uma tabela do tipo *Excel*. As abas contêm os dados de cada forma, a saber: *você*, *vocês*, *ocê*, *ocês*, *cê*, *cês*.

As colunas estão assim divididas:

- I. forma (coluna B na fig.1);
- II. nome do arquivo;

¹⁰⁷ Definimos esta etapa como *corpus-driven*, diferenciando-a da *corpus-based* porque, apesar de termos estabelecido critérios de busca, será o *corpus* que indicará os caminhos subsequentes, visto que não é possível prever os casos que serão encontrados. Com os dados à mão se poderá estabelecer o que é mais ou menos relevante para a continuação da pesquisa.

- III. número do enunciado;
- IV. *token* (enunciado ou trecho de texto para contextualização);
- V. especificação da posição, segundo os critérios apresentados anteriormente;
- VI. dados sociolinguísticos dos informantes: sigla;
- VII. dados sociolinguísticos dos informantes: número de código;
- VIII. dados sociolinguísticos dos informantes: gênero;
- IX. dados sociolinguísticos dos informantes: idade;
- X. dados sociolinguísticos dos informantes: escolaridade;
- XI. dados sociolinguísticos dos informantes: origem;
- XII. dados sociolinguísticos dos informantes: profissão;
- XIII. anotações de vários tipos para uso do pesquisador: possíveis topicalizações, subordinadas, conferência da transcrição etc.

Após a varredura do *corpus*, que será descrita abaixo, cada item foi preenchido e o resultado final possibilitou visualizar um quadro geral, como o da figura (1), da situação atual das formas pronominais de segunda pessoa no PB, como encontrado no C-ORAL-BRASIL.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	CLUT	ARQUIVO	n. en.	TOKEN		FUNÇA	INFOR	COD	GE	DA	ESC	ORIGEM	PROFISS	OBS.
2	1	cê	bfamcv01 bt 38	*LEO: [38] como o Chub fala / es vão pegar os caras		sub PV	*LEO	146	M	A	2	BH	Student	
3	1	cê	bfamcv01 bt 90	// [90] <cê se> sente numa <prisão / e tal hhh> //		sub PV	*EVN	143	M	A	2	BH	Student	
4	1	cê	bfamcv01 bt 128	*EVN: [128] <ê> / concordo / isso aí que cê falou //		sub PV	*EVN	143	M	A	2	BH	Student	
5	1	cê	bfamcv01 bt 139	*LUI: [139] <eu acho que / o nome / do> [1] do [1] de		sub PV	*LUI	145	M	A	2	BH	Student	
6	1	cê	bfamcv02 bt 48	*RUT: [48] uai / e cê acha isso velho / <Terezinha> //		sub PV	*RUT	10	F	C	2	BH	Retired Teacher	
7	1	cê	bfamcv02 bt 95	// [95] cê nú viu que o> da Tiane eu nem <entrei> na		sub PV	*RUT	10	F	C	2	BH	Retired Teacher	
8	1	cê	bfamcv02 bt 193	*RUT: [193] <lá em casa> / <aviso se cê for lá> //		sub PV	*RUT	10	F	C	2	BH	Retired Teacher	
9	1	cê	bfamcv02 bt 202	// [202] porque lá em casa / cê sabe / presente da m		sub PV	*RUT	10	F	C	2	BH	Retired Teacher	
10	1	cê	bfamcv02 bt 218	// [218] cê pensa que ele / participa da [1] & [1] de		sub PV	*RUT	10	F	C	2	BH	Retired Teacher	
11	1	cê	bfamcv02 bt 260	*RUT: [259] é // [260] cê tem que ir chique //		sub PV	*RUT	10	F	C	2	BH	Retired Teacher	
12	1	cê	bfamcv02 bt 267	*JAE: [267] inda bem que cê vai tar co seu cabelo cu		sub PV	*JAE	9	F	D	2	BH	Retired Teacher	
13	1	cê	bfamcv02 bt 273	// [273] o' p' cê <ver> //		sub PV	*RUT	10	F	C	2	BH	Retired Teacher	
14	1	cê	bfamcv02 bt 300	*TER: [300] <mas quando a Paulinha> casar / cê <v		sub PV	*TER	11	F	C	1	BH	Homemaker	
15	1	cê	bfamcv02 bt 329	// [329] aí cê acha que eu devo convidar o Guilherme		sub PV	*TER	11	F	C	1	BH	Homemaker	
16	1	cê	bfamcv02 bt 335	*RUT: [335] convite de casamento cê pode mandar		sub PV	*RUT	10	F	C	2	BH	Retired Teacher	
17	1	cê	bfamcv02 bt 361	*RUT: [361] <cê convidou> +		sub PV	*RUT	10	F	C	2	BH	Retired Teacher	
18	1	cê	bfamcv03 bt 32	// [32] o' p' cê ver //		sub PV	*CEL	215	M	B	1	Passatempo_Merchant		
19	1	cê	bfamcv03 bt 35	*CEL: [34] uai // [35] olha p' cê ver //		sub PV	*CEL	215	M	B	1	Passatempo_Merchant		
20	1	cê	bfamcv03 bt 50	// [50] aí cê coloca o <resto> //		sub PV	*REN	241	M	B	2	Contagem-MC Police officer		
21	1	cê	bfamcv03 bt 57	*REN: [57] cê <erra / nós tava fodido> //		sub PV	*REN	241	M	B	2	Contagem-MC Police officer		
22	1	cê	bfamcv03 bt 59	*CAR: [59] cê erra <também / Toninho / aí ficava o s		sub PV	*CAR	213	M	B	3	Inhapim-MG Teacher		
23	1	cê	bfamcv03 bt 60	*CEL: [60] <ô Toninho / cê tem que matar o cinco / s		sub PV	*CEL	215	M	B	1	Passatempo_Merchant		
24	1	cê	bfamcv03 bt 61	*REN: [61] agora cê mata o onze aí //		sub PV	*REN	241	M	B	2	Contagem-MC Police officer		
25	1	cê	bfamcv03 bt 81	*CEL: [81] <cê devia pegar> outro //		sub PV	*CEL	215	M	B	1	Passatempo_Merchant		
26	1	cê	bfamcv03 bt 121	// [121] melhor coisa que cê faz //		sub PV	*CEL	215	M	B	1	Passatempo_Merchant		
27	1	cê	bfamcv03 bt 129	// [129] olha p' cê ver / vai matar ainda //		sub PV	*CEL	215	M	B	1	Passatempo_Merchant		
28	1	cê	bfamcv03 bt 130	// [130] cê vê / voltando aqui o' //		sub PV	*CEL	215	M	B	1	Passatempo_Merchant		

Figura 1: imagem da tabela *Excel* preenchida com os dados sobre a forma *cê* e sua análise.

3.1.2. Análise acústica: descrição do método utilizado

A maior inovação metodológica que esta pesquisa propõe reside no fato de que parte dos dados acima apresentados foram submetidos a uma análise prosódica. A hipótese subjacente é que questões como tonicidade ou atonicidade não podem ser definidas com base em fatores de tipo sintático, mas dependem de fatores acústicos. Isso significa que a taxonomia que foi aplicada e descrita na primeira parte da verificação dos dados serviu como base para uma análise mais profunda.

Os estudos prosódicos assinalam três parâmetros que devem ser observados quando se analisa a produção da fala e, pelo que nos interessa, a noção de tonicidade:

- I. a duração, ou seja, a diferença de tempo entre dois eventos;
- II. a frequência fundamental (f_0), é a primeira frequência, ou ciclo de vibrações, produzida na glote (que é o espaço onde se situam as pregas vocais). A f_0 está ligada às variações de altura e intensidade;
- III. e a intensidade, entendida como a energia de vibração da fonte que emite o som.

Esses três parâmetros são relacionados, pela percepção da fala, às noções de:

- I. duração percebida;
- II. altura;
- III. e volume.

Na realização do acento há de se considerar também a frequência dos formantes (F_1 , F_2 e F_3), medida em Hz. Os formantes são as ressonâncias da modulação dos sons no trato vocal. As vogais podem ser caracterizadas pelos primeiros três formantes. O termo frequência indica o número de ciclos, as vibrações, que as pregas vocais perfazem em um segundo. Sua frequência é que determina a qualidade das vogais. A duração das vogais reduzidas é bem menor do que aquela de suas contrapartes plenas, *ceteris paribus*. Como são as vogais que carregam primariamente o acento, as dimensões acústicas a serem observadas são a duração, a f_0 e a intensidade.

Quando uma sílaba é acentuada, há um aumento na duração e na intensidade relativa da sílaba saliente (cf. ABERCROMBIE, 1968; MAJOR, 1985; MORAES, 1998; CAGLIARI, 1999; HAYEARD, 2000; BARBOSA, 2006; FERREIRA NETO, 2007 entre outros). Em cada língua existe a predominância de algum desses parâmetros na sinalização do acento (cf. LAVER, 1994). No PB, os estudos de natureza fonético-acústica apontam para a duração como principal correlato do acento, seja daquele lexical (MASSINI, 1991; BARBOSA, 2000 e 2002; FONTES, 2013), seja daquele frasal (BARBOSA, 2000 e 2002).

Neste trabalho optou-se, portanto, por considerar a duração como principal medida a ser observada na análise acústica.

Esta análise mensura a duração silábica das formas pronominais *cê(s)*, *ocê(s)* e *você(s)*, comparando-as com as sílabas dos enunciados e das unidades tonais a que pertencem. A partir dos resultados obtidos é possível estabelecer se as formas são longas ou curtas com relação à média do termo de comparação e, portanto, fornecer um parâmetro para avaliá-la como tônicas ou átonas sem recorrer a pressupostos de que a forma segmental ou a função sintática possam decidir sobre um parâmetro que é de natureza prosódica. Neste estudo, pretende-se testar a hipótese de que o que marca a função das formas pronominais em PB seja a prosódia e não a sua forma segmental, o que é possível graças aos recursos de pesquisa oferecidos pelo *corpus*, e pelo aparato tecnológico hoje acessível, que permite avaliar o que até pouco tempo atrás carecia dos recursos necessários. Lembramos especificamente que este estudo foi possível somente graças aos recursos oferecidos por *corpora* como o C-ORAL-BRASIL e aqueles que seguem a arquitetura do C-ORAL-ROM.

A sensação de duração percebida está diretamente associada ao cruzamento dos três parâmetros prosódicos acima como um todo (BARBOSA P. A., 1999). Ao se estudar a fala, a análise prosódica é geralmente entendida como o:

estudo da duração observada, deixando-se de lado os papéis desempenhados pela frequência fundamental e pela intensidade em sua percepção (BARBOSA P. A., 1999, p. 23)

A análise acústica de enunciados de fala espontânea é uma tarefa complexa por vários motivos, entre os quais:

- I. grande variabilidade do conteúdo locutivo de cada enunciado;
- II. possibilidade de mudanças na taxa de elocução, até daquela de um mesmo falante;
- III. interferência de vários tipos de ruído de fundo que podem introduzir fatores de maior complicação;
- IV. sobreposições de funções diferentes (por ex. funções informacionais diferentes).

Com a finalidade de minimizar esses fatores foi realizada a comparação das formas pronominais com as sílabas de um mesmo enunciado e de uma mesma unidade tonal.

Através de tal procedimento é possível verificar se a(s) sílaba(s) da forma analisada encontra(m)-se entre aquelas mais longas ou entre aquelas mais curtas do mesmo enunciado e unidade tonal.

Os critérios de segmentação fonética envolvem decisões que devem ser tomadas na escolha das fronteiras entre um evento e outro. Há também de se definir por qual, dentre a variedade de fenômenos observáveis no sinal de fala, deve-se optar.

Os resultados da segmentação que apresentaremos foram obtidos seguindo-se as seguintes etapas:

- I. foi efetuada uma busca manual nas planilhas do tipo Excel *supra* especificadas dos enunciados de interesse;
- II. através do programa *WinPitch*, foram aferidos os enunciados selecionados. O *software* permite visualizar o alinhamento do texto ao som e conferir a qualidade do espectro acústico e a boa audibilidade do som. Em caso positivo, o trecho em questão foi selecionado, copiado em formato *.wav* e transferido para uma pasta separada;
- III. leitura do arquivo de áudio no programa *Praat* (BOERSMA; WEENINK, 2013) e segmentação do enunciado em palavras, sílabas e fones;
- IV. extração dos dados relativos à duração das sílabas de cada enunciado em formato *.txt*, e conversão em formato compatível com uma planilha do tipo *Excel* para análise estatística.

Em relação aos critérios de segmentação, deve-se atentar ao fato de que estes requerem do pesquisador uma tomada de decisão sobre qual evento do sinal acústico há de ser utilizado como fronteira para delimitação entre o início e o fim do segmento mensurado. Sempre haverá certa dose de incerteza, devido principalmente aos fenômenos de coarticulação (BARBOSA P. A., 1999). Barbosa P. A. (1999) define o termo segmentação como:

a marcação dos eventos que foram escolhidos para a caracterização da duração.

Nesta pesquisa, os principais indicadores de segmentação foram: a análise do espectrograma e a percepção do sinal acústico.

O trabalho de segmentação é bastante complexo e demorado: o enunciado inteiro é ouvido, através do *software Praat*, e, com base em uma primeira impressão, são marcadas em

camadas de anotação (*TextGrid*) as fronteiras entre palavras. Vários indícios permitem assinalar as fronteiras entre fones, citamos alguns:

- I. a observação dos formantes, ou seja, das frequências de ressonância em volta das quais se formam picos de amplitude no espectro, permite localizar as vogais, com referência na segmentação a F2;
- II. as oclusivas são precedidas por um período de oclusão após o qual há uma “plosão”, assinalada no espectrograma com uma barra vertical bastante reconhecível;
- III. as sibilantes não vozeadas, fundamentais para o reconhecimento do *cê(s)*, possuem um espectro não harmônico com F1 alto.

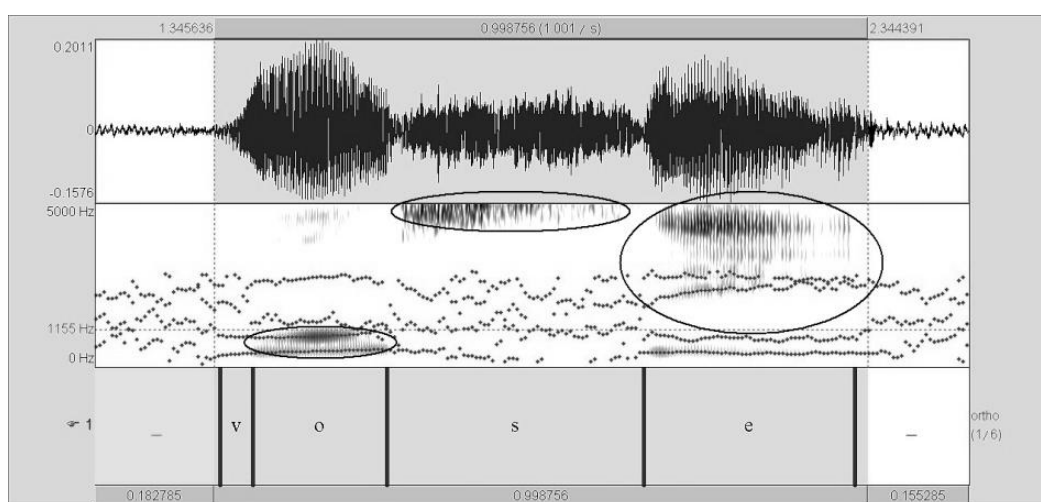


Figura 2: Espectro do pronome “você” segmentado no programa Praat: é possível visualizar os formantes, em específico as vogais [o] e [e] e a sibilante não vozeada [s] descritas acima.

Diversos fenômenos (ruído de fundo e coarticulação são os mais relevantes), às vezes, impossibilitam detectar todas as fronteiras. É por isso que são importantes a observação e a interpretação do espectrograma, que fornece um “retrato” das ondas acústicas produzidas na fala.

A detecção dos fones é importante para efetuar sua junção em sílabas, que é a parte que foi analisada. Em um primeiro momento, elas foram transcritas ortograficamente, com poucas modificações. Justamente para preservar a legibilidade imediata dos enunciados. Em seguida as sílabas foram todas transcritas em transcrição fonética larga¹⁰⁸, em formato *ASCII*, para serem analisadas pelo programa *SGdetector* (BARBOSA P.A., 2006), como será

¹⁰⁸ “broad phonetic transcription”, Barbosa P. A. (2013, p. 86).

explicado adiante. Deixando de lado a questão do formato da transcrição, as fronteiras permanecem as mesmas, e as análises foram feitas em cima dessas segmentações¹⁰⁹.

O *script* referido extrai automaticamente a duração de cada sílaba e, como foi dito, os dados são inseridos em uma planilha do tipo *Excel* para efeito de comparação.

3.1.2.1. Definição dos critérios de escolha dos enunciados para análise

As características definidas inicialmente como prioritárias dos enunciados objeto de análise acústica eram:

- I. conter pelo menos uma ocorrência das formas *cê(s)/ocê(s)/você(s)*;
- II. possuir uma alta qualidade acústica (segundo os critérios do C-ORAL-BRASIL qualidade A ou AB¹¹⁰);
- III. não fazer parte de enunciados com sobreposição de fala.

Contudo, após várias análises, notou-se que os enunciados que contêm as formas mais interessantes para a pesquisa, especialmente as pós-verbais, e que ocorrem em menor número, nem sempre se encontram em enunciados que possuem as características acima.

Muitas vezes as ocorrências situam-se justamente em textos com muitas sobreposições de fala ou de qualidade acústica inferior. Como a prioridade é justamente medir as várias formas nas diferentes funções, foi necessário modificar um pouco os critérios de busca e, evitando justamente as sobreposições de fala que impossibilitariam a análise acústica, ver, entre os textos de qualidade acústica mais baixa, quais enunciados ainda possibilitavam uma análise.

Foram submetidas à análise acústica 133 formas diferentes e seus relativos contextos. A tabela abaixo apresenta as formas analisadas:

¹⁰⁹ Agradecemos a Horácio Queiroz por nos ter introduzido à análise acústica e, pacientemente, ter ilustrado o funcionamento do programa *Praat* e as modalidades de segmentação acústica. Agradecemos também a disponibilidade do professor Plínio Barbosa em conferir muitas das segmentações e instruir-nos sobre a utilização do *script Sgdetector*.

¹¹⁰ Para os critérios de avaliação da qualidade acústica das gravações remetemos ao parágrafo 3.1 “A qualidade acústica” do cap. 2 “O corpus C-ORAL-BRASIL” (RASO, 2012).

Forma	Ocorrências
CÊ PRÉ-VERBAL	37
CÊ PÓS-VERBAL	35
OCÊ PRÉ-VERBAL	8
OCÊ PÓS-VERBAL	16
VOCÊ PRÉ-VERBAL	24
VOCÊ PÓS-VERBAL	13
TOTAL	133

Tabela 1: ocorrências das formas pronominais de segunda pessoa singular submetidas a análise acústica.

Foram estudados 62 textos diferentes: 52 pertencentes ao contexto familiar, dos quais 22 conversações, 17 diálogos e 13 monólogos. Outros 9 textos fazem parte do contexto público: 3 conversações, 5 diálogos e um monólogo.

Em relação à qualidade acústica dos textos veja-se a tabela abaixo:

QUALIDADE ACÚSTICA	NÚMERO DE TEXTOS
A	16
AB	21
B	10
BC	7
C	8
TOTAL	62

Tabela 2: ocorrências dos textos analisados em base à sua qualidade acústica.

Foram analisados enunciados de 16 textos diferentes, com qualidade acústica A, muito alta. Outros 21 textos são de qualidade AB, também alta. Dez textos são de qualidade B, médio alta. Sete são de qualidade BC, médio baixa. Por fim, 8 textos são de qualidade C, considerada baixa.

É importante esclarecer que, quando citamos a qualidade acústica, esta se refere à que foi estabelecida para o texto inteiro. Contudo, isso não corresponde necessariamente à qualidade acústica de um enunciado específico. Pode acontecer que a qualidade do texto seja alta, mas no trecho escolhido haja ruídos de fundo ou uma sobreposição de fala, o que compromete a compreensão, ou o caso oposto, em que em um texto com baixa qualidade seja possível encontrar um enunciado adequado que seja perfeitamente compreensível e analisável.

Os resultados das análises serão reportados no capítulo 4.

3.1.2.2. Normalização dos dados.

A análise acústica permite medir a duração das sílabas e as comparar com as outras do mesmo enunciado ou unidade tonal. Os dados obtidos mostram o aproximar-se das sílabas das formas pronominais objeto de estudo, das sílabas mais longas e daquelas mais curtas do enunciado. Desta forma, pode-se afirmar, com certo grau de confiabilidade, que determinadas sílabas são longas ou são curtas com relação ao enunciado ou unidade tonal e fazer uma hipótese sobre sua tonicidade.

Contudo, os dados assim obtidos ainda são considerados “brutos”. Cada vogal e consoante possui um seu valor intrínseco de duração (por ex., a vogal “a” é mais longa do que a “e”, por exemplo). A complexidade silábica pode dar a impressão de que uma sílaba curta, mas mais complexa sob o ponto de vista segmental, pareça mais longa. Deve-se lembrar também que sílabas em final de enunciado geralmente tendem ao alongamento. Sílabas que apresentam a mesma duração, mensurada em milissegundos, podem na realidade possuir valores diferentes, dependendo das vogais e das consoantes nelas contidas. Faz-se, portanto, necessário implementar um processo de normalização das sílabas que leve em conta tais fatores e permita comparar as sílabas diferentes de forma mais adequada.

Uma possível crítica ao nosso trabalho é que na parte do PB não levamos em conta as características sociolinguísticas dos informantes. Na realidade isto é verdade somente em parte: na etapa inicial do trabalho, quando foram separadas as formas com os respectivos enunciados, fez-se um apanhado geral de todos os dados relativos aos informantes. Contudo, visto o grande dispêndio de tempo que nos requereu a etapa de análise acústica, não nos foi possível computar e separar todas estas informações de maneira adequada. Todavia, pudemos observar que as várias formas e suas funções não parecem ter um peso determinante para as finalidades desta pesquisa.

3.1.2.3. O software *SGdetector*: funcionamento e princípios básicos

O *SGdetector* (BARBOSA P. A., 2006) é um *script* realizado para ser rodado no programa *Praat*. Ele se baseia na concepção de que a duração é um parâmetro fundamental para assinalar acento, proeminência e fronteiras prosódicas em várias línguas entre as quais o PB. O programa foi implementado em 2004 e aperfeiçoado em 2009 e 2010. Ele permite detectar de maneira semiautomática picos de duração de sílabas e marca as palavras proeminentes via pico de duração local em seu interior.

Seu funcionamento requer que a segmentação seja feita seguindo uma transcrição fonética larga utilizando o código ASCII, e não o padrão IPA.

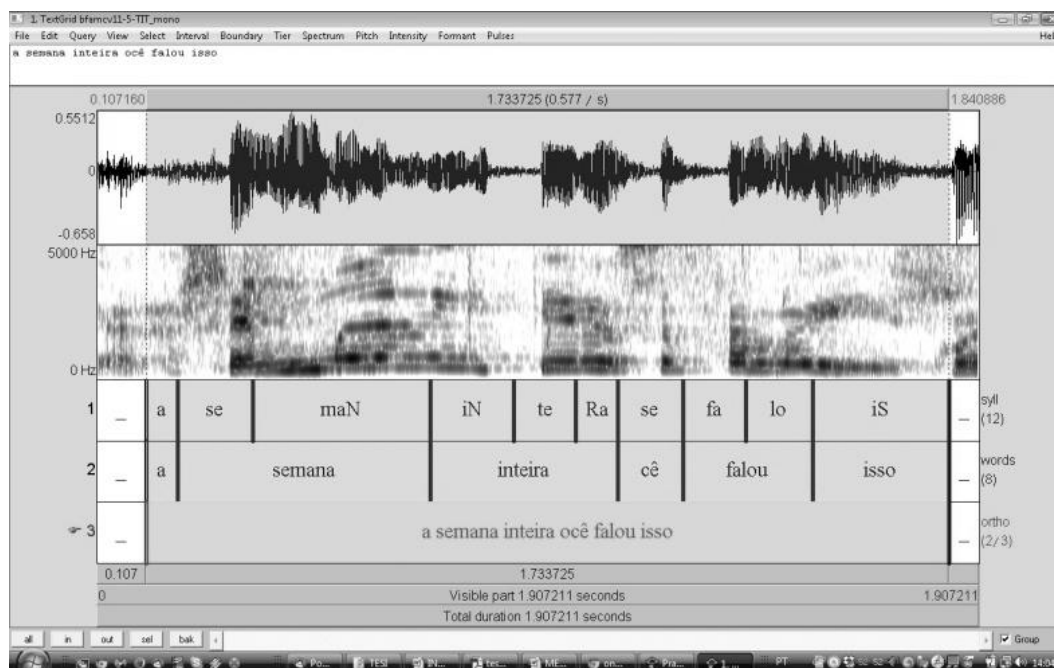


Figura 3: Tela do Praat com segmentação em código ASCII (primeira linha).

O *TextGrid* da segmentação é inserido no programa juntamente com o arquivo *TableOfReal* que contém as médias e os desvios padrão em milissegundos da duração dos fones da língua em questão, neste caso o PB¹¹¹.

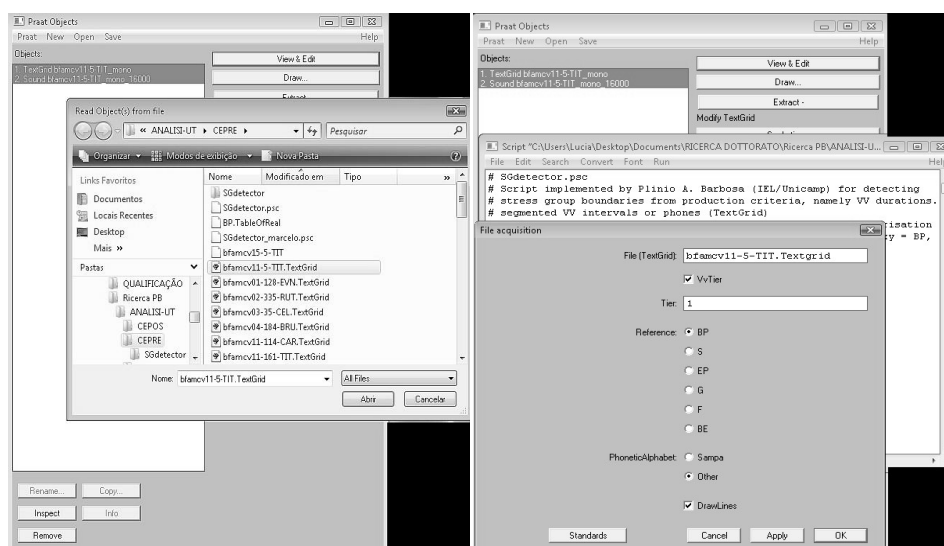


Figura 4: Abertura do script *SGDetector* e inserção dos dados para normalização.

O script gera dois arquivos de texto:

¹¹¹ Para os detalhes técnicos sobre as equações utilizadas pelo programa remetemos a Barbosa (2013).

- I. O primeiro contém cinco colunas que se referem a: a) a transcrição silábica; b) a duração bruta em milissegundos; c) o índice estatístico de desvios padrão (em *z-scores*) da duração bruta; d) uma suavização de cinco pontos do *z-score*; e) um valor binário (0 e 1) que indica se a posição é um pico local de *z-score* suavizado ou não.
- II. O segundo contém duas colunas. A primeira mostra a duração bruta em milissegundos do grupo acentual relacionado à duração. A segunda o número de unidades silábicas no grupo acentual correspondente.

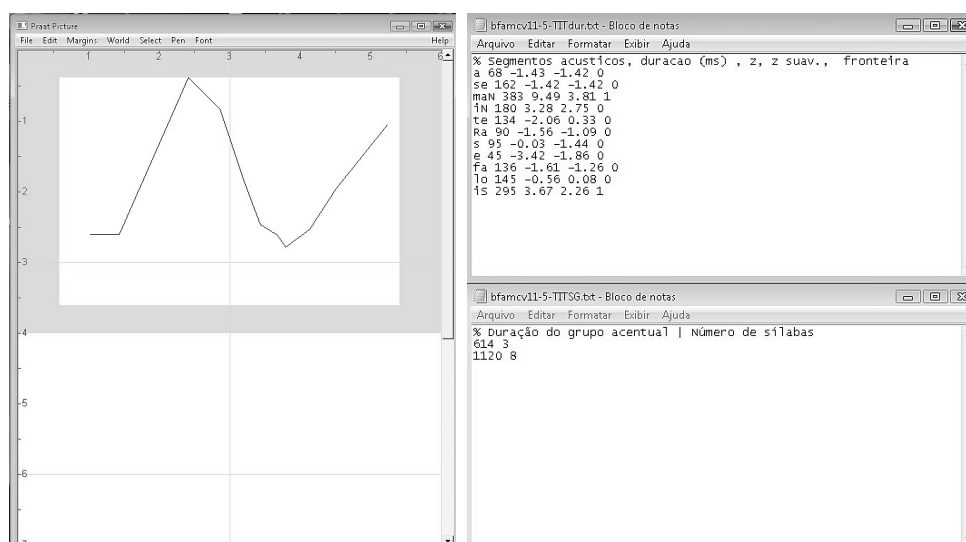


Figura 5: Saída de dados do *SGDetector*: gráfico e arquivos de texto.

O primeiro arquivo de texto é o que mais nos interessa, pois a coluna “c)” indica a duração normalizada das sílabas e, dessa forma, é possível realizar as análises.

O *script* é capaz de gerar outros dados que podem ser interessantes para análises futuras, como poder visualizar os picos de duração do enunciado e ver em que palavras se apoiam, entre outras funções.

3.1.3. Validação estatística

Ao final da etapa de análise, foi implementada uma validação estatística dos resultados obtidos¹¹². Devido ao valor reduzido da amostra e como os dados não apresentam distribuição normal¹¹³ optou-se por utilizar o teste de *Wilcoxon-Mann-Whitney* (cf. SPRENT &

¹¹² Agradecemos a Marcelo Vieira pela orientação sobre os testes mais adequados a serem utilizados.

¹¹³ Considera-se distribuição normal, em estatística, uma distribuição de probabilidade contínua geralmente utilizada como primeira aproximação para a descrição das variáveis casuais e os valores reais que se concentram em volta de um único valor médio.

SMEETON, 2001; CONOVER, 1999 entre outros), ideal para dados que não apresentam distribuição normal¹¹⁴.

3.2. Análise do vernáculo florentino

3.2.1. O *Corpus Stammerjohann*: a coleta original

O *Corpus Stammerjohann* é considerado o primeiro exemplo de *corpus* de fala espontânea em italiano. Antes mesmo da publicação das pesquisas sobre estudos sociolinguísticos de Labov, o pesquisador alemão Harro Stammerjohann esteve em Florença, de 29 de janeiro a 26 de fevereiro de 1965. Ali efetuou 30 gravações, num total de mais de quarenta horas¹¹⁵. O interesse do então doutorando em romanística era coletar dados para um estudo sobre o italiano falado em Florença. O então Presidente da *Accademia della Crusca*¹¹⁶, Giovanni Nencioni, introduziu o jovem pesquisador à comunidade de artesãos do bairro popular de *Santo Spirito* e a famílias e profissionais das redondezas. Os critérios de coleta, através de gravações de interações naturais da língua viva, produzida espontaneamente, buscando suas variações de uso mais significativas, eram algo totalmente inédito para a época. Esta é uma das grandes vantagens dos dados do *corpus*, o que permitiu também sua remodelagem sob uma ótica atual, comparável com dados recentes, como será visto adiante.

As gravações foram feitas com um equipamento muito sofisticado para a época¹¹⁷, o que garantiu sua preservação até o ano de 2001, quando foram remasterizadas. Metade do tempo de gravação é composto por gravações ambiente. O gravador foi posto em uma posição oculta dentro de uma barbearia em *via Faenza*¹¹⁸ enquanto os fregueses entravam e saíam: há diversas situações, diferentes temas de discussão e participantes variados, além de longos momentos de silêncio. O restante das gravações compõe-se cada uma por um texto, sem mudanças de situação ou falantes e temas. Foram efetuadas em ambiente familiar/privado ou

¹¹⁴ Este teste verifica, na presença de dados com distribuição contínua, se duas amostras estatísticas provêm da mesma população.

¹¹⁵ Pela precisão: 41 horas 15 minutos e 34 segundos

¹¹⁶ Considerada a maior instituição italiana para o amparo e estudo da língua italiana.

¹¹⁷ Um gravador Rk 34 da Philips com bobinas de quatro pistas e velocidade de 4,75 cm por segundo.

¹¹⁸ Uma rua do centro de Florença.

público e tratam de vários assuntos. Os gêneros predominantes são a entrevista e a conversação em contexto familiar¹¹⁹.

Como é possível notar, apesar de não poder ser considerado um verdadeiro *corpus*, pois não possuía em seu planejamento originário tal objetivo ou arquitetura, os dados coletados mostram-se extremamente úteis e significativos para uma pesquisa como aquela aqui proposta, de estudo da língua espontânea, seja como quadro de uma situação, em sincronia, seja em uma análise diacrônica. A remasterização e a nova compilação dos dados efetuada pela equipe do Laboratório *Lablita* possibilitou esse tipo de trabalho.

3.2.2. O trabalho de remasterização, transcrição e etiquetagem

O *corpus* nunca chegou a ser completamente transcrito ou publicado¹²⁰ pelo autor. Foi doado a Emanuela Cresti para que fosse arquivado junto aos *corpora Lablita* e utilizado pela comunidade científica.

A digitalização das bobinas ocorreu em 2001¹²¹: os arquivos foram tratados digitalmente e salvos em formato *.wav PCM*¹²². Este material de áudio foi ouvido integralmente e avaliado em seu conteúdo: desta forma, os pesquisadores envolvidos puderam subdividir os textos das gravações ambiente, e em geral fazer uma divisão com base na duração e no número de palavras de todos os textos, além de classificar os metadados.

O que é interessante para a finalidade de nossa pesquisa é que houve uma forte preocupação em redistribuir os textos com base nas variações sociolinguísticas e assegurar que houvesse um balanceamento em termos quantitativos. Isso significa que as características sociolinguísticas e as dimensões dos textos foram adequadas aos padrões do C-ORAL-ROM (informal: 1500 e 4500 palavras; formal: 3000 palavras), possibilitando comparações.

Foram transcritas cerca de 100.000 palavras do total de mais de 40 horas de gravação. Uma parte das transcrições feitas por Stammerjohann foi reaproveitada, após sua revisão e conversão no formato utilizado para todo o *corpus*, o CHAT, com anotação prosódica. Para a escolha dos textos houve a preocupação de encontrar trechos com boa

¹¹⁹ Uma descrição detalhada de todos os textos, participantes e duração das gravações encontra-se em: Scarano e Signorini (2003). As estudasas foram também entre as responsáveis pela atualização do *corpus*, como será visto adiante.

¹²⁰ Há somente uma publicação do autor: Stammerjohann (1970), na revista *Studi di Filologia Italiana*, em que aparecem algumas transcrições do *corpus*.

¹²¹ O trabalho de remasterização e os estudos subsequentes fazem parte do projeto “L’italiano orale in diacronia”, financiado pelo *FIRB (Fondo per la Ricerca di Base)*, do governo italiano.

¹²² Tal operação foi bastante complexa, pois foi necessário encontrar um gravador da época ainda em função e submeter o material a dois tipos de tratamento, um para o estudo imediato e a formatação em *corpus*, e o outro para o arquivo do laboratório. Para os detalhes, cf. Signorini e Tucci (2004).

qualidade acústica, mas que, ao mesmo tempo, obedecessem aos critérios de correto balanceamento e representatividade diafásica. A divisão dos textos segue aquela da família C-ORAL-ROM:

- I. textos informais subdivididos em familiar, privado e público;
- II. textos formais subdivididos em contexto natural e transmitido;
- III. todos os textos obedecem à classificação do evento comunicativo (monólogo, diálogo e conversação)¹²³.

Todos são alinhados ao som através do software *WinPitch*.

Pode-se observar que graças a estas características é possível pesquisar os dados gravados em 1965 utilizando os recursos atuais de busca. E ainda mais importante, é possível verificar facilmente o áudio através de uma pesquisa textual.

A escolha deste *corpus*, portanto, parece-nos óbvia, sobretudo pela ausência de pesquisas sobre a morfologia e aspectos prosódicos nele baseadas¹²⁴.

Finalizando esta parte, apresentamos brevemente o *corpus* de forma mais detalhada¹²⁵:

¹²³ Maiores detalhes em: Scarano e Signorini (2003, p. 10).

¹²⁴ Até hoje há somente um trabalho publicado que utiliza este *corpus*, uma pesquisa de análise diacrônica sobre o léxico florentino (Moneglia *et al.*, 2006).

¹²⁵ A divisão entre regulado e não regulado responde à qualidade da interação: regulada, quando há normas sociais que determinam a troca de turno entre os falantes, como em situações familiares ou de trabalho; não regulada quando não há como se prever a resposta do outro falante e a troca de turno acontece de forma casual.

CONTEXTO	TIPO	INTERAÇÃO	NÚMERO DE TEXTOS
FAMILIAR 22.517 palavras	NÃO-REGULADO 22.517 palavras	MONÓLOGO: 0 palavras	0
		DIÁLOGO: 9.310 palavras	2
		CONVERSAÇÃO: 13.207 palavras	4
PRIVADO 36.817 palavras	NÃO-REGULADO 25.606 palavras	MONÓLOGO: 0 palavras	0
		DIÁLOGO: 0 palavras	0
		CONVERSAÇÃO: 25.606 palavras	8
	REGULADO 11.211 palavras	MONÓLOGO: 3.023 palavras	1
		DIÁLOGO: 8.188 palavras	4
		CONVERSAÇÃO: 0 palavras	0
PÚBLICO 23.196 palavras	NÃO-REGULADO 16.278 palavras	MONÓLOGO: 0 palavras	0
		DIÁLOGO: 0 palavras	0
		CONVERSAÇÃO: 16.278 palavras	4
	REGULADO 6.918 palavras	MONÓLOGO: 1.561 palavras	3
		DIÁLOGO: 5.357 palavras	4
		CONVERSAÇÃO: 0 palavras	0
TELEFÔNICO 3.722 palavras		3.722 palavras	2
TOTAL		86.252 PALAVRAS	32

Quadro 6: Estrutura do *corpus Stammerjohann*, com número de palavras e textos para cada campo da estrutura.

Da gravação original foi retirada toda a parte de registro de emissões radiofônicas, por não serem representativas da fala espontânea, e menos ainda daquela do vernáculo florentino. Do restante das gravações, algumas estavam em condições não recuperáveis. daquelas que responderam aos critérios apresentados acima, foram extraídos os textos mostrados na tabela, que se referem ao *corpus Stammerjohann* recuperado pelo laboratório *Lablita*.

O *corpus* compõe-se, desta forma, por 86.252 palavras: está subdividido hierarquicamente em três contextos: familiar, privado e público. Estes são por sua vez divididos com base no tipo de interação: regulada e não regulada, cada qual subdividida com base na estrutura do evento comunicativo: monólogo, diálogo e conversação.

O contexto familiar apresenta cerca de um quarto do *corpus*, com 22.517 palavras, sendo todas as interações não reguladas e divididas entre dois diálogos (para um total de 9.301 palavras) e quatro conversações (para um total de 13.207 palavras).

O contexto privado é o maior de todos, com 36.817 palavras: a parte não regulada contém somente conversações, oito no total, para um montante de 25.606 palavras. As interações reguladas dividem-se entre um monólogo, de 3.023 palavras, e quatro diálogos, somando 8.188 palavras.

O contexto público perfaz 23.196 palavras: a parte não regulada é composta por quatro conversações, totalizando 16.278 palavras. A parte regulada divide-se entre três monólogos, que juntos somam 1.561 palavras, e quatro diálogos, para um total de 5.357 palavras.

No final foram transcritos, e fazem parte do *corpus*, dois diálogos telefônicos que perfazem 3.722 palavras.

Quanto aos participantes, há profissionais liberais, artesãos, professoras, crianças, donas de casa entre outros: todos pertencentes à comunidade local de um dos bairros mais tradicionais de Florença.

Estas características são primordiais para este tipo de pesquisa: como se vê, a estrutura do *corpus* privilegia diálogos e conversações, sendo os monólogos uma parte menor do total. Isto garante maior interação e espontaneidade da fala gravada.

3.2.3. O Corpus per il Confronto Diacronico: critérios e seleção dos textos

Para poder utilizar de maneira mais proveitosa os dados do *Corpus Stammerjohann*, a equipe do Laboratório *Lablita* organizou, em 2005, o *Corpus per il Confronto Diacronico*, pensado e realizado com a finalidade de ser utilizado em pesquisas sobre a fala em perspectiva diacrônica (SCARANO, 2005; MONEGLIA *et al.*, 2008; SCARANO & SIGNORINI, 2003; MONEGLIA & SCARANO, 2008). O *corpus* compõe-se de uma amostragem do *corpus Lablita*, replicando a estrutura, o número de palavras totais e em todas as suas subdivisões, e a tipologia dos textos do *Corpus Stammerjohann*. A estrutura criada para esse novo *corpus* busca reproduzir com fidelidade a variação diafásica do *corpus* mais antigo: há uma variedade maior de situações justamente para poder compensar a diversidade obtida nas gravações ambiente originais. Não sendo possível obter textos com idênticas características de falantes, preferiu-se optar por uma escolha que levasse em conta situações cujas características diafásicas, diastráticas e de conteúdo pudessem garantir um uso mais próximo possível da variedade lexical regional.

Os textos são transcritos no formato *CHAT* e a qualidade acústica é geralmente alta, A e B, mas em alguns casos C, mais próxima àquela apresentada pelo *Corpus Stammerjohann*.

O quadro abaixo apresenta o *corpus* de maneira mais detalhada:

CONTEXTO	TIPO	INTERAÇÃO	NÚMERO DE TEXTOS
FAMILIAR 22.784 palavras	NÃO-REGULADO 22.784 palavras	MONÓLOGO: 0 palavras	0
		DIÁLOGO: 9.415 palavras	6
		CONVERSAÇÃO: 13.369 palavras	7
PRIVADO 36.647 palavras	NÃO-REGULADO 25.840 palavras	MONÓLOGO: 0 palavras	0
		DIÁLOGO: 0 palavras	0
		CONVERSAÇÃO: 25.840 palavras	15
	REGULADO 10.655 palavras	MONÓLOGO: 3.475 palavras	3
		DIÁLOGO: 7.180 palavras	5
		CONVERSAÇÃO: 0 palavras	0
PÚBLICO 23.134 palavras	NÃO-REGULADO 16.100 palavras	MONÓLOGO: 0 palavras	0
		DIÁLOGO: 0 palavras	0
		CONVERSAÇÃO: 16.100 palavras	9
	REGULADO 7.034 palavras	MONÓLOGO: 1.219 palavras	1
		DIÁLOGO: 5.815 palavras	4
		CONVERSAÇÃO: 0 palavras	0
TELEFÔNICO 3.343 palavras		3.343 palavras	5
TOTAL		85.908 PALAVRAS	55

Quadro 7: Estrutura do *corpus Corpus per il Confronto Diacronico*, com número de palavras e textos para cada campo da estrutura.

Comparando os dados acima com aqueles do quadro 6 é perfeitamente visível que foi feito um grande trabalho de seleção para obter números praticamente idênticos em totais e parciais para cada campo do *corpus*. Conseguiu-se manter a estrutura em todos os detalhes diferindo apenas no número de textos, que passou de 32 para 55.

Como foi dito acima, o *corpus* foi pensado para estudos de análise lexical. A pesquisa que aqui apresentamos aborda a questão da expressão do sujeito e do uso dos clíticos no vernáculo florentino que, como será detalhado adiante, apresenta peculiaridades que o distinguem até mesmo da região rural em volta da cidade de Florença, e mais ainda do vernáculo toscano e do italiano padrão. Por este motivo, optou-se por refinar o *Corpus per il Confronto Diacronico* retirando todos os textos em que houvesse falantes não florentinos.

Desta forma, tentamos contornar o risco de coletar dados que não sejam rigorosamente de falantes do vernáculo em análise.

Foram extraídos 40 textos, para um total de 58.299 palavras, tamanho um pouco inferior, mas ainda comparável com o *Corpus Stammerjohann*. Quanto ao balanceamento, não foi possível mantê-lo, porque foi necessário excluir os textos que apresentavam falantes não florentinos.

O quadro abaixo expõe este novo *corpus* que chamaremos de *Corpus de Florentinos*:

CONTEXTO	TIPO	INTERAÇÃO	NÚMERO DE TEXTOS
FAMILIAR 21.232 palavras	NÃO-REGULADO 21.232 palavras	MONÓLOGO: 0 palavras	0
		DIÁLOGO: 9.415 palavras	6
		CONVERSAÇÃO: 11.817 palavras	6
PRIVADO 24.862 palavras	NÃO-REGULADO 16.641 palavras	MONÓLOGO: 0 palavras	0
		DIÁLOGO: 0 palavras	0
		CONVERSAÇÃO: 16.641 palavras	12
	REGULADO 8221 palavras	MONÓLOGO: 1.958 palavras	2
		DIÁLOGO: 6.263 palavras	4
		CONVERSAÇÃO: 0 palavras	0
PÚBLICO 8.862 palavras	NÃO-REGULADO 7.308 palavras	MONÓLOGO: 0 palavras	0
		DIÁLOGO: 0 palavras	0
		CONVERSAÇÃO: 7.308 palavras	4
	REGULADO 1.554 palavras	MONÓLOGO: 0 palavras	0
		DIÁLOGO: 1.554 palavras	1
		CONVERSAÇÃO: 0 palavras	0
TELEFÔNICO 3.343 palavras		3.343 palavras	5
TOTAL		58.299 PALAVRAS	40

Quadro 8: Estrutura do *Corpus de Florentinos*, com número de palavras e textos para cada campo da estrutura.

Pode-se observar que o contexto familiar foi quase que totalmente mantido, tendo sido retirado somente um texto, ou 1.552 palavras do total. O contexto privado acabou sendo desmembrado de forma substancial com a retirada de quatro textos, equivalentes a 11.785 palavras, sendo mais forte a perda na parte não-regulada. O contexto público foi o que mais sofreu com a ausência dos não-florentinos: passou-se de 23.134 palavras para 8.862 palavras.

O *Corpus de Florentinos* representa, em termos de número de palavras, cerca de 68% do *Corpus per il Confronto Diacrônico*.

Do ponto de vista da distribuição interna dos textos e números de palavras nos dois *corpora*, o balanceamento permanece muito próximo ao original no contexto privado: no *Corpus per il Confronto Diacrônico* este representa 42,55% do número de palavras, enquanto no *Corpus de Florentinos* é 42,65% do total. Internamente a este contexto também os valores mantêm-se próximos ao balanceamento ideal: as interações não reguladas representam 30,13% do total de palavras no primeiro corpus e 28,54% no segundo; aquelas reguladas 12,42% no primeiro e 14,10% no segundo.

As divergências maiores encontram-se nos contextos familiares e públicos, como mostra a tabela abaixo:

CONTEXTO	TIPO	INTERAÇÃO
FAMILIAR <i>Corpus per il Confronto</i> <i>Diacrônico: 26,57</i> Corpus de Florentinos: 36,42	NÃO-REGULADO <i>Corpus per il Confronto</i> <i>Diacrônico: 26,57</i> Corpus de Florentinos: 36,42	MONÓLOGO: 0 palavras
		DIÁLOGO <i>Corpus per il Confronto Diacrônico:10,98</i> Corpus de Florentinos: 16,15
		CONVERSAÇÃO <i>Corpus per il Confronto Diacrônico:15,59</i> Corpus de Florentinos: 20,27
PRIVADO <i>Corpus per il Confronto</i> <i>Diacrônico:42,55</i> Corpus de Florentinos:42,65	NÃO-REGULADO <i>Corpus per il Confronto</i> <i>Diacrônico:30,13%</i> Corpus de Florentinos: 28,54	MONÓLOGO: 0
		DIÁLOGO: 0
		CONVERSAÇÃO <i>Corpus per il Confronto Diacrônico:30,13</i> Corpus de Florentinos: 28,54
	REGULADO <i>Corpus per il Confronto</i> <i>Diacrônico:12,42</i> Corpus de Florentinos: 14,10	MONÓLOGO <i>Corpus per il Confronto Diacrônico:4,05</i> Corpus de Florentinos: 3,36
		DIÁLOGO: <i>Corpus per il Confronto Diacrônico:8,37</i> Corpus de Florentinos: 10,74
		CONVERSAÇÃO: 0
PÚBLICO <i>Corpus per il Confronto</i> <i>Diacrônico:26,97</i> Corpus de Florentinos: 15,20	NÃO-REGULADO <i>Corpus per il Confronto</i> <i>Diacrônico:18,77</i> Corpus de Florentinos: 12,54	MONÓLOGO: 0
		DIÁLOGO: 0
		CONVERSAÇÃO <i>Corpus per il Confronto Diacrônico:18,77</i> Corpus de Florentinos: 12,54
	REGULADO <i>Corpus per il Confronto</i> <i>Diacrônico:8,2</i> Corpus de Florentinos: 2,67	MONÓLOGO <i>Corpus per il Confronto Diacrônico:1,42</i> Corpus de Florentinos: 0
		DIÁLOGO <i>Corpus per il Confronto Diacrônico:6,78</i> Corpus de Florentinos: 2,67
		CONVERSAÇÃO: 0
TELEFÔNICO <i>Corpus per il Confronto</i> <i>Diacrônico:3,9</i> Corpus de Florentinos: 5,73		<i>Corpus per il Confronto Diacrônico:3,9</i> Corpus de Florentinos: 5,73
TOTAL		100

Quadro 9: Comparação em termos percentuais entre o *Corpus per il Confronto Diacrônico* e o *Corpus de Florentinos*, para cada campo da estrutura.

Pode-se visualizar, no quadro acima, que no contexto familiar o *Corpus per il Confronto Diacrônico* apresenta 26,57% do total de palavras, enquanto o *Corpus de Florentinos* possui uma representação maior desse tipo de interação, com 36,42% do total. No caso do contexto público a situação é inversa: o *Corpus per il Confronto Diacrônico* apresenta uma porcentagem maior, 26,97% do total de palavras enquanto o *Corpus de Florentinos* exibe 15,2% do total. No caso das interações telefônicas, o número de palavras

permaneceu idêntico, não tendo sido feitos recortes. Contudo, em relação aos *corpora* a que pertencem, no *Corpus per il Confronto Diacrônico* equivale a 3,9% do total enquanto no *Corpus de Florentinos* a 5,73%.

O corte que fizemos do *corpus* original foi bastante grande. Não obstante, acreditamos que tenha sido necessário, tendo em vista as finalidades da pesquisa. Certamente, na análise dos dados, este fator deverá ser levado em conta.

4. RESULTADOS

4.1. Resultados do PB

4.1.1. Primeira varredura do corpus com o software *AntConc*

O C-ORAL-BRASIL possui, entre seus arquivos, todas as transcrições no formato *.txt* em *UTF8*, o mais legível pela maioria dos *softwares*. Após alguns testes com programas como o *TextStat*¹²⁶, que se mostrou pouco confiável, retornando resultados diferentes em suas leituras, e o *Notepad++*¹²⁷, cujas funções são limitadas para a análise requerida, o *software* escolhido para a captação dos dados do *corpus* C-ORAL-BRASIL foi o *AntConc*¹²⁸. Trata-se de um programa gratuito projetado por Laurence Anthony, estudioso de informática com interesses linguísticos. A interface do *software* é bastante intuitiva e seus usos são muitos: desde a busca por termos até aquela por expressões regulares. O *AntConc* permite carregar um simples arquivo ou até mesmo todos os textos de um *corpus*. A janela de visualização dos resultados permite ver parte do contexto e, clicando no termo, obtém-se o texto inteiro.

Após ter carregado os arquivos do *corpus* no *software*, foi efetuada a busca das seis formas pronominais em análise. O resultado de cada uma foi, em seguida, transcrito na tabela *Excel*, previamente preparada, e que foi preenchida com os dados necessários e com as observações sobre distribuição e informantes.

4.1.2. Resultados parciais das formas e ocorrências encontradas

O primeiro resultado obtido com a varredura do *corpus* é o número de ocorrências brutas. Destas foram retiradas seis ocorrências em que foram transcritas letras do alfabeto, como a letra “c”. Todos os enunciados que apresentaram as formas estudadas foram ouvidos atentamente através do programa *WinPitch* (MARTIN, 2005). Os casos de transcrição errada foram corrigidos e computados corretamente. Queremos ressaltar que estes casos foram muito raros e na maioria deles foi possível constatá-los apenas na etapa de análise acústica. De fato, nos seguintes enunciados, em que à oitava ouviu-se a forma *ocê*, somente analisando-se seu

¹²⁶ <http://neon.niederlandistik.fu-berlin.de/textstat/>

¹²⁷ <http://notepad-plus-plus.org/>

¹²⁸ http://www.antlab.sci.waseda.ac.jp/antconc_index.html

espectro foi possível perceber que se tratava da forma *cê*: bfamcv04-453; bfamcv05-450; bfamcv13-34; bfamcv25-248¹²⁹.

Desta maneira, obtiveram-se todos os dados passíveis de análise efetiva. Lembramos que nesta etapa a análise é feita em relação às classificações de distribuição citadas na parte metodológica. Em seguida, a análise acústica avaliará as durações das formas em relação ao resto do enunciado e às unidades tonais.

A tabela abaixo apresenta os resultados desta primeira etapa:

POSIÇÕES		<i>CÊ</i>	<i>CÊS</i>	<i>OCÊ</i>	<i>OCÊS</i>	<i>VOCÊ</i>	<i>VOCÊS</i>
PRÉ-VERBAL		1974	121	177	10	225	16
PÓS-VERBAL	OBJETO DE VERBO	1	0	14***	3	5	4
	OBJETO DE PREPOSIÇÃO	36	1	51	14	46	11
	SUJEITO POSPOSTO	1	0	10	0	11	2
	TOTAL	38	1	75	17	62	17
EM ISOLAMENTO EM UNIDADE TONAL		13*	0	8	1	6	1
EM ISOLAMENTO ABSOLUTO		1**	0	0	0	6	2
CONTRASTIVO		0	0	2****	0	7	0
ENUNCIADO INTERROMPIDO		89	3	5	2	8	3
RETRACTING		27	1	9	1	5	0
NÚCLEO SINTÁTICO DO ENUNCIADO		1	0	2	0	0	0
TOTAL		2129	126	277	31	319	39

Tabela 3: apresentação das ocorrências e distribuição das formas pronominais de segunda pessoa, subdivididas por posições e funções, encontradas no C-ORAL-BRASIL.

*4 são também objeto de preposição e 9 sujeito.

** também objeto de preposição.

***um deles é um caso de duplo acusativo.

**** um dos dois é também objeto de preposição.

A tabela evidencia o altíssimo número de ocorrências da forma *cê(s)*, singular e plural, com 2255 casos, enquanto *ocê(s)* foi encontrado 308 vezes e *você(s)* 358. Em termos percentuais significa que, no *corpus* C-ORAL-BRASIL a forma *cê(s)* equivale a 77,3% do total de ocorrências das formas de segunda pessoa, *ocê(s)* a 10,5% e *você(s)* a 12,2%, como mostrado pelo gráfico abaixo.

¹²⁹ É este mais um fator que comprova a grande confiabilidade das transcrições do *corpus*, pois somente uma análise muito minuciosa foi capaz de detectar tais incorreções. Veja-se a este respeito o capítulo sobre a validação estatística em Mello *et al.* (2012). Adiante esses textos serão retomados e será inserido seu áudio, para apreciação do leitor.

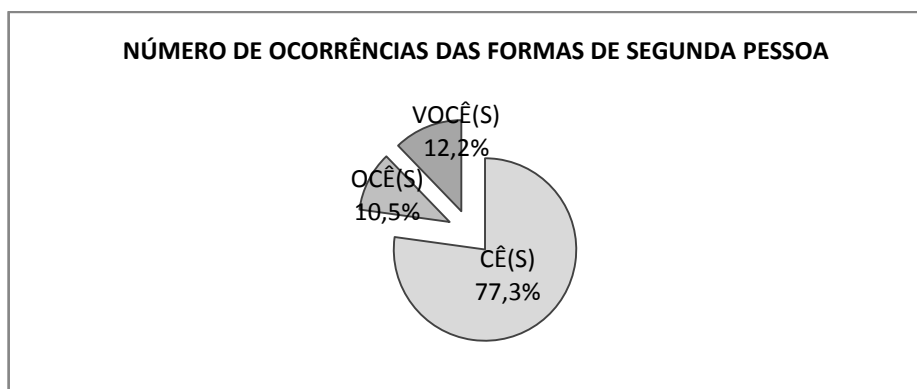


Gráfico 1: visualização das porcentagens de uso das formas pronominais de segunda pessoa encontradas no C-ORAL-BRASIL.

Como os estudos anteriores haviam previsto, mas não haviam quantificado, há um uso muito grande da forma reduzida em relação às bissilábicas.

Antes de passar a uma análise das posições ocupadas por essas formas, apontamos para a desproporção apresentada entre os singulares e os plurais.

Observando-se os resultados em forma de gráfico tem-se:

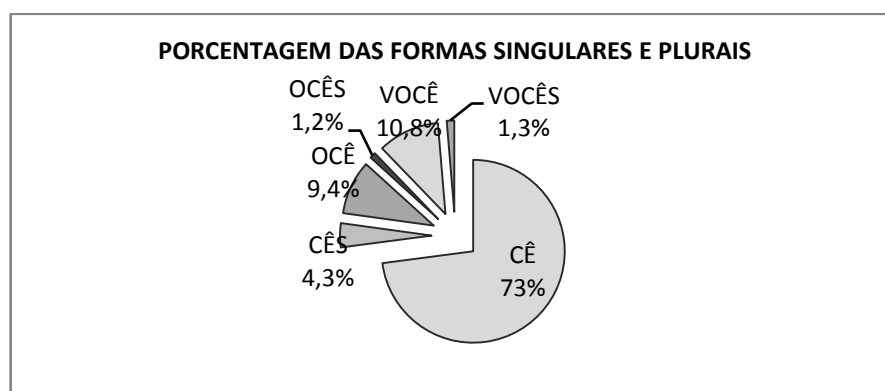


Gráfico 2: visualização das porcentagens de uso das formas pronominais de segunda pessoa, singulares e plurais, encontradas no C-ORAL-BRASIL.

Nota-se de imediato que em todas as formas o singular é prevalente sobre o total: enquanto o *cê* no singular apresenta 73% das ocorrências, seu plural é de somente 4,3%. O mesmo fenômeno ocorre com as formas *você*, com 11,8% no singular e 1,3% no plural e *ocê*, 9,4% no singular e 1,2% no plural.

Visto de outra perspectiva, esses dados significam que, do total da forma *cê(s)*, somente 5,5% representa o plural. As formas bissilábicas apresentam uma porcentagem um pouco maior dos plurais: do total de *ocê(s)*, 10% são da forma no plural, e do total de *você(s)* 10,9% são plurais. Não é objetivo do presente trabalho, mas seria interessante buscar entender as razões de tal comportamento.

Os dados, descartando-se os casos de enunciados com interrupção e *retracting*¹³⁰ são relacionados abaixo em quadros separados.

POSIÇÕES	<i>CÊ</i>	<i>CÊS</i>	<i>OCÊ</i>	<i>OCÊS</i>	<i>VOCÊ</i>	<i>VOCÊS</i>
TOTAL	2129	126	277	31	319	39
INTERRUPÇÕES	89	3	5	2	8	3
NOVO TOTAL	2040	123	272	29	311	36

Tabela 4: total das ocorrências das formas pronominais de segunda pessoa excluindo-se os casos de interrupções do enunciado.

A exclusão dos casos de interrupção não altera substancialmente os dados. Visto em forma de gráfico e em termos percentuais isto é ainda mais evidente:

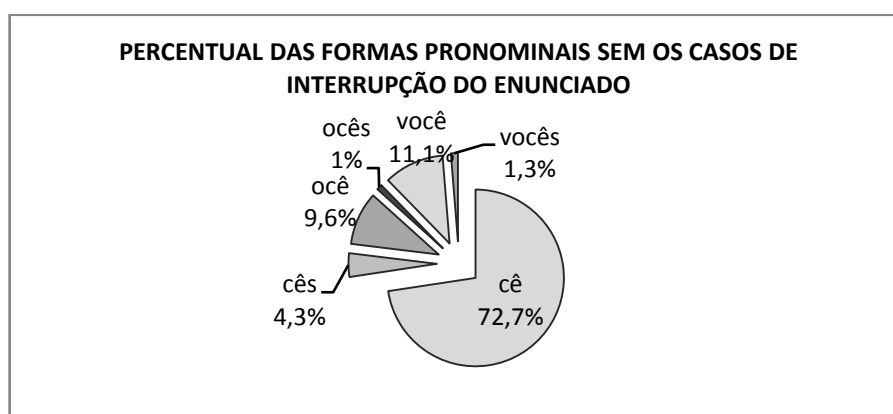


Gráfico 3: visualização das porcentagens de uso das formas pronominais de segunda pessoa, singulares e plurais, excluídos os casos de interrupção do enunciado, encontradas no C-ORAL-BRASIL.

Como é possível observar, as porcentagens não são idênticas, mas estão bem próximas dos valores totais mostrados anteriormente.

Relacionamos agora um quadro em que excluímos da contagem os casos de *retracting*:

POSIÇÕES	<i>CÊ</i>	<i>CÊS</i>	<i>OCÊ</i>	<i>OCÊS</i>	<i>VOCÊ</i>	<i>VOCÊS</i>
TOTAL	2129	126	277	31	319	39
RETRACTING	27	1	9	1	5	0
NOVO TOTAL	2102	125	268	30	314	39

Tabela 5: total das ocorrências das formas pronominais de segunda pessoa excluindo-se os casos de *retracting*.

Mesmo neste caso, em que parece ocorrer uma alteração do quadro geral, observando a situação em termos percentuais, notamos que não há mudanças substanciais.

¹³⁰ A este respeito, veja-se parágrafo 3.1.1.3.

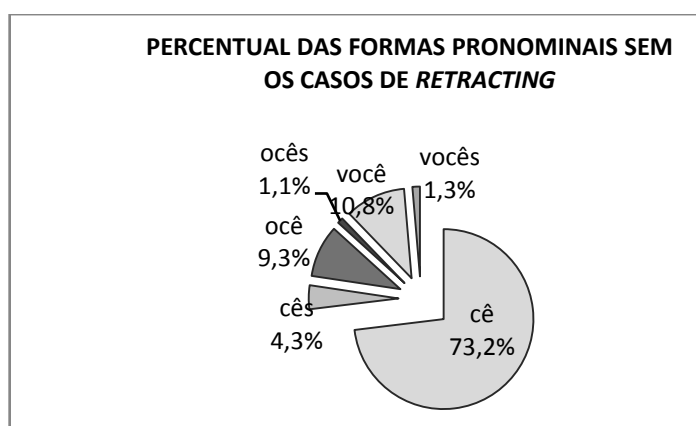


Gráfico 4: visualização das porcentagens de uso das formas pronominais de segunda pessoa, singulares e plurais, excluídos os casos de *retracting*, encontradas no C-ORAL-BRASIL.

Percebe-se que as porcentagens continuam muito próximas àquelas dos gráficos anteriores.

Por fim, subtraímos do total das formas encontradas todos os casos, seja de interrupção do enunciado, seja de *retracting*.

POSICÕES	<i>CÊ</i>	<i>CÊS</i>	<i>OCÊ</i>	<i>OCÊS</i>	<i>VOCÊ</i>	<i>VOCÊS</i>
TOTAL	2129	126	277	31	319	39
INTERRUPÇÕES	89	3	5	2	8	3
<i>RETRACTING</i>	27	1	9	1	5	0
NOVO TOTAL	2013	122	263	28	306	36

Tabela 6: total das ocorrências das formas pronominais de segunda pessoa excluindo-se os casos de interrupção do enunciado e de *retracting*.

Novamente os dados surpreendem pela consistência: uma visualização em gráfico e em termos percentuais mostra o seguinte.

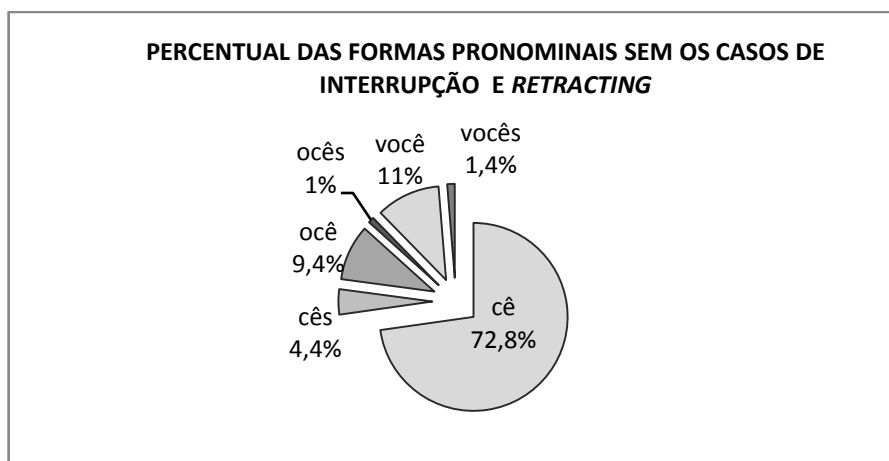


Gráfico 5: visualização das porcentagens de uso das formas pronominais de segunda pessoa, singulares e plurais, excluídos os casos de interrupção e de *retracting*, encontradas no C-ORAL-BRASIL.

As porcentagens continuam se mantendo muitíssimo próximas àquelas obtidas em todos os outros casos: desde os dados brutos, até as várias subtrações de tipos diferentes de disfluência.

Passemos agora à análise das formas com base na primeira classificação feita. Para não criar confusões, a partir deste momento utilizaremos os totais das várias formas subtraídos dos casos de interrupção e *retracting*. É possível depreender do quadro acima que, para todas as formas analisadas, *cê(s)*, *ocê(s)* e *você(s)*, a posição favorita é a pré-verbal, associada à função de sujeito.

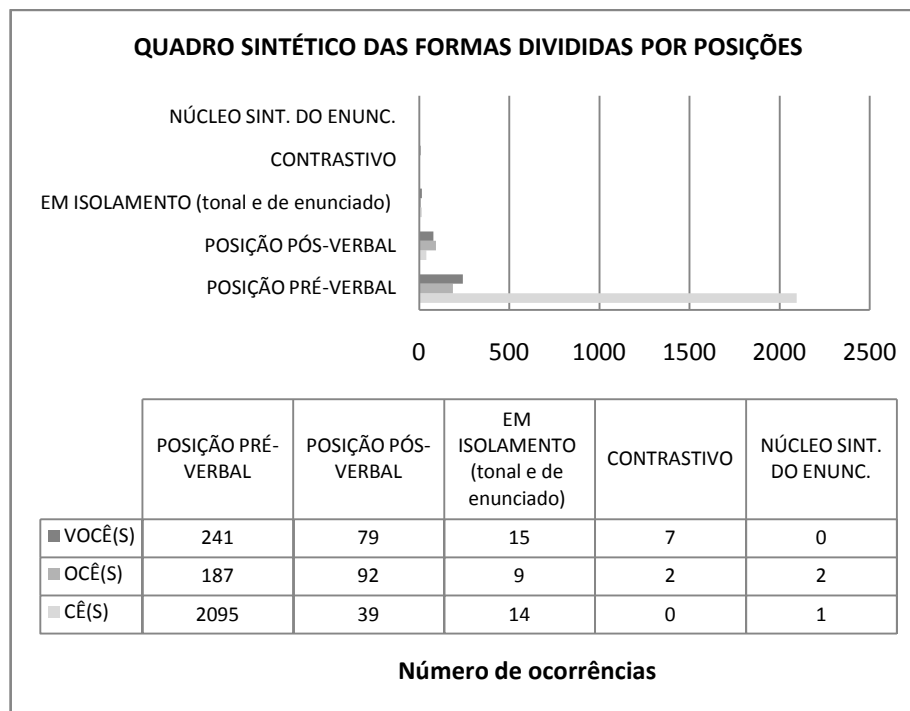


Gráfico 6: exposição sintética das ocorrências das formas pronominais de segunda pessoa divididas por posições, extraídas do C-ORAL-BRASIL.

Com respeito à distribuição, o quadro acima retoma os dados da tabela 1, e mostra a preferência pelo uso das três formas em posição pré-verbal, naquela que é a posição canônica do sujeito. A forma *cê(s)* ocorre como tal 2095 vezes, *ocê(s)* 187 vezes e *você(s)* 241 vezes. Em termos percentuais a posição pré-verbal fica assim dividida:

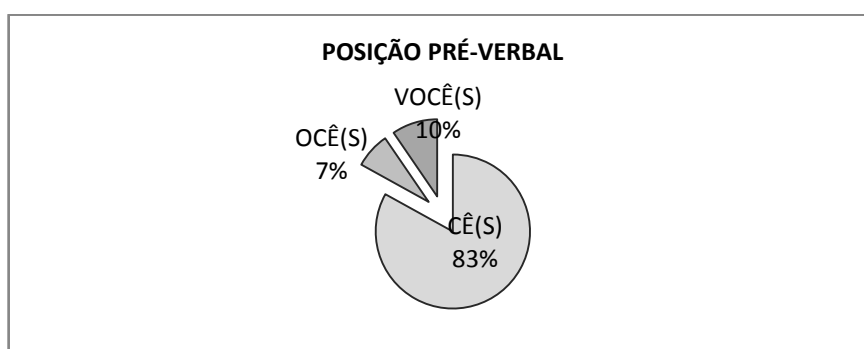


Gráfico 7: visualização das porcentagens de uso das formas pronominais de segunda pessoa na posição pré-verbal.

Este primeiro quadro mostra com clareza a predominância da forma *cê(s)* em posição pré-verbal, com 83% das ocorrências, em relação às formas *você(s)*, em segundo lugar, com 10% das ocorrências, e *ocê(s)*, com 7% das ocorrências.

O gráfico abaixo permite visualizar a situação no caso da posição pós-verbal, que será detalhada em pormenores caso a caso logo em seguida.

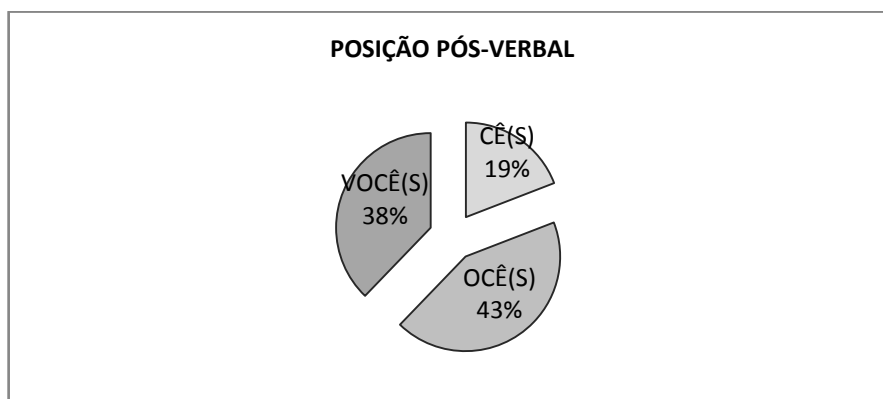


Gráfico 8: visualização das porcentagens de uso das formas pronominais de segunda pessoa na posição pós-verbal.

A situação muda de maneira clara: a maior porcentagem de uso refere-se à forma *ocê(s)*, com 43% das ocorrências, seguida da forma *você(s)*, com 38% dos casos e *cê(s)*, com 19% das ocorrências. Podemos também considerar as formas *ocê(s)* e *você(s)* como uma única forma bissilábica: desse modo, elas constituiriam 81% dos casos em posição pós-verbal.

Observemos agora mais em detalhe a distribuição das formas: o gráfico abaixo ilustra a subdivisão em posição pós-verbal:

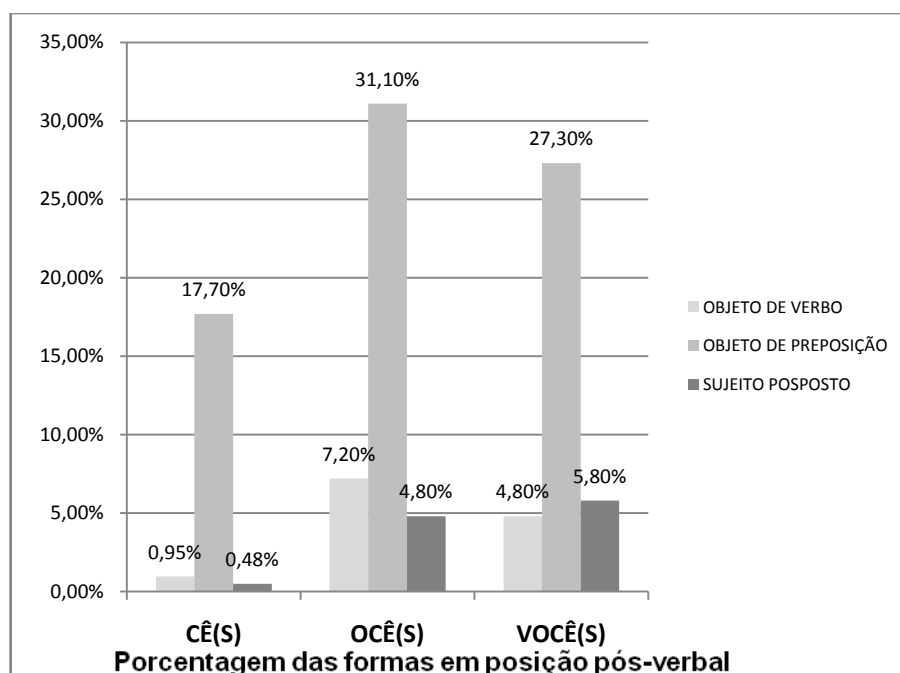


Gráfico 9: subdivisão em percentual das formas pronominais de segunda pessoa na posição pós-verbal.

Contrariamente às afirmações de boa parte da literatura, foram encontradas no *corpus* duas ocorrências de *cê* como objeto direto e uma como sujeito posposto. Como objeto de preposição, o C-ORAL-BRASIL apresentou 37 ocorrências da forma *cê(s)*. Esses dados são significativos, pois, como foi visto na seção de revisão da literatura, grande parte dos autores negam tais possibilidades, mas a análise de dados colhidos em contexto de uso real e espontâneo contradiz essa afirmação.

A forma *ocê(s)* ocorre 16 vezes como objeto de verbo, 10 como sujeito posposto, e é aquela mais utilizada como objeto de preposição, com 66 ocorrências.

A forma *você(s)* apresentou 10 ocorrências como objeto de verbo e 12 como sujeito posposto. Como objeto de preposição ocorreu 57 vezes.

O gráfico acima aponta quão significativo é a presença no *corpus* da forma *cê(s)*: como objeto de preposição, apresenta 17,7% das ocorrências totais das três formas na posição pós-verbal. No caso de considerarmos as três formas somente na função de objeto de preposição, o peso da forma *cê(s)* aumenta, representando 23,10% das ocorrências contra *ocê(s)*, com 41,25% e *você(s)*, com 35,65%. Lembramos que essas porcentagens e esses dados não podem ser considerados simplesmente acidentes na realização linguística do falante, mas devem ser considerados como seu projeto de fala.

O fato de ter encontrado em um *corpus* de fala espontânea ocorrências de *cê* como objeto direto, indireto e sujeito posposto, já por si, apesar de sua exiguidade, confirma com dados empíricos algo que suspeitávamos ser possível, merecendo ser aprofundado e não podendo ser ignorado.

Passemos agora a uma análise das formas em isolamento, distinguindo entre isolamento em unidade tonal e isolamento absoluto, ou seja, isolamento no enunciado. A tabela abaixo inclui tanto as formas não preposicionadas quanto as preposicionadas, desde que estejam isoladas em unidade tonal ou no enunciado.

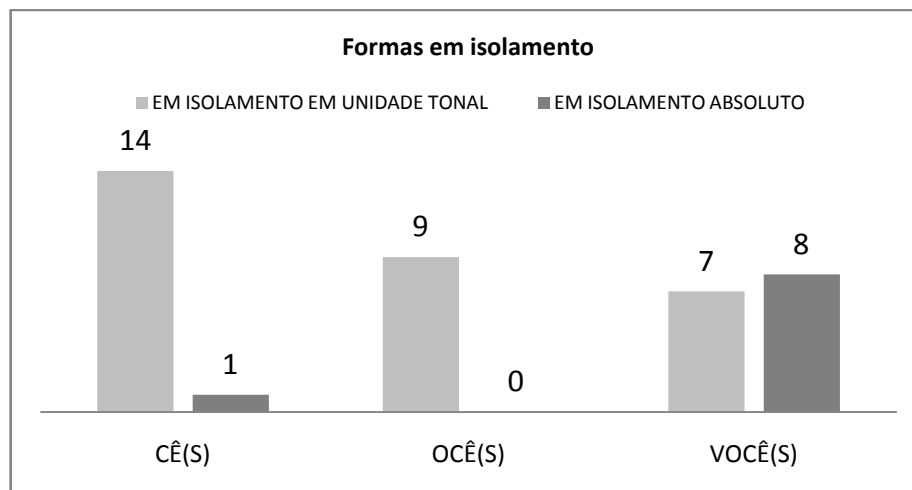


Gráfico 10: subdivisão das formas pronominais de segunda pessoa em isolamento.

Em termos de número de ocorrências, a forma *cê(s)*, em isolamento em unidade tonal, é aquela que apresenta o maior número de casos, 14. Da forma *ocê(s)* foram encontrados nove casos e da forma *você(s)*, sete. Contudo, esta última forma é aquela que apresenta o maior número de ocorrências de isolamento absoluto no enunciado, com oito casos, enquanto com a forma *cê(s)*¹³¹ foi encontrado apenas uma única vez e com a forma *ocê(s)* nenhum caso foi encontrado. Na literatura não se considera nunca a relação entre as formas e a sua posição relativa à quebra prosódica e, portanto, à função informacional. A nosso ver, essa relação é muito relevante e pode nos ajudar muito a entender a natureza do fenômeno como um todo. Essa questão será aprofundada na seção seguinte, a de análise dos dados.

Antes de concluir esta etapa, vejamos os mesmos dados em termos percentuais.

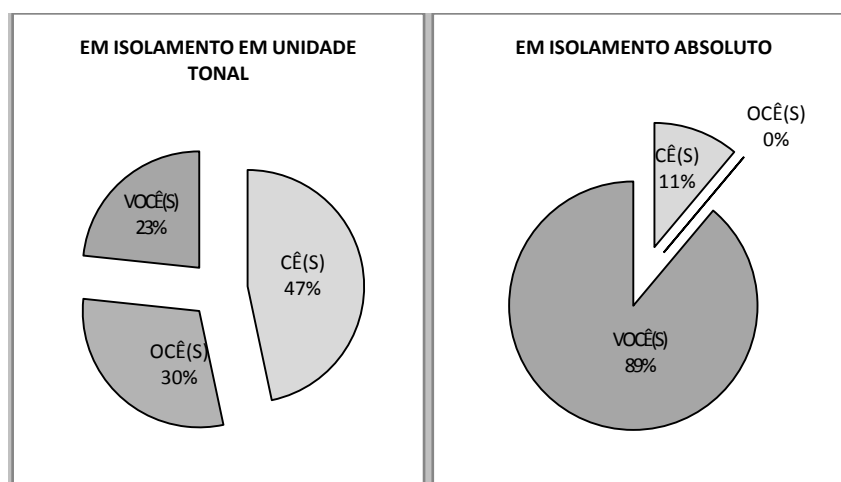


Gráfico 11: subdivisão das formas pronominais de segunda pessoa em isolamento, em termos percentuais.

¹³¹ Na realidade trata-se de um *cê* acompanhado por preposição (*p' cê*), portanto, a rigor, o isolamento não será absoluto, mas o classificamos desta forma por se tratar de um caso tão específico.

A situação mostra-se mais clara: no caso de isolamento em unidade tonal a forma *cê(s)* está presente em quase a metade dos casos (47%), *ocê(s)* em um terço (30%) e *você(s)* em pouco mais de um quarto (23%). Em relação ao isolamento absoluto, a situação é completamente diferente: 89% dos casos encontrados são da forma *você(s)*; *cê(s)* ocorre em 11% dos casos (o que, como dissemos, equivale a uma única ocorrência e após preposição) e *ocê(s)* não foi encontrado.

4.1.3. Resultados da forma *você(s)*

Cada forma será agora examinada em detalhes individualmente. Reportamos os dados da tabela 3 para que seja possível lembrar as ocorrências, retirando as formas em *retracting* e interrompidas:

POSIÇÕES		VOCÊ	VOCÊS
PRÉ-VERBAL		225	16
PÓS-VERBAL	OBJETO DE VERBO	5	4
	OBJETO DE PREPOSIÇÃO	46	11
	SUJEITO POSPOSTO	11	2
	TOTAL	62	17
EM ISOLAMENTO EM UNIDADE TONAL		6	1
EM ISOLAMENTO ABSOLUTO		6	2
CONTRASTIVO		7	0
NÚCLEO SINTÁTICO DO ENUNCIADO		0	0
TOTAL		306	36

Tabela 7: apresentação das ocorrências e distribuição da forma *você(s)*, encontradas no C-ORAL-BRASIL.

As ocorrências totais das formas *você(s)*, retirados os casos de disfluência que foram relatados acima, são de 306 casos no singular e 36 no plural. Isto significa que 89% dos casos são encontrados no singular e 11% no plural.

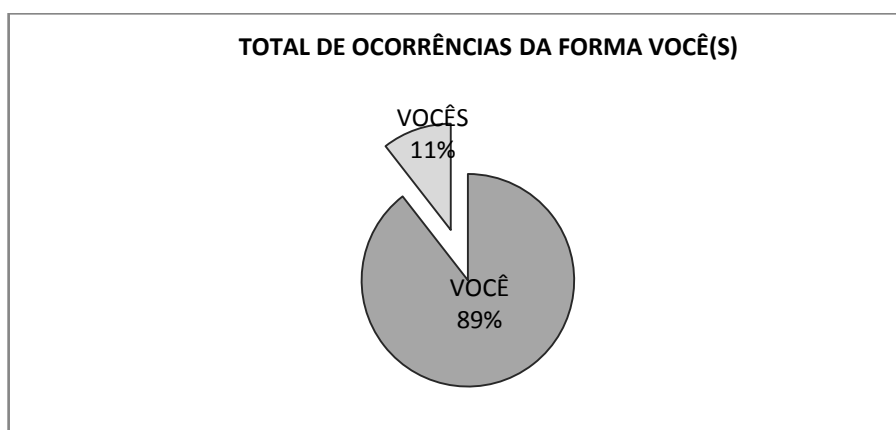


Gráfico 12: visualização das porcentagens de uso das formas pronominais de segunda pessoa *você(s)*.

Tal desproporção entre singular e plural nas formas estudadas já foi relatada mas, como dissemos, não será objeto do trabalho.

Conforme foi visto, a distribuição da forma *você(s)*, assim como as outras, mostra uma frequência maior na posição pré-verbal. Foram encontradas, nesta posição, 225 das 306 ocorrências de *você* no singular, e 16 das 36 de *vocês* no plural.

A posição pós-verbal merece uma avaliação mais detalhada, pois, seguindo os critérios aqui adotados, nela é necessário distinguir entre ocorrências de vários tipos. De maneira geral, as ocorrências de *você(s)*, em posição pós-verbal, somam 62 casos no singular e 17 no plural. Quando o pronome é objeto de verbo, foram encontradas cinco ocorrências no singular e quatro no plural. Na posição de objeto de preposição, está presente o maior número dessas formas: 46 no singular e 11 no plural. Os casos de sujeito posposto são dez no singular e duas no plural.

Apresentamos os dados em termos percentuais para melhor visualização:

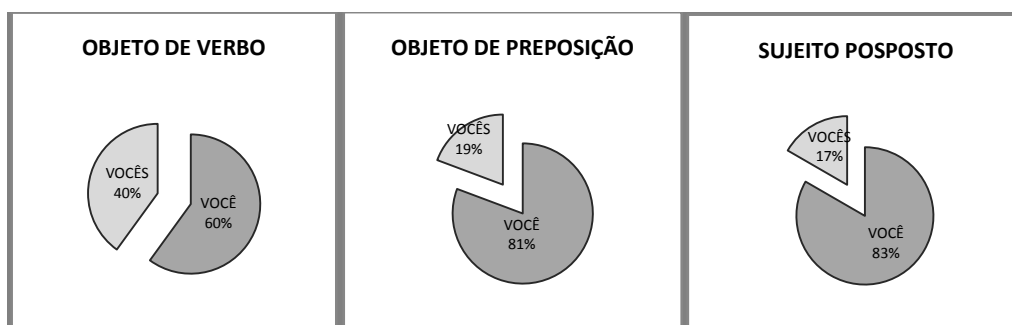


Gráfico 13: subdivisão da forma *você(s)* na posição pós-verbal em termos percentuais.

Como objeto de verbo a forma apresenta uma pequena prevalência do singular, com 60% das ocorrências, contra 40% no plural. A desproporção entre singular e plural é, ao contrário, muito evidente quando o pronome é objeto de preposição (81% dos casos no singular e 19% no plural) e quando o pronome é um sujeito posposto (83% das ocorrências no singular e 17% no plural).

Observemos agora as ocorrências de *você(s)* em casos de isolamento em enunciado ou em unidade tonal. Em relação ao isolamento em unidade tonal, a forma apresenta seis ocorrências no singular e uma no plural. Já em isolamento absoluto foram encontrados seis casos no singular e dois no plural.

O gráfico abaixo detalha os dados em termos percentuais.

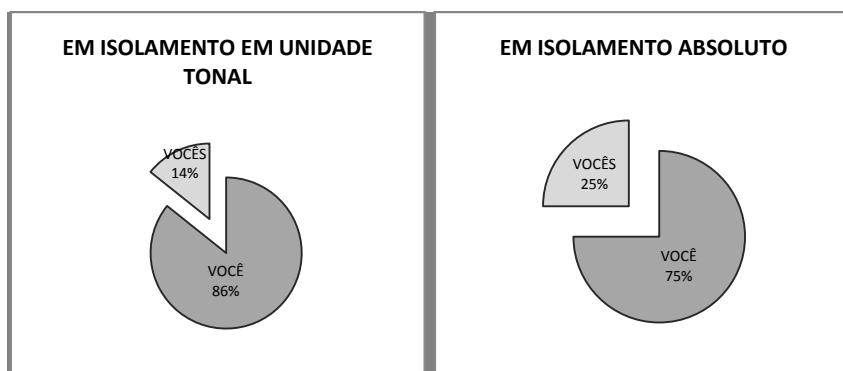


Gráfico 14: subdivisão da forma *você(s)* em isolamento, em termos percentuais.

Observemos, por fim, os casos de *você(s)* com valor contrastivo. A forma consta de sete ocorrências no singular e nenhuma no plural.

4.1.4. Resultados da forma *ocê(s)*.

Ao se analisar a forma *ocê(s)*, será reportada novamente parte da tabela 3, retirando as formas em *retracting* e interrompidas, para lembrar os dados.

POSIÇÕES		<i>OCÊ</i>	<i>OCÊS</i>
PRÉ-VERBAL		177	10
PÓS-VERBAL	OBJETO DE VERBO	14*	3
	OBJETO DE PREPOSIÇÃO	51	14
	SUJEITO POSPOSTO	10	0
	TOTAL	75	17
EM ISOLAMENTO EM UNIDADE TONAL		8	1
EM ISOLAMENTO ABSOLUTO		0	0
CONTRASTIVO		2***	0
NÚCLEO SINTÁTICO DO ENUNCIADO		2	0

Tabela 8: apresentação das ocorrências e distribuição da forma *ocê(s)*, encontradas no C-ORAL-BRASIL.

*um é um caso de duplo acusativo.

*** um dos dois é também objeto de preposição.

Neste caso, a tabela não apresenta a soma dos valores, pois em dois há coincidência de duas classificações de uma mesma forma e o pronome seria contado em mais de uma posição na tabela. O caso será exposto no exemplo 25, ao longo da discussão.

Preferimos, portanto, apresentar somente uma tabela com as várias posições discriminadas.

A forma *ocê(s)* apresenta 263 ocorrências no singular e 28 no plural. Em termos percentuais isso significa que *ocê* é responsável por 90,3% dos casos e *ocês* por 9,7%.

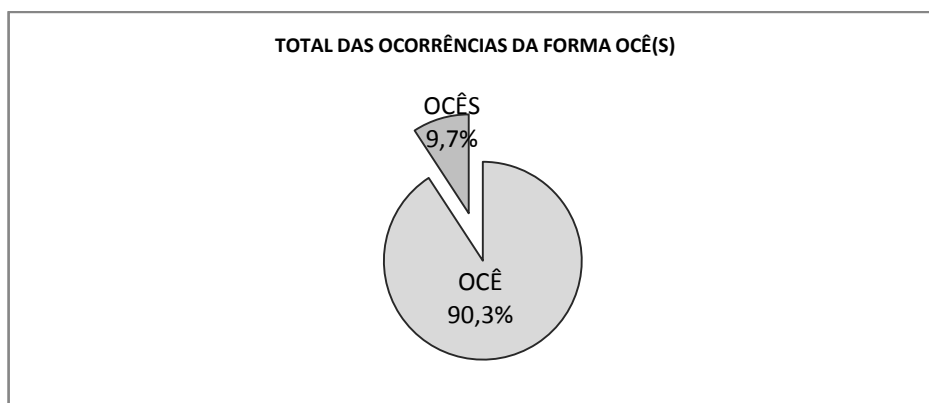


Gráfico 15: visualização das porcentagens de uso das formas pronominais de segunda pessoa *ocê(s)*.

Não vamos repetir a necessidade de uma análise futura mais detalhada sobre as causas da desproporção entre a forma no singular e no plural.

Quanto à sua distribuição, a primeira observação é que, mais uma vez, na posição pré-verbal a frequência do singular, 177 ocorrências, é muito maior que aquela no plural, com 10 ocorrências.

A posição pós-verbal merece destaque: como objeto de preposição, *ocê(s)* perfaz o número mais alto de ocorrências entre as três formas estudadas, apresentando 52 ocorrências no singular e 14 no plural. Também como objeto de verbo é a forma que apresenta o maior número de casos: 14 no singular e três no plural. Enfim, como sujeito posposto, aparece com 10 ocorrências, todas no singular.

O gráfico abaixo mostra estes dados em termos percentuais:

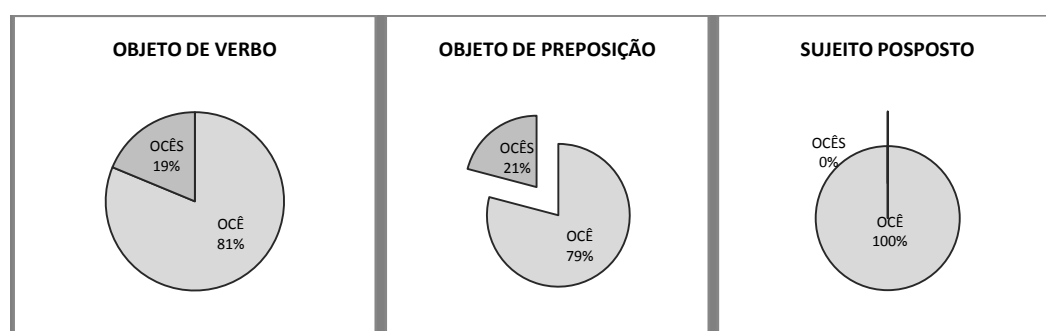



Gráfico 16: subdivisão da forma *ocê(s)* na posição pós-verbal em termos percentuais.

É evidente pelas imagens que a forma no singular prevalece em todos os casos, em 100% quando sujeito posposto, em 81% quando objeto de verbo e em 79% quando objeto de preposição.

É interessante, pela unicidade da ocorrência, citar o caso em que *ocê* é um duplo acusativo:

- (24) *ALO: [100] *aí / quando é três anos atrás mais ou menos / chega um filho dela / perto de mim / falou assim / ô Aloysio // [101] minha mãe morreu // [102] devia **ocê uma conta** na loja //*

(bfammn03-102) 

Pode-se notar que *ocê* é objeto direto do verbo “dever”, assim como “uma conta”.

Vejamos os casos de isolamento em unidade tonal ou enunciado. Como relatado acima, a forma *ocê(s)* não ocorre em nenhum caso em situação de isolamento absoluto. Quando em isolamento em unidade tonal, foram encontradas oito ocorrências no singular e uma no plural.

O gráfico traduz isto em termos percentuais:

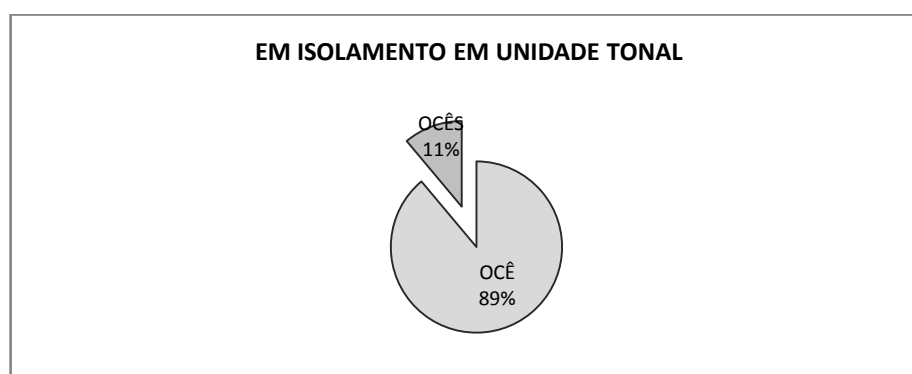



Gráfico 17: subdivisão da forma *ocê(s)* em isolamento em unidade tonal, em termos percentuais.

No singular as ocorrências equivalem a 89% dos casos enquanto no plural a 11%.

A forma *ocê(s)* aparece também em dois casos com valor contrastivo, sempre no singular: em um dos dois, contudo, é também objeto de preposição. Reportamos abaixo os casos:

- (25) *ONO: *eu cheguei primeiro que **ocê** //*

(bfamcv11-30) 


- (26) // *mas c' **ocê** e' nũ vai / eu / falo co Antônio Carlos todo dia / ele nem dá a mínima //*

(bfamcv10-8) 


No exemplo (25) tem-se uma situação com valor contrastivo. No (26), além do valor contrastivo, há a presença da preposição reduzida “c”.

Concluimos a exposição da forma *ocê(s)* com os únicos dois exemplos, somente no singular, de quando é núcleo sintático do enunciado:

(27) *LUC: [164] *contraventora / <cê fraga> // *MIC: [165] <Nossa / mas isso> <é desde nova // *FLL: [166] <eu só tenho amiga bandida> / sô / a Elisa é igualzinha / todo mundo é assim // *MIC: [167] *cê é assim / Elisa // **ocê com essa cara / de super nerd <sua> //****

(bfamcv26-168) 

(28) *REG: [52] *na hora o menino nasce de nove meses hoje em dia / a primeira coisa que es fazem / é todo um preparo e já põem na incubadora / e ali a criança fica no [1] um <tempão na> incubadora / até eles arrumarem / pôr roupinha / e nũ sei o quê // *BAL: [53] <hum hum> // *REG: [54] *em observação pra ver aquea primeira hora como que tá // *REG: [55] **ocê de sete meses e meio / sem / noção do que que eu ia tar passando / né // *REG: [56] *ai / ele assim / pra onde // *REG: [57] *falei assim / entra aqui no Barreiro // *REG: [58] *porque eu vi aqueas torres do [1] do &Le [2] &he / eu lembrei ali do Leroy / falei assim / No' / cê entra aqui / porque Barreiro tem muito hospital //*******

(bfammn28-55) 

4.1.5. Resultados da forma *cê(s)*.

Retomemos pela última vez parte da tabela 3 para visualizar os dados da forma *cê(s)*.

POSIÇÕES		<i>CÊ</i>	<i>CÊS</i>
PRÉ-VERBAL		1974	121
PÓS-VERBAL	OBJETO DE VERBO	1	0
	OBJETO DE PREPOSIÇÃO	36	1
	SUJEITO POSPOSTO	1	0
	TOTAL	38	1
EM ISOLAMENTO EM UNIDADE TONAL		13*	0
EM ISOLAMENTO ABSOLUTO		1**	0
CONTRASTIVO		0	0
NÚCLEO SINTÁTICO DO ENUNCIADO		1	0

Tabela 9: apresentação das ocorrências e distribuição da forma *cê(s)*, encontradas no C-ORAL-BRASIL.

*4 são também objeto de preposição e 9 sujeito.

** também objeto de preposição.

Também neste caso não foi posta a soma das ocorrências, pelos motivos expostos acima. Os exemplos de duplicidade serão evidenciados ao longo da discussão.

O total de ocorrências da forma *cê(s)* é de 2013 no singular e 122 no plural. Como já foi dito, das três formas estudadas, é a que mais ocorre no *corpus*. O primeiro fato que chama a atenção é a desproporção entre singular e plural, ainda mais forte que nas outras formas. Vejamos as porcentagens das ocorrências totais.

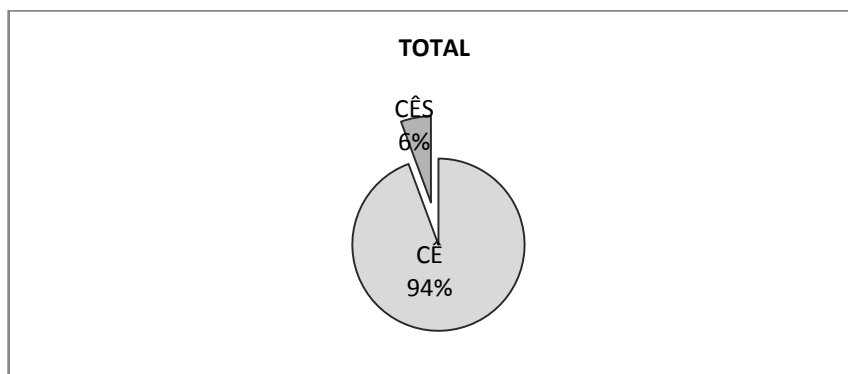


Gráfico 18: visualização das porcentagens de uso das formas pronominais de segunda pessoa *cê(s)*.

A forma *cê* comparece em 94% dos casos enquanto *cês* em 6%, bem menos em relação aos plurais *ocês* e *vocês*. O uso da forma no plural está quase toda concentrada na posição pré-verbal, com 121 ocorrências. O singular nesta posição apresenta 1974 ocorrências.

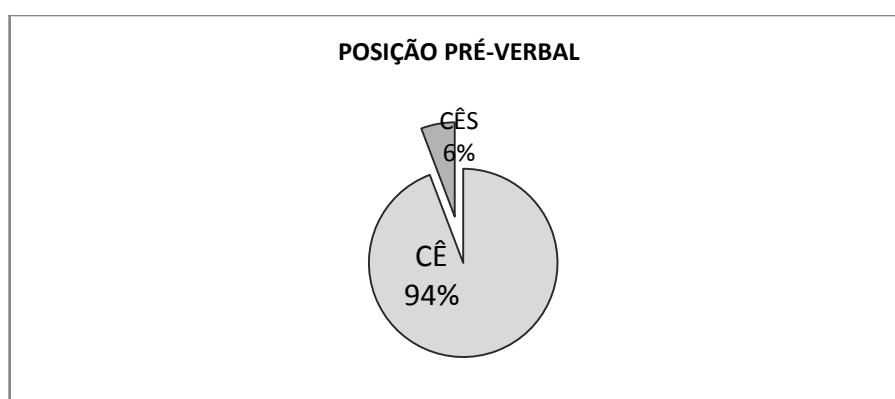


Gráfico 19: visualização das porcentagens de uso das formas pronominais de segunda pessoa *cê(s)* em posição pré-verbal.


A forma *cê* comparece novamente em 94% dos casos e *cês* em 6%.

A posição que é mais interessante observar, contudo, é aquela pós-verbal, pelo grande debate que gera, e do qual se discutiu na revisão da literatura. Posto que somente uma


ocorrência no plural foi encontrada, como objeto de preposição, o C-ORAL-BRASIL apresentou no singular apenas um caso de objeto de verbo e um de sujeito posposto. Já como objeto de preposição há 36 ocorrências desta forma.

Reportamos os casos de objeto de verbo e o do sujeito posposto em uma construção clivada, pelo interesse especial que despertam:

(29) *CEL: *chamou cê não / sô //*

(bfamcv05-240) 

(30) *BRU: [453] *agora é cê que faz //*

(bfamcv04-453) 

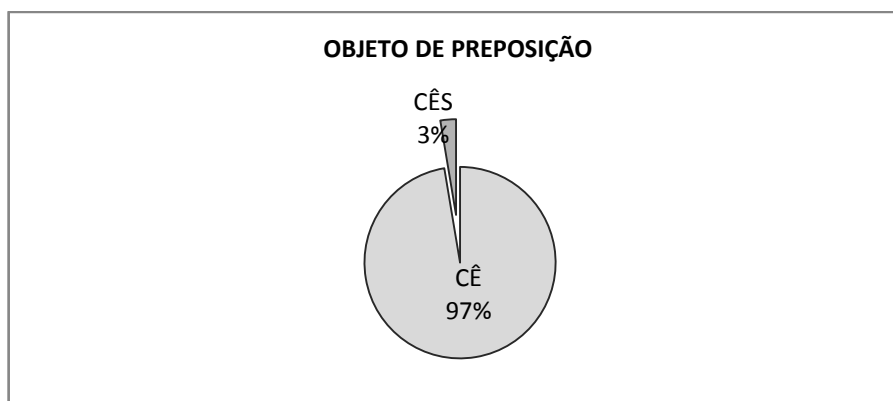



Gráfico 20: visualização das porcentagens de uso das formas pronominais de segunda pessoa *cê(s)* como objeto de preposição.

Como objeto de preposição, o singular aparece em 97% dos casos, no plural em 3%, lembrando que este dado corresponde a uma ocorrência apenas. Contudo, vale ressaltar que este único caso foi verificado através de análise acústica: a forma *cês* tinha sido transcrita como *ocês*. Mais uma vez isto mostra a importância desta etapa, dado que se trata de segmentos curtos e de difícil percepção sem a comprovação de um espectrograma.

Reportamos abaixo o caso citado:


(31) *HER: [66] *vou descontar no salário d' cês //*

(bfamcv18-66) 

A preposição utilizada é *de*, na forma contraída *d'*. Este foi o único caso encontrado com a preposição *de*. Nos casos no singular foram encontradas as seguintes preposições:


I. *com*: 28 ocorrências;

(32) // *a mamãe conversa **com cê*** //

(bfamcv17-5) 


II. *c'*, contração de *com*: 2 ocorrências;

(33) // *e [1] e lá [1] lá em Lagoa Santa o xxx falou **c' cê*** //

(bfamdl17-201) 

III. *p'*, contração de *para*: 6 ocorrências;


(34) **MAC: coa o café aí pra mim // fazendo favor / eu levo o arroz lá **p' cê*** //

(bfamcv18-18/19) 


Passemos agora à análise da forma *cê(s)* quando encontrada em isolamento, em unidade tonal, e no enunciado.

No plural não foram verificadas ocorrências dessa forma na situação descrita. No singular, há 13 ocorrências em isolamento em unidade tonal: quatro delas são também objeto de preposição e nove sujeito. Reportamos dois exemplos deste tipo:

(35) // *porque na verdade / foi o que eu falei / **com cê*** //

(bfamcv23-3) 


(36) // [47] ***cê** / mandava [1] &re [1] correspondia por carta / <assim / né>* //

(bfammn02-47) 

No exemplo (35), *com cê* encontra-se isolado em uma unidade tonal, mas ao mesmo tempo tem-se o caso de pronome precedido por preposição. No exemplo (36), *cê* está isolado em unidade tonal, mas encontra-se também em posição pré-verbal como sujeito de verbo. Este último caso é para nós de especial importância porque, após a análise prosódica, pudemos perceber se tratar de um tópico, o que comprovaria a necessidade de se analisar no nível suprasegmental, já que a forma *cê* revelou-se longa.


Em relação aos casos de isolamento absoluto, há somente uma ocorrência. Esta também combina a situação de isolamento com o fato de ser objeto de preposição. O exemplo é o seguinte:

(37) // [90] *acho que ea fica melhor do que da outra* // [91] **p' cê** //

(bpubdl01-91) 

Para finalizar a apresentação dos resultados de *cê(s)*, citamos o único exemplo da forma como núcleo sintático do enunciado que foi encontrado no *corpus*. Relembramos que, além deste caso com a forma *cê*, há somente outros dois com a forma *ocê*, sendo, portanto, uma situação bastante rara.

(38) *AVI: [38] *ái o cara falou assim / não* // [39] *eu nã posso não* // [40] *de jeito nenhum* // [41] *nã posso dormir aqui hoje não / sô* // [42] *uai / mas por quê* // [43] *nã sei o que que tem / tal* // [44] *falou / de jeito nenhum* // [45] *cê tá ficando doido* // [46] ***cê casadinho de novo / e tal / a mulher bonita desse jeito aí / e tal / uma caminha de meio casal*** // [47] *como é que eu vou dormir aqui / meu Deus* //

(bfamdl19-46) 

4.1.6. Resultados da análise acústica

A metodologia adotada para a análise acústica já foi detalhada na seção 3.1.2. Neste parágrafo, indicaremos os resultados obtidos e o que se pode sugerir a partir deles. Foram submetidas à análise acústica 133 formas diferentes e seus relativos contextos. Alguns enunciados apresentaram mais de uma ocorrência da mesma forma ou de formas diferentes e de interesse do trabalho.

A tabela abaixo apresenta os pronomes analisados acusticamente:

Forma	Ocorrências
CÊ PRÉ-VERBAL	37
CÊ PÓS-VERBAL	35
OCÊ PRÉ-VERBAL	8
OCÊ PÓS-VERBAL	16
VOCÊ PRÉ-VERBAL	24
VOCÊ PÓS-VERBAL	13
TOTAL	133

Tabela 10: ocorrências das formas pronominais de segunda pessoa singular submetidas a análise acústica.

A análise versou sobre 133 pronomes, todos no singular, a saber 37 da forma *cê* em posição pré-verbal e 35 em posição pós-verbal, oito da forma *ocê* em posição pré-verbal e 16 em posição pós-verbal; 24 da forma *você* em posição pré-verbal e 13 em posição pós-verbal. As ocorrências da forma *cê* em posição pós-verbal foram analisadas em praticamente sua totalidade, faltando três exemplos que não puderam ser estudados devido à baixa qualidade acústica e excesso de sobreposição de fala, que deixaria ampla margem de dúvida à pesquisadora.

Lembramos que, justamente durante esta etapa foram descobertos dois casos da forma *cê* que haviam sido transcritos como *ocê*.

Foram varridos 62 textos diferentes: 52 pertencentes ao contexto familiar, dos quais 22 conversações, 17 diálogos e 12 monólogos. Outros nove textos fazem parte do contexto público: três conversações, cinco diálogos e um monólogo.

As 133 formas foram pronunciadas por 74 falantes: 36 homens e 38 mulheres.

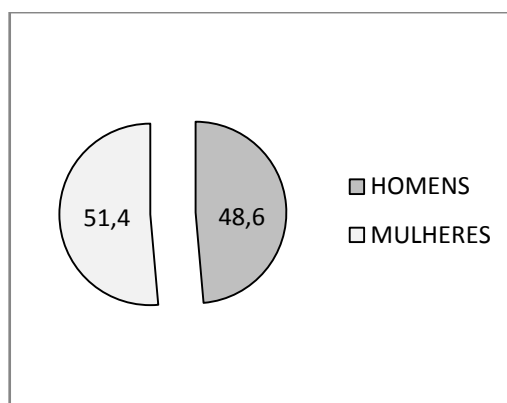


Gráfico 21: Dados dos informantes: gênero em termos percentuais.

Foram verificadas as características sociolinguísticas dos informantes: entre os homens sete pertencem à faixa A (entre 18 e 25 anos), 11 à B (entre 26 e 40 anos), nove à C (entre 41 e 60 anos), seis à D (maior de 60 anos). Um informante é menor de idade e dois são de idade desconhecida. Entre as mulheres 12 pertencem à faixa A, 11 à B, 13 à C, uma à D e de outra é desconhecida a idade.

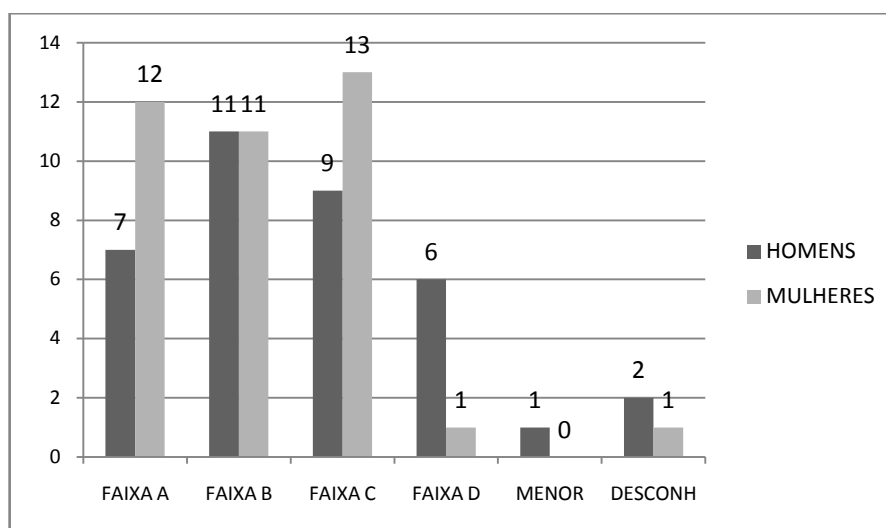


Gráfico 22: Dados dos informantes: idade em termos absolutos.

Quanto ao nível escolar, entre os homens seis possuem ausência de escolarização ou primário incompleto (faixa 1); 10 pertencem à faixa 2, que inclui uma escolarização até o terceiro grau completo mas cuja profissão não requeira o ensino superior; 12 pertencem à faixa 3, com título de terceiro grau, e cuja profissão o exija. De três falantes não conhecemos a escolaridade. As mulheres apresentaram uma escolarização mais alta: seis pertencem à faixa 1, 22 à faixa 2 e nove à faixa 3. De uma delas não se conhece a escolaridade.

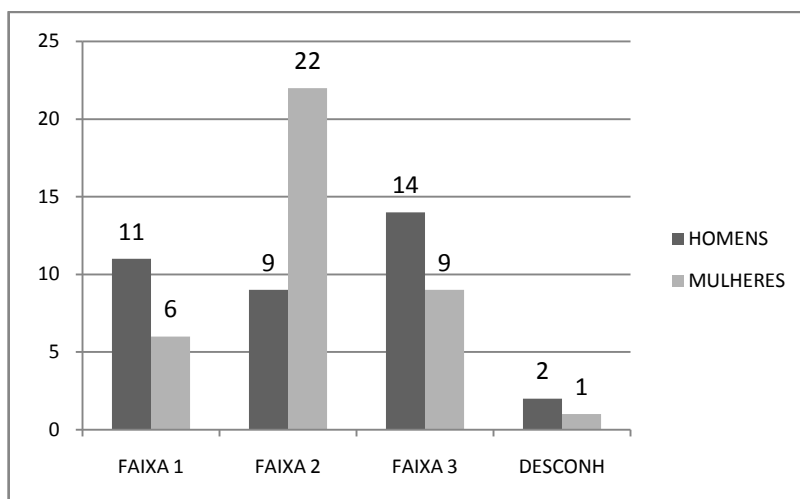


Gráfico 23: Dados dos informantes: escolaridade em termos absolutos.

Em relação à proveniência, 13 homens são da cidade de Belo Horizonte, nove da Grande Belo Horizonte, dez do interior de Minas Gerais (das cidades de: Ipanema, Ponte Nova, Itaúna, Juiz de Fora, Coluna, Santa Bárbara, Três Corações, Lavras e Presidente Bernardo) e um da cidade de São Paulo. De três não é conhecida a proveniência. As mulheres são em sua maioria da capital: 20 são da cidade de Belo Horizonte, quatro da Grande Belo Horizonte, nove do interior do estado (das cidades de Brasília de Minas, Ponte Nova, Alpercata e Itamarandiba) e quatro de outros estados: de Vitória (Espírito Santo), Rio de Janeiro, São Miguel do Oeste (Santa Catarina) e Santo Expedito (São Paulo). De uma mulher não se possuem dados sobre a origem.

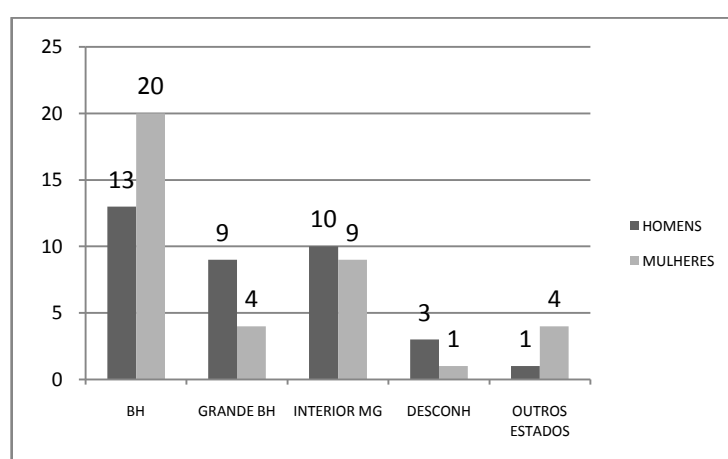


Gráfico 24: Dados dos informantes: origem em termos absolutos.

Como já foi dito anteriormente os textos foram escolhidos em primeiro lugar entre aqueles que apresentam melhor qualidade acústica. Contudo, especialmente no caso das formas *cê(s)* pós-verbais, foi necessário recorrer a textos com qualidade acústica inferior, por

se tratar de formas pouco frequentes. Tentamos evitar ao máximo textos com sobreposição de fala, para minimizar a margem de dúvida, contudo alguns destes foram inseridos entre aqueles submetidos à análise acústica por apresentar casos especialmente interessantes e raros. A fala espontânea, com a presença de ruídos de fundo, dificultou o processo de segmentação dos enunciados, que foi mais demorado do que o previsto, limitando assim o número total de formas analisadas e a opção por limitarmos-nos àquelas no singular.

A análise fonética procedeu da seguinte forma:

- foram extraídos do *corpus* alinhado ao áudio os enunciados que contêm as formas analisadas e seus enunciados limítrofes através do programa *WinPitch*;
- em seguida, o áudio e a transcrição da fala foram inseridas no programa *Praat* para a segmentação;
- foi efetuada uma segmentação fonética larga em ASCII das sílabas fonéticas de cada enunciado;
- extraíram-se os valores das durações brutas dos segmentos;
- todos os valores foram normalizados com o auxílio do *script SGdetector*;
- Os valores de *z-score*, ou seja, o desvio da média em unidade de desvio-padrão, da sílaba *cê* de cada forma foram comparados com a média das sílabas do enunciado a que pertenciam.
- Também para as formas *ocê* e *você* foi analisada somente a segunda sílaba, possibilitando comparar entre si todas as formas.

Esta primeira etapa forneceu-nos uma série de dados que fortaleceram a hipótese de que haveria uma forte correlação entre duração das formas e suas funções e não uma correlação entre forma segmental e função. A dizer, os segmentos curtos equivaleriam à posição pré-verbal e à função de sujeito, enquanto os segmentos longos equivaleriam à posição pós-verbal e à função de complemento, de verbo ou de preposição, além de sujeito posposto. A explicação seria que antes do verbo o segmento *cê* é átono, enquanto após o verbo carrega acento.

Mas por que falamos em sílaba *cê* se estamos analisando as três formas da segunda pessoa? É porque a única maneira de compará-las entre si seria justamente comparando a sílaba fonética [‘se]. Para termos maior certeza do peso desta sílaba, foram averiguados todos os valores da duração bruta das duas partes do segmento: [s] e [e]. Queríamos ter certeza de que não fosse a vogal [e] que determinaria a maior ou menor duração da sílaba. Segundo nossa análise, não há um peso substancial da vogal, mas parece que o fator determinante seja a duração da consoante [s].

Após esta etapa percebemos que havia valores não previstos. Percebemos que não seria possível continuar com nossa análise sem levar em conta a natureza informacional do segmento.

4.1.6.1. Função sintática e função informacional

O objetivo deste trabalho é estabelecer se nos pronomes *cê*, *ocê* e *ocê* a duração é um parâmetro mais confiável para marcar a função sintática do que a forma segmental. A hipótese é que as restrições para uma determinada forma segmental associadas a uma determinada função sintática não se confirmam nos dados de fala espontânea, ou se reduzem fortemente. Certamente, a forma *cê* é usada também em função de objeto de preposição; exemplos isolados nos deixam pensar que ela possa ser usada, mesmo com baixa frequência, também como objeto pós-verbal em frase clivada, como objeto direto e como tópico. Isso indica que um *corpus* maior do que o C-ORAL-BRASIL poderia apresentar um número de dados maior para essas funções. A função seria, portanto, marcada por características prosódicas e não pela forma segmental, apesar de a forma bissilábica ser mais frequente em função de objeto ou de sujeito posposto. Uma confirmação indireta da nossa hipótese é que também as formas bissilábicas mudam a sua duração em dependência da sua função sintática. Em síntese, o PB permitiria o uso de mais formas segmentais nas mesmas posições e funções sintáticas (embora com probabilidades diferentes), mas a duração seria sempre associada à função: breve em função de sujeito pré-verbal e longa em função de sujeito pós-verbal e objeto.

Todavia, para analisar a relação entre duração e função sintática, é necessário eliminar os fatores informacionais que podem condicionar a duração das formas. De fato, a função informacional é de ordem superior àquela sintática, e a duração é um elemento fundamental para marcar a função informacional. Somente dentro do mesmo domínio informacional podemos avaliar as relações sintáticas. Essa posição, que é assumida como base deste trabalho (CRESTI, 2000; RASO, 2012; MONEGLIA & RASO, 2014; CRESTI, 2014), nos levou a tomar algumas decisões metodológicas explicadas a seguir.

Em primeiro lugar, existe uma fortíssima tendência à isomorfia entre unidade informacional e unidade tonal. Por essa razão, o domínio mais apropriado de comparação entre a duração silábica é a unidade tonal e não o enunciado. De fato, o conteúdo locutivo de uma unidade tonal sempre faz parte da mesma unidade informacional. Em uma quantidade relativamente pequena de casos pode acontecer que uma mesma unidade informacional seja escansionada em duas (raramente mais) unidades tonais, mas nunca acontece que uma mesma

unidade tonal cubra mais de uma função informacional. Por essa razão, junto com a comparação da duração silábica do *cê* com as sílabas de todo o enunciado, colocamos também a comparação da duração silábica apenas com as sílabas da mesma unidade tonal na qual aparece o pronome objeto de estudo. A nosso ver, os valores realmente significativos para a comparação são esses últimos.

Em segundo lugar, a avaliação da função informacional nos leva a fazer considerações específicas em alguns casos em que o próprio pronome é marcado informacionalmente e em outro em que o pronome ocorre dentro da unidade informacional de tópico. Analisaremos esses casos um a um. Trata-se das situações seguintes:

- seis casos em que o pronome possui função ilocucionária, dos quais duas com *cê* objeto de preposição, três com *ocê* e uma com *você*;
- um caso em que o *cê* e a preposição que o rege formam um parentético, sem nenhum outro elemento dentro da unidade tonal;
- cinco casos de unidades tonais escansionadas constituídas apenas pelo pronome, das quais quatro com *cê* e uma com *você*;
- um caso em que o pronome *cê* aparece isolado em unidade tonal e com aparente função de tópico; na realidade, não temos dúvida sobre a função de tópico do pronome, mas, já que esse seria um caso contrário ao que normalmente é afirmado na literatura e o exemplo é único, nos parece correto não asseverar isto com certeza. Em alternativa, esse também seria um caso de escansão como os anteriores;
- três (talvez quatro) pronomes motivados pragmaticamente, um com *ocê*, dois (ou três) com *você*.

4.1.6.1.1. O valor ilocucionário

Para a avaliação das funções informacionais nos baseamos na *Language into Act Theory* (L-AcT) (CRESTI, 2000; MONEGLIA 2005 e 2011; RASO 2012; MONEGLIA & RASO, 2014). O *corpus* C-ORAL-BRASIL (como também os *corpora* LABLITA, C-ORAL-ROM e *Stammerjohann* nos quais foram estudados os pronomes italianos) são segmentados com base nos critérios da L-AcT.

A ilocução é definida como o núcleo informacional do enunciado, que, por sua vez, é definido como a menor unidade autônoma pragmaticamente (e prosodicamente). Isso significa que o enunciado constitui a unidade de referência da fala, a unidade mínima que pode ser

interpretada em isolamento. Por sua vez, isso quer dizer que esse é o domínio dos fenômenos linguísticos de base. Para que isso seja possível, é necessário que o enunciado possua uma unidade com função ilocucionária. A unidade ilocucionária é, portanto, necessária e suficiente para a realização de um enunciado, já que é ela que transmite o valor comunicativo acional e é prosodicamente interpretável. É a unidade ilocucionária que confere interpretabilidade ao enunciado como um todo. Podemos, portanto, ter enunciados formados apenas pela unidade ilocucionária e enunciados formados pela unidade ilocucionária e outra(s) unidade(s) com diferente(s) valor(es) informacional(is).

Antes de examinar os exemplos, precisamos também dizer que a unidade de referência da fala pode se configurar sob forma de enunciado, com uma única unidade ilocucionária (COM - *Comment*) ou duas unidades ilocucionárias padronizadas entre elas (CMM-CMM - *Multiple Comment*), mas pode também ter a estruturação chamada *Stanza* (CRESTI, 2010; RASO 2012; MONEGLIA & RASO, 2014). Nesses casos, o enunciado se dilata e cada subpadrão termina com um sinal prosódico de continuidade que indica que a unidade de referência continua, estruturando-se em volta de mais unidades ilocucionárias processuais (ou seja, não padronizadas) da mesma classe: isso é frequente em textos monológicos, com sequências narrativas, expositivas ou argumentativas que se organizam em forma de *Stanza*. Nesses casos, as ilocuções que carregam o sinal de continuidade são etiquetadas como COB (*Bound Comment*).

Analisaremos agora um por um os exemplos em que os pronomes são de natureza ilocucionária:

(39) bfamcv18-167: *HER: [167] quer que eu quebro / **p' cê** //

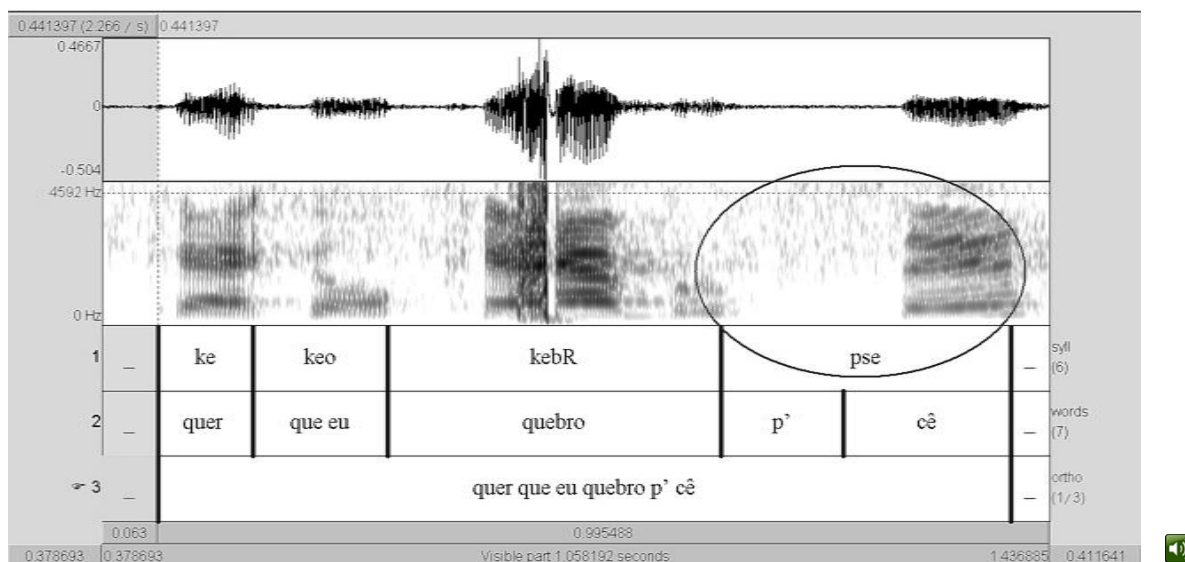


Figura 6: Tela do Praat com o texto (bfamcv18-187): é mostrado o espectrograma e nas camadas de anotação: o texto, sua segmentação em palavras e em código ASCII. É evidenciada a ilocução.

O caso acima poderia também ser segmentado como dois enunciados; a discussão teórica que leva a decidir entre uma ou outra segmentação não é influente para o nosso objetivo:

(40) bfamcv18-167: *HER: [167] quer que eu quebro // **p' cê** //

De fato, as duas unidades são ambas interpretáveis em isolamento e o sinal de continuidade entre a primeira e a segunda não é evidente. Isso, contudo, não mudaria a nossa análise. O que conta realmente é que *p' cê* constitui uma ilocução, ocupando a unidade tonal e, talvez, um enunciado inteiro. Não faria sentido, portanto, comparar a duração silábica da forma pronominal com a duração de algo que não está na mesma unidade e talvez nem no mesmo enunciado. A função ilocucionária produz valores de duração longos pela própria natureza, carregando um núcleo prosódico com função ilocucionária que é o que veicula a função acional e a interpretabilidade da unidade.

No enunciado acima, pode-se perceber que a duração do segmento *p'cê*, com *cê* preposicionado em posição pós-verbal é alongado, não somente pela posição em final de

enunciado ou pela função de objeto preposicional, mas principalmente pela função ilocucionária, que é o que prevalece.

Mesmo assim, a comparação da duração com as sílabas da primeira unidade leva aos seguintes resultados:

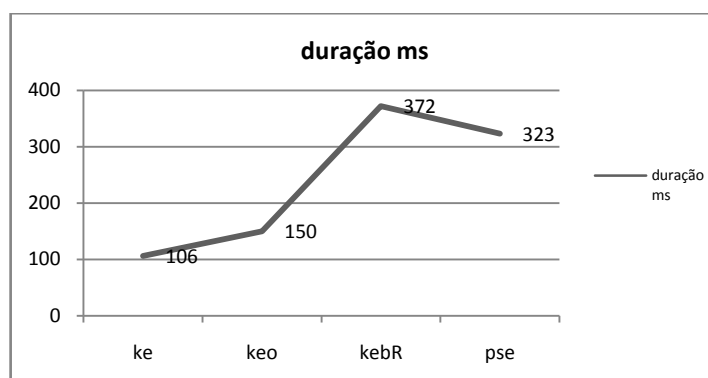


Gráfico 25: duração das sílabas fonéticas do texto (bfamcv18-187).

A sílaba final ['pse] é a segunda mais longa do enunciado, depois da sílaba final (e nuclear ilocucionariamente) da unidade/enunciado anterior.

(41) bpubd101-90-91: *PAU: [90] *acho que ea fica melhor do que da outra // [91] p' cê //*

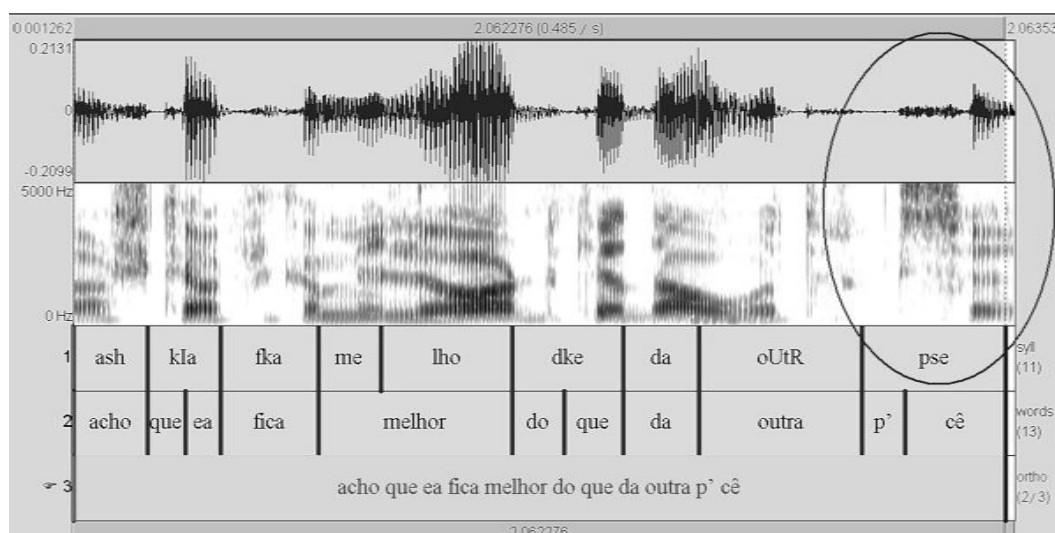


Figura 7: Tela do Praat com o texto (bpubd101-90-91): é evidenciado o último enunciado, ilocutivo.

Não muito diferente o caso seguinte, salvo o fato de que aqui não temos dúvida que o pronome constitui um enunciado à parte. De qualquer forma, mostramos sua duração

juntamente com o enunciado anterior para que o leitor tenha um parâmetro em relação à duração da sílaba que compõe a ilocução.

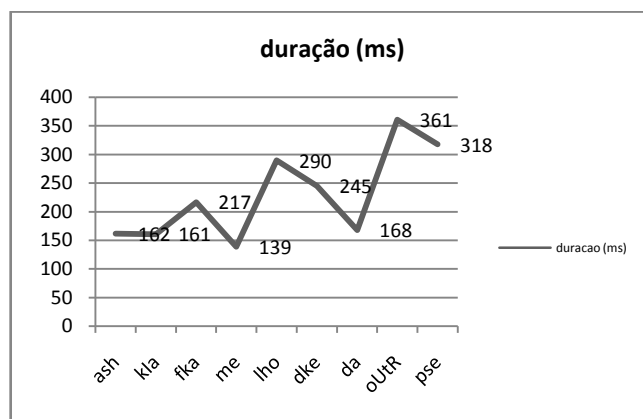


Gráfico 26: duração das sílabas fonéticas do texto (bpubd101-90-91).

O valor da sílaba fonética ['pse] é de 318 ms, muito longa, dado que aquela final do enunciado anterior é de 361 ms.

(42) bfamcv13-33: *ATA: [33] e *ocê* / *como é que arrumou isso aí* //

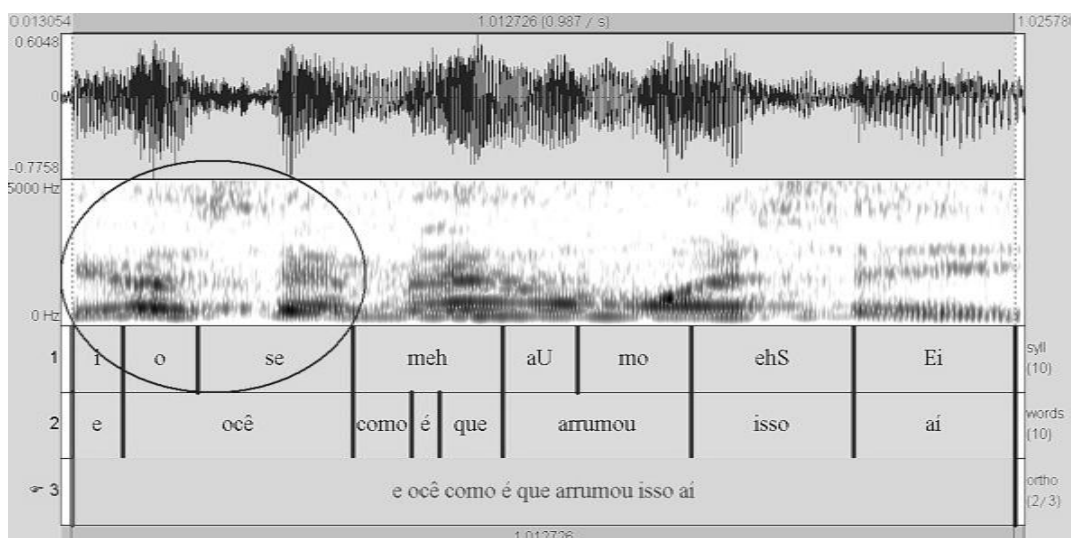


Figura 8: Tela do *Praat* com o texto (bfamcv13-33): é evidenciado o ilocutivo.

Neste caso, a ilocução encontra-se no início do enunciado. Sua duração bruta é de 167 ms, como pode ser deduzido pelo gráfico abaixo. Aqui também é provável que *e ocê* deva ser interpretado como COB, mas não há dúvida sobre a sua natureza ilocucionária.

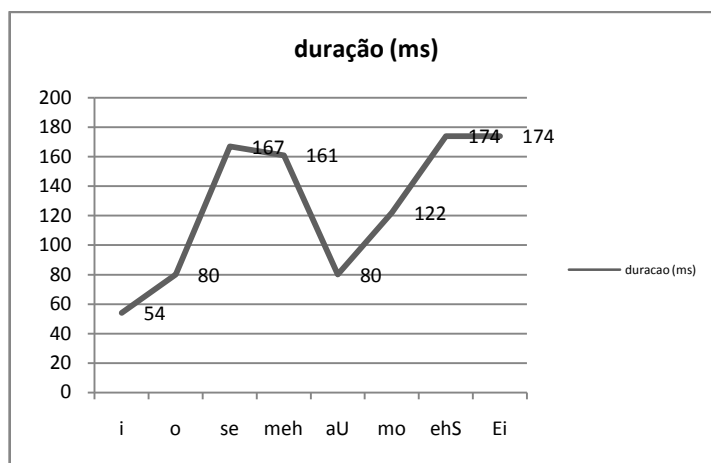


Gráfico 27: duração das sílabas fonéticas do texto (bfamcv13-33).

Os valores de duração confirmam o que se espera de uma unidade ilocucionária, ou seja, uma duração maior.

(43) bfamcv03-247: *TON: [247] *é ocê / Onofre / que tá cagando assim //*

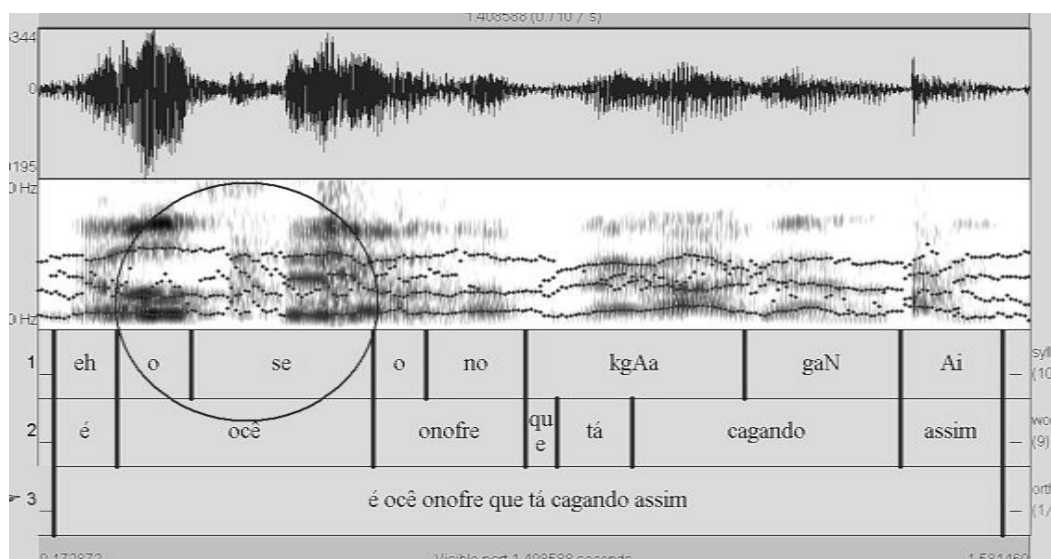


Figura 9: Tela do Praat com o texto (bfamcv03-247): é evidenciado o ilocutivo “é ocê”.

Discurso análogo para o exemplo acima.

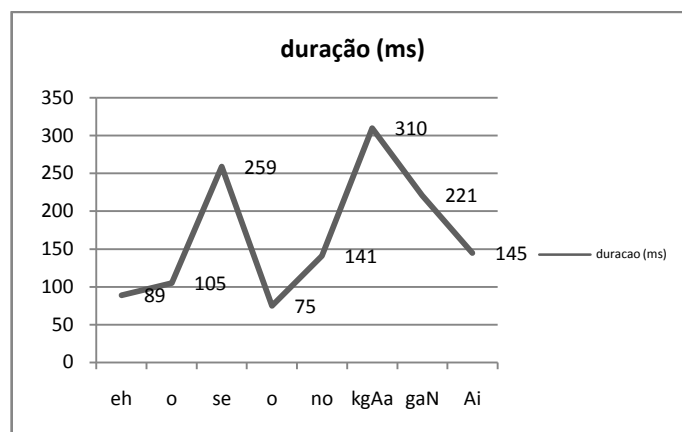


Gráfico 28: duração das sílabas fonéticas do texto (bfamcv03-247).

A duração bruta da sílaba ['se] é de 259 ms, a segunda mais longa do enunciado.

(44) bpubd111-241: *LUC: [241] *como é que foi esse projeto pr' ocê //*

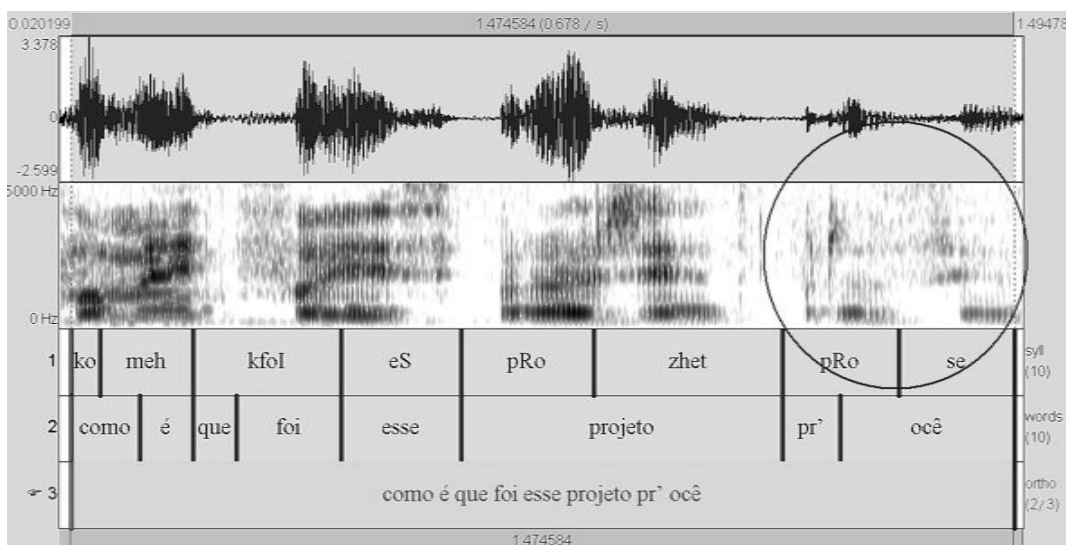


Figura 10: Tela do Praat com o texto (bpubd111-241): é evidenciado o ilocutivo “pr'ocê”.

A duração bruta da sílaba *cê* é de 180 ms, como é possível observar pelo gráfico.

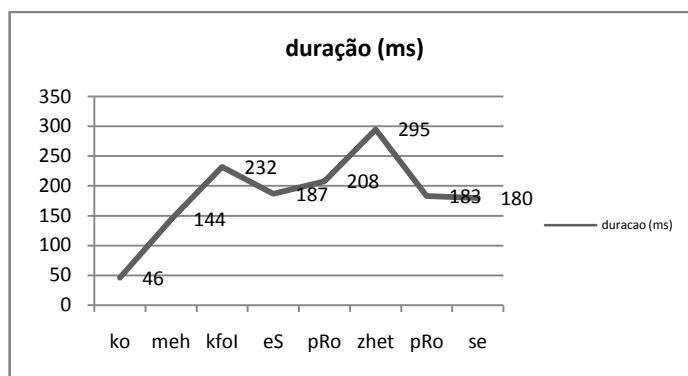


Gráfico 29: duração das sílabas fonéticas do texto (bpubdl11-241)

(45) bpubcv04-61: *WIL: [61] e *você* / *Hélida* / *veio só assessorar* //

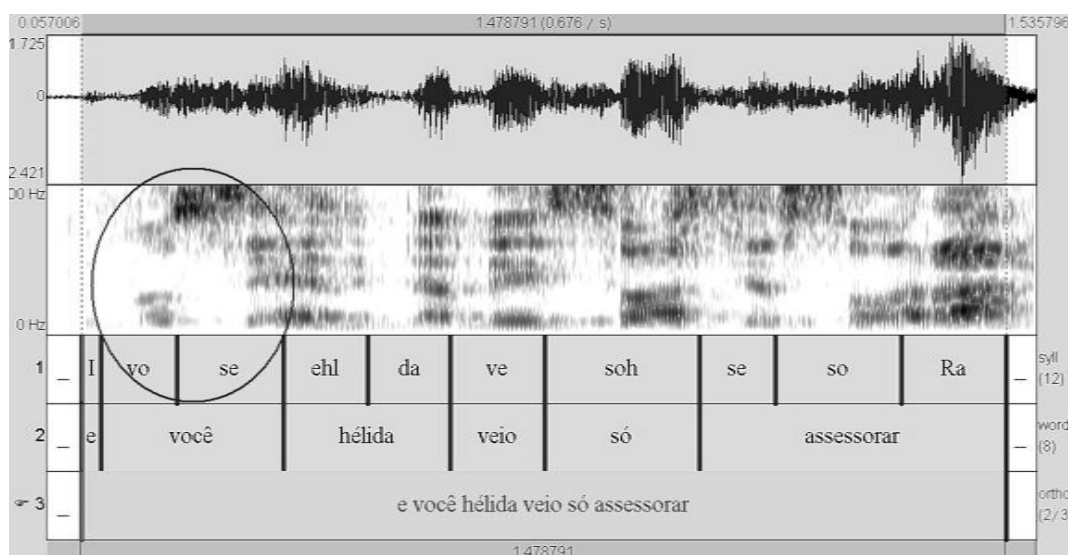


Figura 11: Tela do Praat com o texto (bpubcv04-61): é assinalada a forma *você*; o ilocutivo é “e você Hélida”.

A duração bruta da sílaba fonética [se] (no pronome *você*) neste caso é de 170 ms. Novamente estamos frente a um caso em que o pronome constitui por si só uma ilocução. Nesses casos, a posição não conta aos nossos fins, assim como não se pode falar de função sintática, já que o pronome constitui uma unidade autônoma do ponto de vista acional e não um elemento composicional com o verbo.

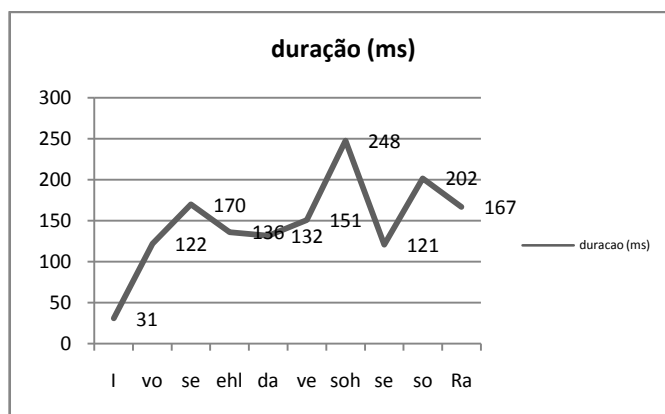


Gráfico 30: duração das sílabas fonéticas do texto (bpubcv04-61).

4.1.6.1.2. Parentético

Quando o enunciado é complexo, ou seja, além da ilocução, apresenta outras unidades tonais com valor informacional não ilocucionário, cada tipologia de unidade informacional apresenta parâmetros prosódicos que têm como função principal sinalizar a função informacional. Em outras palavras, em uma unidade complexa temos diversas unidades informacionais; em cada uma delas os parâmetros prosódicos possuem como função principal a de veicular a função específica daquela unidade informacional. Por isso, não faria sentido comparar os valores de duração silábica que veiculam uma função informacional com aqueles de outra unidade, que veiculam uma outra função informacional. Algumas unidades informacionais são veiculadas através de parâmetros que, além de outros, implicam o alongamento de certas sílabas; outras funções informacionais são veiculadas através de parâmetros que implicam, além de outros, o encurtamento das sílabas.

O caso do exemplo seguinte é muito significativo. Trata-se de uma unidade informacional com valor de parentético, uma função informacional conhecida e sobre a qual existe certo consenso na literatura (cf. TUCCI, 2003, 2009; VALE, 2009). Funcionalmente o parentético (PAR) fornece instruções ou integrações necessárias para a correta interpretação do enunciado ou parte dele. Essa função é veiculada através de uma redução geral dos valores prosódicos: abaixamento de f_0 (raramente aumento e sempre contraste com o resto do enunciado), redução de intensidade e de duração, além de (frequentemente, mas não sempre) um aumento da taxa de elocução. No exemplo abaixo, *com cê* constitui o conteúdo locutivo do parentético. A redução da duração silábica é, portanto, motivada informacionalmente, mesmo porque não faz sentido (em nossa perspectiva teórica) assumir composicionalidade sintática entre elementos separados por quebra, ou seja, entre elementos pertencentes a diferentes unidades informacionais. A quebra estabelece exatamente a fronteira entre o que é

composicional sintática e semanticamente (o que faz parte da mesma unidade) e o que é em relação informacional, veiculada por formas prosódicas. Dentro dessa visão, o fato de *com cê* ter duração silábica mais curta daquela das outras unidades do mesmo enunciado é exatamente o esperado.

(46) bfamcv13-261: *ATA: [261] <Mané / o' / eu fiquei> feliz de encontrar [/2] cê encontrou / dois gente boa aí / <com cê / no quarto> / batendo papo //

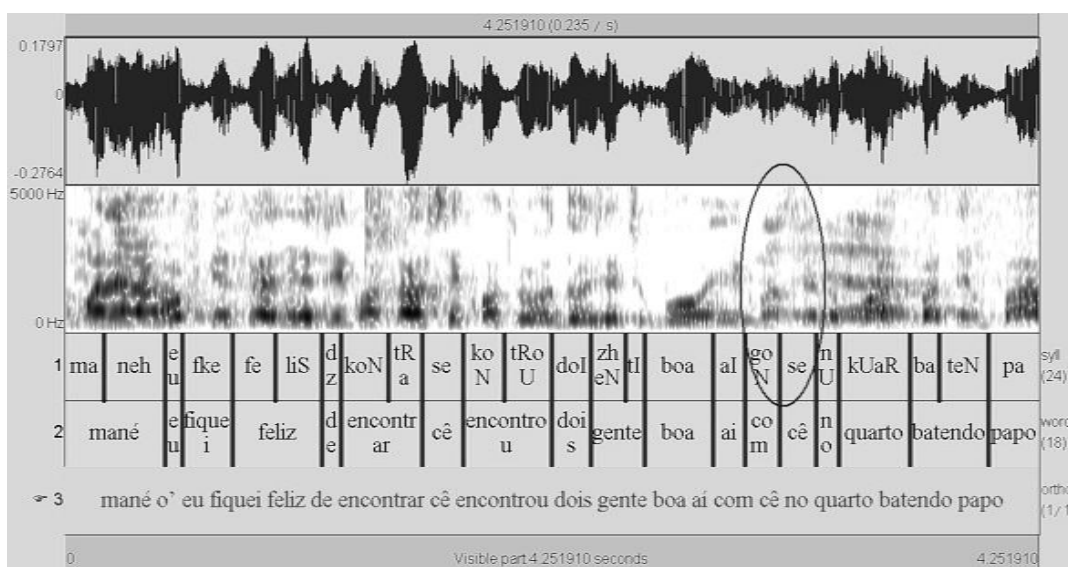


Figura 12: Tela do *Praat* com o texto (bfamcv13-261): é evidenciado o parentético “com cê”.

Na figura seguinte eliminamos, através de uma edição de som, a parte do parentético, justamente para mostrar de oitiva como a eliminação de algo que se diferencia do andamento prosódico do resto em contraste com ele não gera nenhuma sensação de falta:



Figura 13: Tela do *Praat* com o texto (bfamecv13-261): note-se que foi retirado o parentético “com cê”.

4.1.6.1.3. As unidades escansionadas

Já dissemos acima que, às vezes, uma mesma unidade informacional é realizada por duas (ou raramente mais) unidades tonais. Nesse caso, o perfil da unidade tonal é neutro, e apenas a última unidade tonal da unidade informacional carrega os valores prosódicos que veiculam a sua função. A escansão de uma unidade informacional em mais unidades tonais pode se dar por diferentes razões. A mais comum é relacionada a uma imperícia, de base ou ocasional, do falante. De fato, as escansões são muito frequentes em falantes muito jovens ou de diastratia baixa (CRESTI, 2005), e são frequentemente acompanhadas por fenômenos de disfluência como *retractings* ou interrupções e mudanças de projeto. Outras vezes, principalmente em contexto formal, as escansões são devidas a programas informacionais demasiadamente longos do ponto de vista silábico, de modo que a divisão do programa informacional em mais unidades tonais é devida a razões articulatórias (em média, uma unidade tonal não supera as 6-7 sílabas fonéticas). Por fim, e em poucos casos, a escansão pode ser motivada por razões enfáticas: o falante quer dividir o programa informacional em mais unidades para marcar enfaticamente seu conteúdo locutivo. Isso aparentemente acontece só nas unidades ilocutivas.

Nos exemplos a seguir aparece o pronome *cê* em uma unidade escansionada. É importante dizer que o fato de um elemento aparecer em uma unidade escansionada pode ter ou não efeitos sobre sua duração silábica: sendo a escansão com frequência ligada a fenômenos de disfluência, ela pode ser acompanhada por encurtamentos (devidos à não completa realização do conteúdo locutivo) como a alongamentos (devidos a hesitações). Portanto achamos oportuno, do ponto de vista metodológico, excluir esses casos da nossa

pesquisa. Contudo, os apresentamos para que o leitor possa entender melhor a nossa escolha e fazer suas avaliações.

(47) bfamcv31-187: *LUC: [...] [187] o ruim / é / tipo / *cê* / acertar direitinho / porque / a / <precisão dele é> muito ruim //

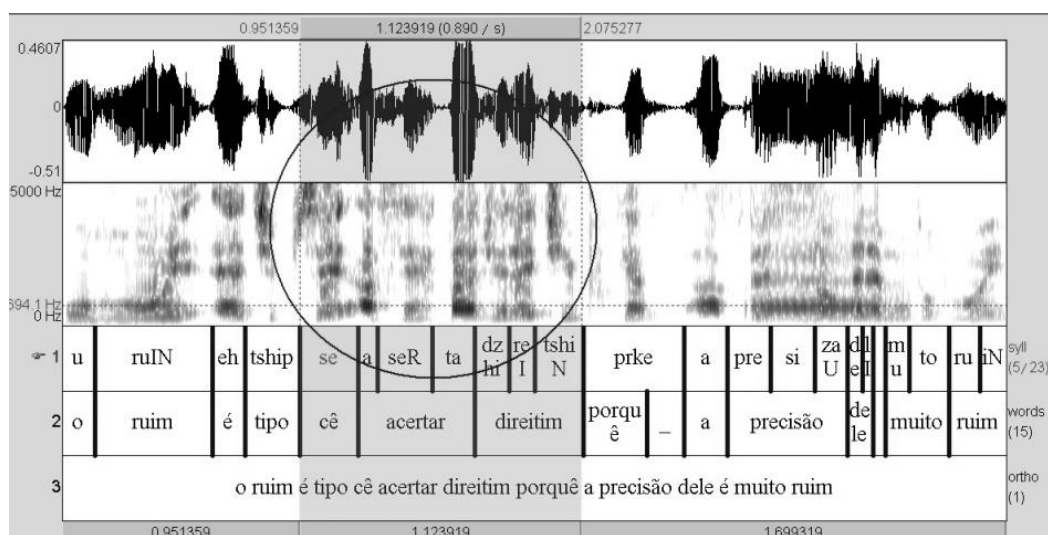


Figura 14: Tela do Praat com o texto (bfamcv31-187): é evidenciada a escansão “cê acertar direitim”.

A duração bruta da sílaba fonética ['se] neste caso é de 256 ms.

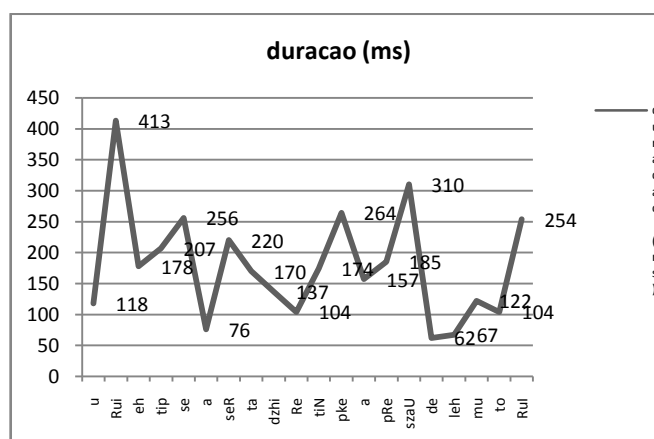


Gráfico 31: duração das sílabas fonéticas do texto (bfamcv31-187).

(48) bfamcv32-105: *HEL: [105] então / &he / &he / cê / faz aquilo que cê / deveria ter feito / agora / nesse / né //

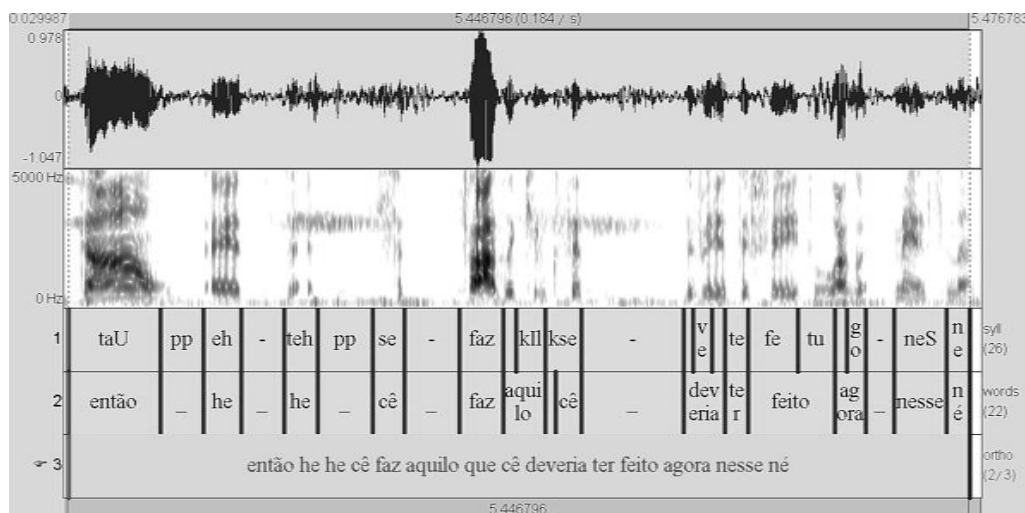


Figura 15: Tela do Praat com o texto (bfamcv32-105).

Este caso apresenta muitas hesitações, ocorrem dois *cê*, mas aquele escansionado é o segundo, cuja duração é de 220 ms, sendo a sílaba ['kse].

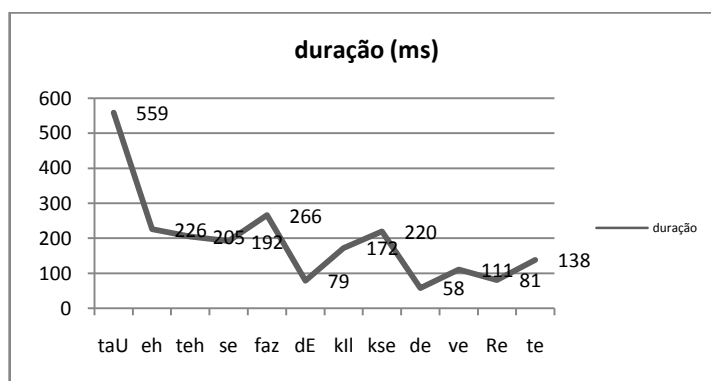


Gráfico 32: duração das sílabas fonéticas do texto (bfamcv32-105).

(49) bfammn27-64: *DOM: [64] *cê / higieniza / limpa / mata os funguinhos //*

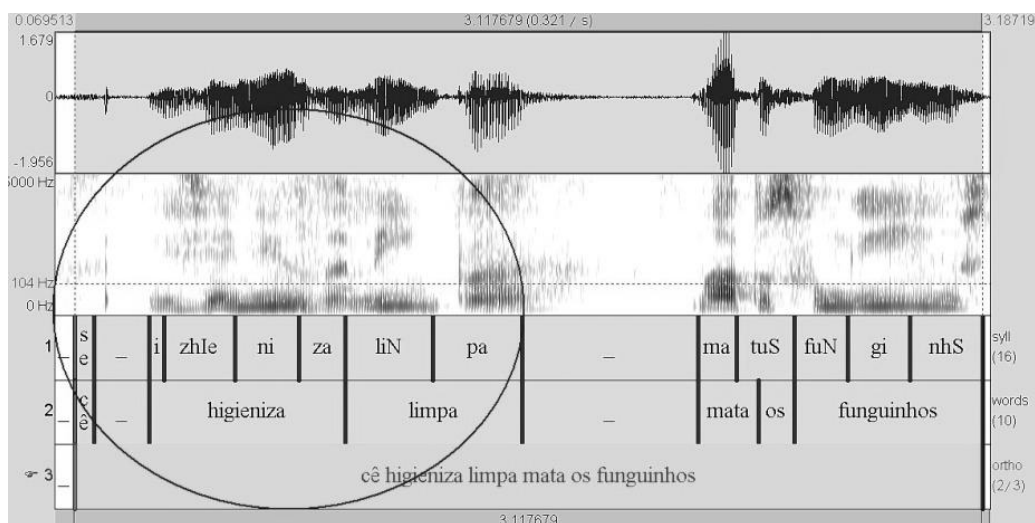


Figura 16: Tela do *Praat* com o texto (bfammn27-64).

Aqui a duração de *cê* é de 67 ms.

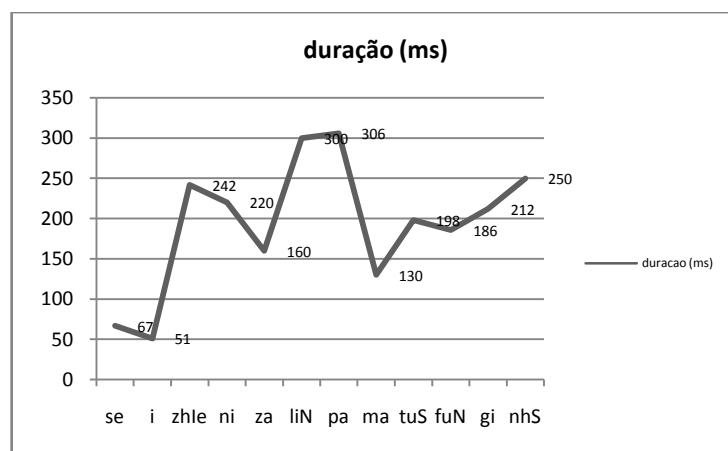


Gráfico 33: duração das sílabas fonéticas do texto (bfammn27-64).

(50) bpubmn11-24: *MMM: [24] *cê / falou que seria uma coisa diferente / né //*



Figura 17: Tela do *Praat* com o texto (bpubmn11-24).

A duração apresentada por *cê* é de 103 ms.

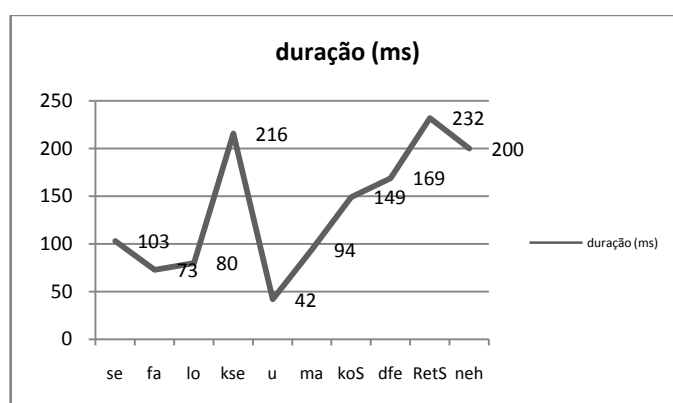


Gráfico 34: duração das sílabas fonéticas do texto (bpubmn11-24).

(51) bfamd105-185: *CES: [185] *você / é responsável pela obra //*

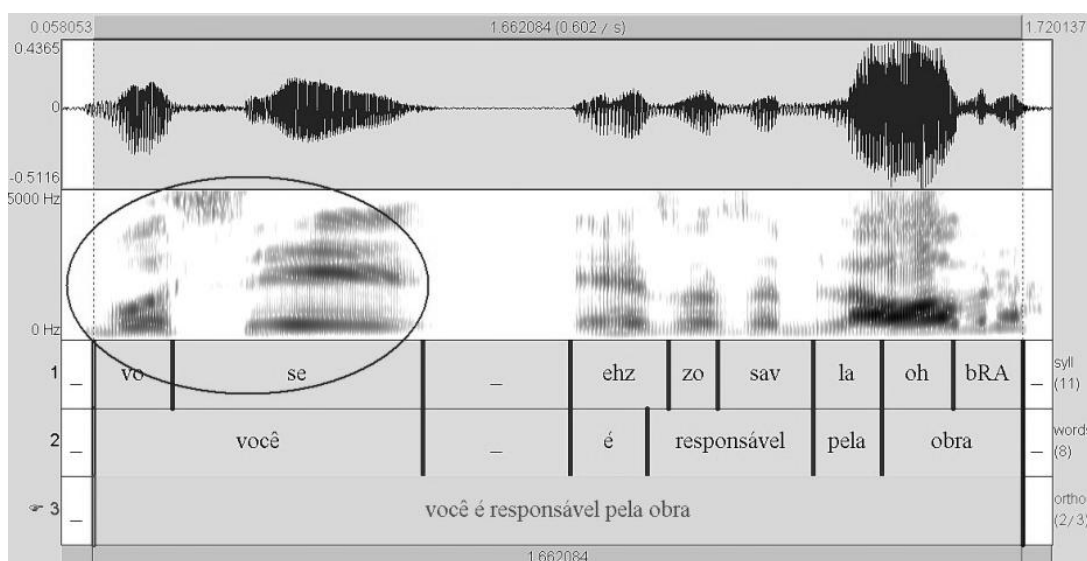


Figura 18: Tela do *Praat* com o texto (bfamd105-185).

Neste caso a duração da sílaba ['se] de *você* é muito longa: 448 ms.

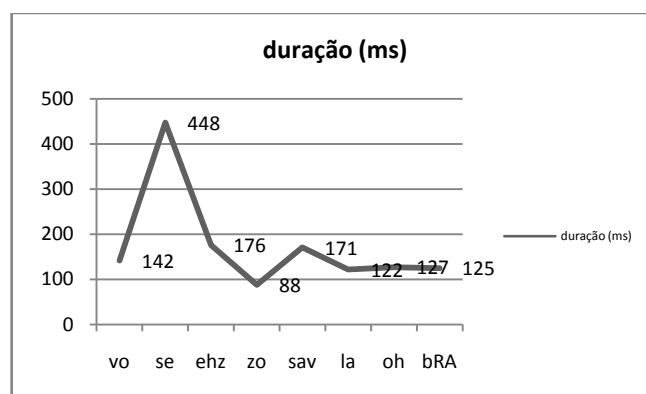


Gráfico 35: duração das sílabas fonéticas do texto (bfamd105-185).

4.1.6.1.4. Tópico

A unidade informacional de tópico, no arcabouço no qual nos baseamos, é definida como o âmbito de aplicação da força ilocucionária. O tópico, portanto, tem a função de definir o domínio cognitivo ao qual a força ilocucionária se refere. Do ponto de vista prosódico, o tópico é caracterizado por um núcleo funcional que veicula a função, sendo o resto dispensável do ponto de vista funcional e dedicado unicamente ao preenchimento locutivo da unidade. Esse núcleo funcional, nas formas de tópico conhecidas para o PB, pode ser de quatro tipos. Dois deles são posicionados à direita da unidade, a partir da última sílaba tônica. Os outros dois são divididos em dois seminúcleos, um posicionado à direita, sempre a partir

da última tônica, e o outro posicionado à esquerda (em um caso nas primeiras duas sílabas, em outro caso podendo ser precedido por uma preparação) (cf. MITTMANN, 2012). Em todas as formas, as partes nucleares são alongadas, sendo que o elemento à direita (seja o núcleo inteiro, seja o segundo seminúcleo) é sempre longo. Quanto ao nosso objetivo, isso tem como consequência que, no único caso em que o pronome *cê* funciona, na nossa opinião, como tópico, sua duração longa seria motivada pela sua natureza informacional. É o caso do exemplo seguinte:

(52) bfamnn02-47: “*LUC: [47] *cê* / mandava /l &re /l correspondia por carta / [...]/”

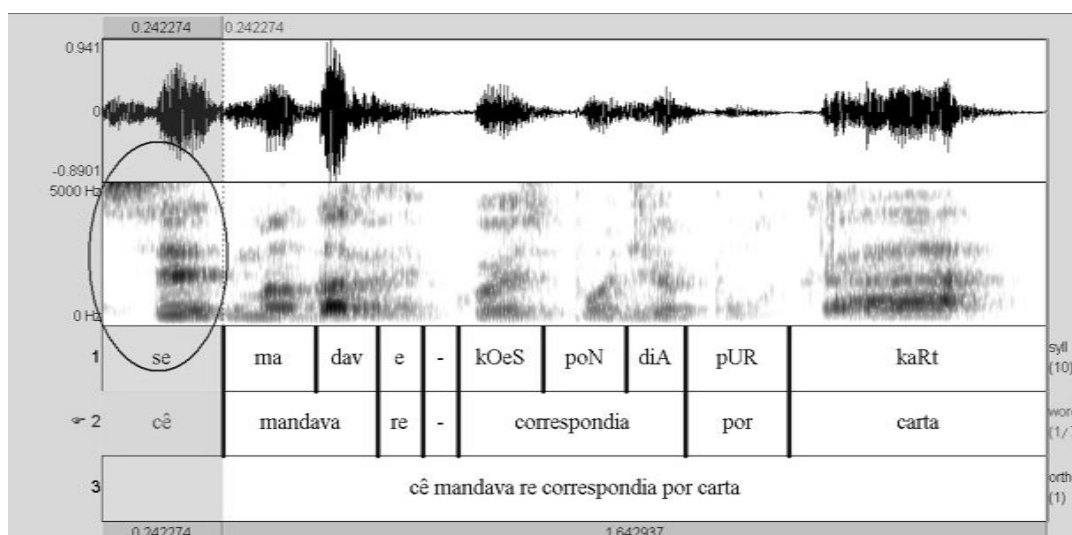


Figura 19: Tela do *Praat* com o texto (bfamnn02-47): é evidenciado o tópico, *cê*.

De fato, a duração de *cê*, de 243 ms é muito longa e, para o leitor ter uma ideia, é inferior somente àquela da sílaba final do enunciado, que é bem alongada.

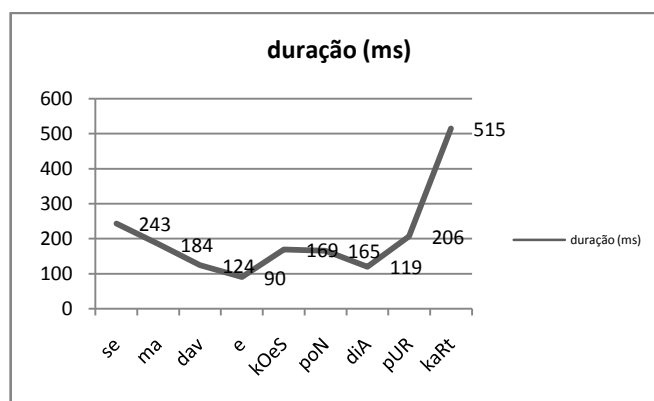


Gráfico 36: duração das sílabas fonéticas do texto (bfamnn02-47).

4.1.6.1.5. Motivação pragmática

Como é notório, a expressão do sujeito pode ter motivação sintática ou motivação pragmática. Nas línguas que são chamadas *não pro-drop*, usando a terminologia gerativista que é também utilizada por estudiosos que não necessariamente se reconhecem nessa vertente teórica, em princípio é necessária a expressão do sujeito, sendo portanto ela motivada sintaticamente. A falta da expressão do sujeito geraria formas agramaticais. Exemplos típicos de línguas *não pro-drop* são o inglês e o francês. Nas línguas chamadas de *pro-drop*, a expressão do sujeito é facultativa, o que seria relacionado com o fato de ela ser veiculada morfológicamente pela desinência verbal. Exemplos típicos de línguas *pro-drop* seriam o italiano (padrão), o PE e o espanhol. Nessas línguas, em princípio, a expressão do sujeito pronominal seria motivada pragmaticamente, já que a morfologia verbal seria suficiente para a individualização da pessoa.

Quanto ao PB, a forte redução das desinências verbais é vista como um elemento relacionado a uma mudança tipológica: o português, originariamente *pro-drop* (como ainda é o PE) estaria em processo de se transformar em uma língua *não pro-drop*, ou seja, em uma língua com sujeito tendencialmente obrigatório e motivado sintaticamente. Isso é frequentemente explicado com a necessidade de identificar o sujeito em ausência de uma morfologia verbal que cumpra essa função. De fato, na fala espontânea de registro coloquial, parece que o que acontece com maior frequência é uma oposição entre a forma de primeira pessoa (por ex. *amo*) em contraposição à forma que vale para todas as outras: (você/ele/nós-a gente/vocês/eles) ama. Isso geraria também uma redução fonética das formas sujeito, já que elas teriam apenas a função de indicar a pessoa verbal: *cê(s)/e'-ea-es-eas/aên*. Uma outra maneira de explicar o mesmo fenômeno, mas de fato equivalente para os nossos fins, é pensar que a marca morfológica que no PE (e nas línguas *pro-drop*) está à direita, teria passado à esquerda nas línguas *pro-drop*. Independentemente das diferentes orientações teóricas e históricas com as quais se explica esse fenômeno, o que nos interessa aqui é que em PB o sujeito é frequentemente motivado sintaticamente, mas pode também ser motivado pragmaticamente. Os exemplos que seguem são evidentemente motivados pragmaticamente. Em (51) o pronome *ocê* funcionalmente poderia ser parafraseado como: *você mesmo*, e é *mesmo você que*:

(53) bfamnn03-115: *ALO: [115] e *ocê* sabe quanto é que é //

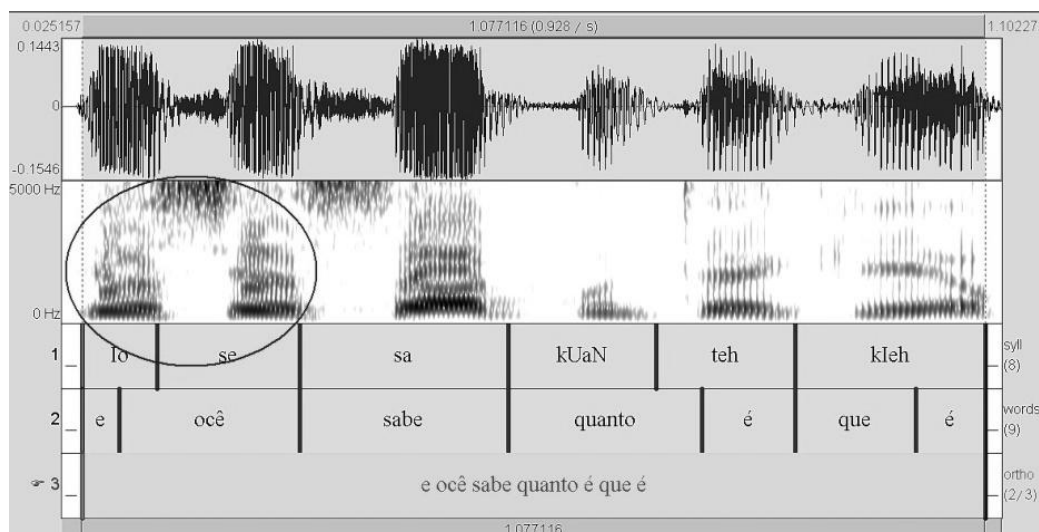


Figura 20: Tela do *Praat* com o texto (bfammn03-115).

Quanto às medidas de duração, a sílaba [se] apresenta uma duração de 170ms.

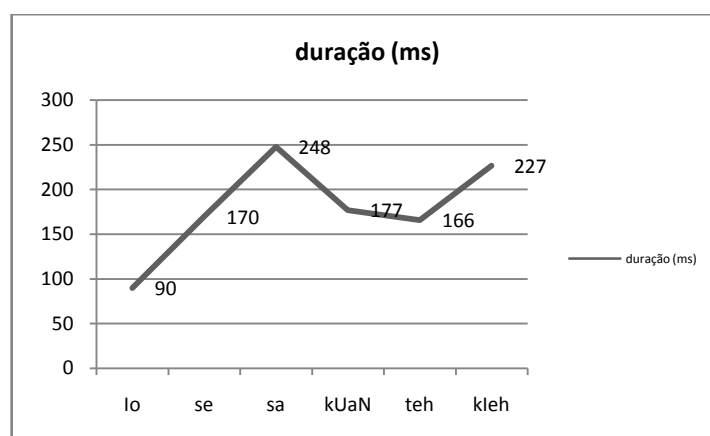


Gráfico 37: duração das sílabas fonéticas do texto (bfammn03-115).

Em (54) e (55), *ocê* poderia ser parafraseado como *ocê, não outra pessoa*.

(54) bfamcv19-27: *PAI: [27] *você que tá bolando aqueas cores //*

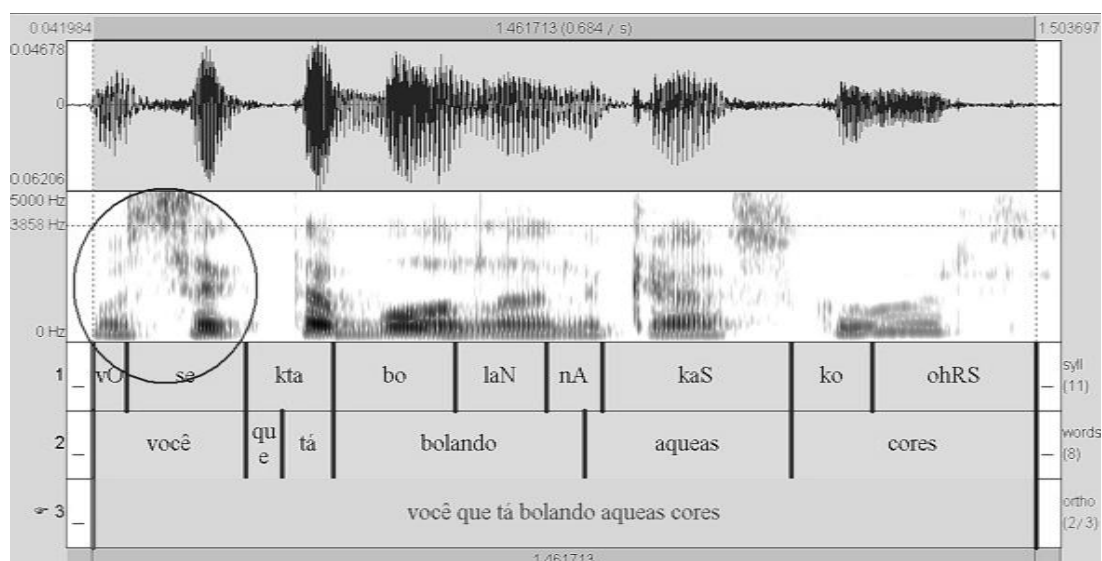


Figura 21: Tela do Praat com o texto (bfamcv19-27).

A duração da sílaba ['se] de *você* é de 171 ms.

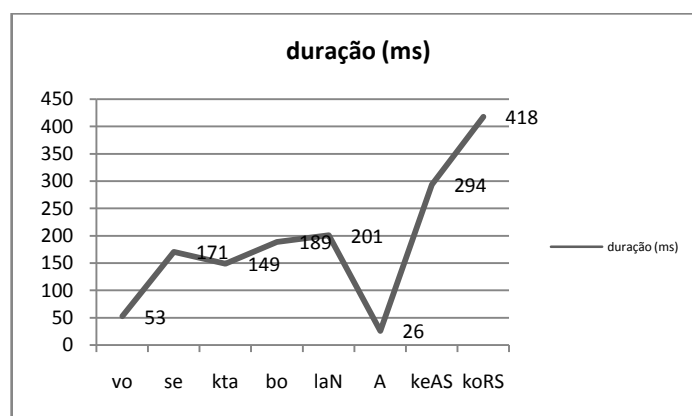


Gráfico 38: duração das sílabas fonéticas do texto (bfamcv19-27).

(55) bfamd105-282: *ANE: [282] *você que me mandou eu ver //*

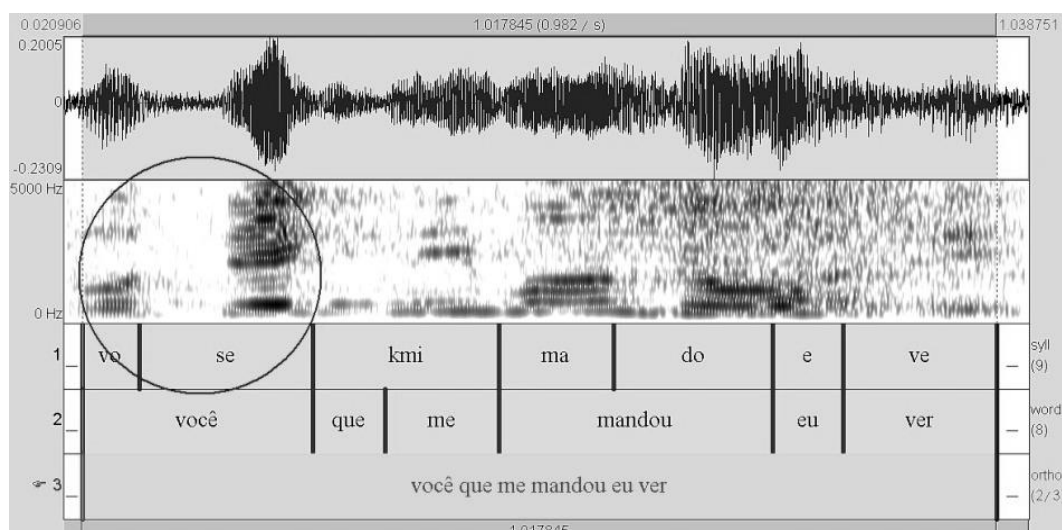


Figura 22: Tela do Praat com o texto (bfamd105-282).

A duração da sílaba ['se] de *você* é de 193 ms.

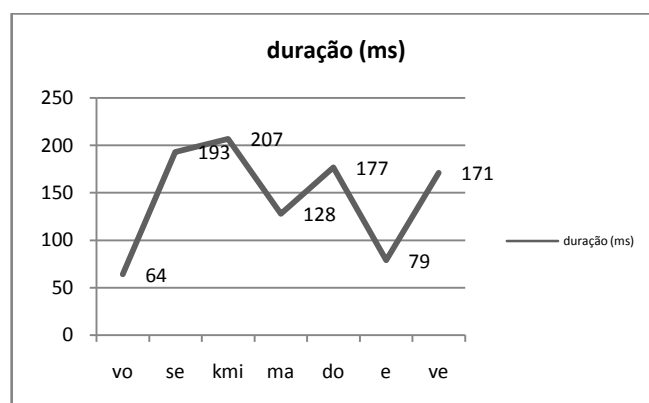


Gráfico 39: duração das sílabas fonéticas do texto (bfamd105-282).

No caso de (55), a oitiva não permite decidir se o pronome é motivado pragmaticamente ou não. Por isso, decidimos colocá-lo junto com a série motivada sintaticamente, mesmo a duração dele sendo longa, ou seja, constituindo um potencial contraexemplo para a nossa proposta, representada por (56) abaixo.

(56) bfamecv20-190: *CAI: [190]// *tipo assim / você botava / blusa sozinha //*

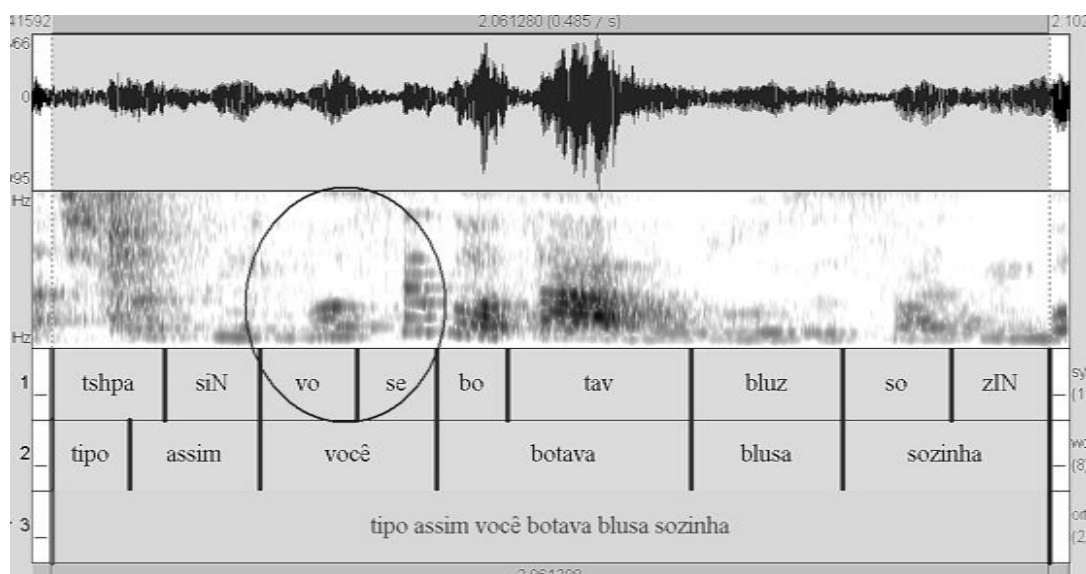


Figura 23: Tela do *Praat* com o texto (bfamecv20-190).

Nota-se, já pela comparação dos gráficos com os valores brutos da duração que, para ['se], é de 200 ms, como neste caso a sílaba em questão é uma das mais curtas.

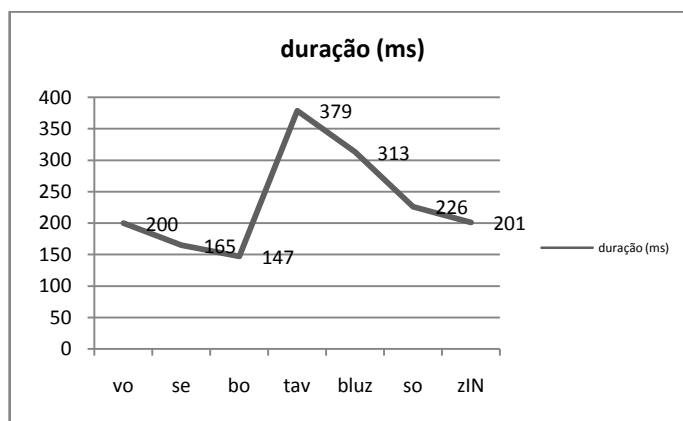


Gráfico 40: duração das sílabas fonéticas do texto (bfamecv20-190).

Naturalmente essas considerações poderiam nos levar a supor explicações à polimorfia dos pronomes de terceira pessoa singular e plural, mas isso foge do escopo do trabalho.

4.1.7. A marcação da função sintática das formas *cê*, *ocê* e *você*

Passemos agora à apresentação dos resultados daqueles dados que não possuem nenhum tipo de condicionamento informacional e que, portanto, são passíveis de uma análise acústica que meça quanto a duração da sílaba *cê* marca sua função sintática.

Excluindo-se os exemplos apresentados no parágrafo anterior, as análises desta etapa versaram sobre 117 casos: 64 da forma *cê*, dos quais 33 pré-verbais e 31 pós-verbais; 20 da forma *ocê*, sendo 6 pré-verbais e 14 pós-verbais; 33 da forma *você*, sendo 20 pré-verbais e 13 pós-verbais.

O gráfico abaixo mostra estas ocorrências em termos percentuais:

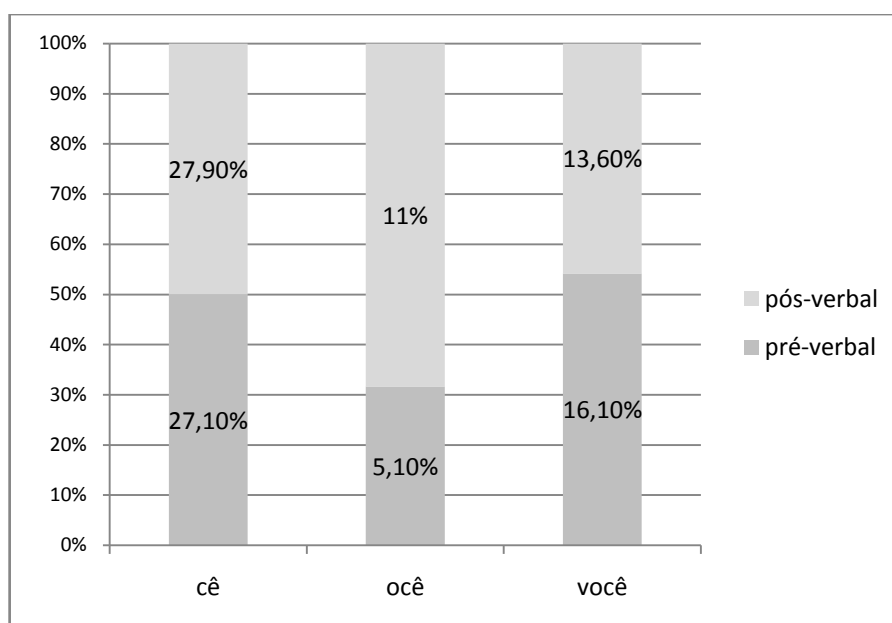


Gráfico 41: porcentagem dos segmentos analisados divididos por formas e posições.

Serão expostos e comentados os dados de cada forma de maneira separada em suas diferentes posições. Temos plena consciência do fato que a análise acústica não conseguiu atingir um número de exemplos ideais, em termos relativos e proporcionais, mas acreditamos que constitua uma base mais do que suficiente para testar e formular hipóteses.

4.1.7.1. Forma *cê* em posição pré-verbal

A forma *cê* é aquela que desperta maior interesse justamente por ser a mais debatida, como foi visto no capítulo da discussão bibliográfica.

Os dados aqui apresentados evidenciaram praticamente todos os mesmos resultados: em posição pré-verbal a sílaba *cê* exibiu resultados normalizados de *z-score* inferiores àqueles

do valor da média dos *z-scores* de todas as sílabas fonéticas do enunciado, e também da unidade tonal à qual o segmento pertence. Isto significa que *cê* ocorreu quase sempre¹³² como sílaba curta. Isto aponta para a atonicidade, confirmando nossa hipótese inicial.

A tabela abaixo expõe os resultados obtidos:

TEXTO	FUNÇÃO SINTÁTICA	DURAÇÃO Z SIL/ENUNCIADO		PROP ENUNC	DURAÇÃO Z SIL /UT		PROP UT
bfamcv01-128-EVN	SUJ	-2,18	-0,16	-1262	-2,18	0,33	-761
bfamcv02-335-RUT	SUJ	-3,44	-0,95	-262	-3,44	-1,28	-169
bfamcv03-35-CEL	SUJ	-3,83	-2,55	-50	-3,83	-2,55	-50
bfamcv04-184-BRU	SUJ	-2,89	-0,25	-1056	-2,89	-2,22	-30
bfamcv11-161-TIT	SUJ	-2,18	-0,88	-148	-2,18	-0,88	-148
bfamcv13-261-ATA-1	SUJ	-0,93	-0,89	-4,50	-0,93	-0,73	-27
bfamcv11-5-TIT	SUJ	-2,15	0,56	-485	-2,15	0,56	-484
bfamcv14-253-AJC	SUJ	-3,96	0,32	-1337	-3,96	-0,9	-340
bfamcv14-255-AJC	SUJ	-3,6	-0,84	-329	-3,6	-0,84	-329
bfamcv16-221-EDE	SUJ	0,85	-0,04	2225	0,85	1,12	-24
bfamcv18-130-HER	SUJ	-2,19	1,61	-236	-2,19	-0,39	-761
bfamcv18-208-HER	SUJ	-2,09	-1,21	-73	-2,09	-1,47	-42
bfamcv18-262-HER	SUJ	-3,69	-0,09	-4000	-3,69	-0,09	-4000
bfamcv20-187-CAI-1	SUJ	-1,54	-0,68	-126	-1,54	-0,84	-83
bfamcv20-187-CAI-2	SUJ	-3,16	-0,68	-365	-3,16	-0,43	-634
bfamcv21-43-FLA-2	SUJ	-3,17	-0,68	-366	-3,17	-0,45	-604
bfamdl02-45-BAL	SUJ	-0,54	0,75	-172	-0,54	-0,58	7
bfamdl02-61-62-BAL	SUJ	-2,28	2,17	-205	-2,28	2,12	-207
bfamdl04-191-SIL	SUJ	-3,43	-0,15	-2187	-3,43	-0,26	-1219
bfamdl04-51-SIL	SUJ	-2,23	-0,73	-205	-2,23	-0,73	-205
bfamdl10-127-HEL	SUJ	-1,46	-1,28	-14	-1,46	-1,28	-14
bfamdl10-198a201-HEL	SUJ	-1,3	-1,26	-3	-1,3	-0,62	-110
bfamdl10-70-HEL	SUJ	-3,18	-2	-59	-3,18	-2,76	-15
bfamdl26-161-MBA	SUJ	-1,11	-0,45	-147	-1,11	-0,45	-147
bfamdl26-284-LAO	SUJ	-3,23	-0,64	-405	-3,23	-0,64	-405
bfamdl28-165-ALV-1	SUJ	-5,41	-0,84	-545	-5,41	1,26	-529
bfamdl28-165-ALV-2	SUJ	-4,32	-0,84	-414	-4,32	-1,75	-147
bfamdl28-165-ALV-3	SUJ	-2,74	-0,84	-226	-2,74	-1,13	-142
bfammn12-144-JUN-2	SUJ	-1,14	-1,9	-40	-1,14	1,14	-200
bfammn35-182-JMA	SUJ	-3,45	-1,08	-219	-3,45	-1,8	-92
bfammn35-232-JMA	SUJ	-2,39	-1,15	-107	-2,39	-0,88	-172
bpubcv02-126-CAR	SUJ	-2,94	-0,56	-425	-2,94	-1,49	-97
bpubcv02-133-CAR	SUJ	-2,24	-2,01	-11	-2,24	-2,07	-8

Tabela 11: resultados das análises acústicas da forma *cê* em posição pré-verbal no *corpus* C-ORAL-BRASIL.

¹³² À exceção de um único exemplo que será discutido adiante.

Na primeira coluna, encontra-se o texto a que a sílaba pertence, com o número do enunciado e a sigla do informante. Se tal sigla é seguida de um número significa que no mesmo enunciado foi encontrada mais de uma forma pronominal, e o número se refere à ordem em que essa forma se encontra: se é a primeira ou a segunda a aparecer no enunciado e assim por diante. Na segunda coluna, é discriminada a função sintática que a forma exerce. Neste caso, o *cê* em posição pré-verbal coincidiu sempre com a função de sujeito. A terceira coluna é reservada ao valor da duração normalizada de *cê* em *z-score* (primeiro valor) comparado com o valor médio da normalização das sílabas fonéticas do enunciado a que a sílaba pertence (segundo valor). A terceira coluna indica a proporção percentual da diferença entre o valor de *cê* e aquele da média do enunciado. Na quarta coluna, tem-se o valor da duração normalizada de *cê* em *z-score* (primeiro valor) comparado com o valor médio da normalização das sílabas fonéticas da unidade tonal a que a sílaba pertence (segundo valor). Na quinta e última coluna, é discriminada a proporção percentual da diferença entre o valor de *cê* e aquele da média da unidade tonal a que ele pertence.

Mostraremos agora alguns exemplos das análises feitas.

(57) bfamcv02-335-RUT: //convite de casamento / *cê* pode mandar pa todo mundo //

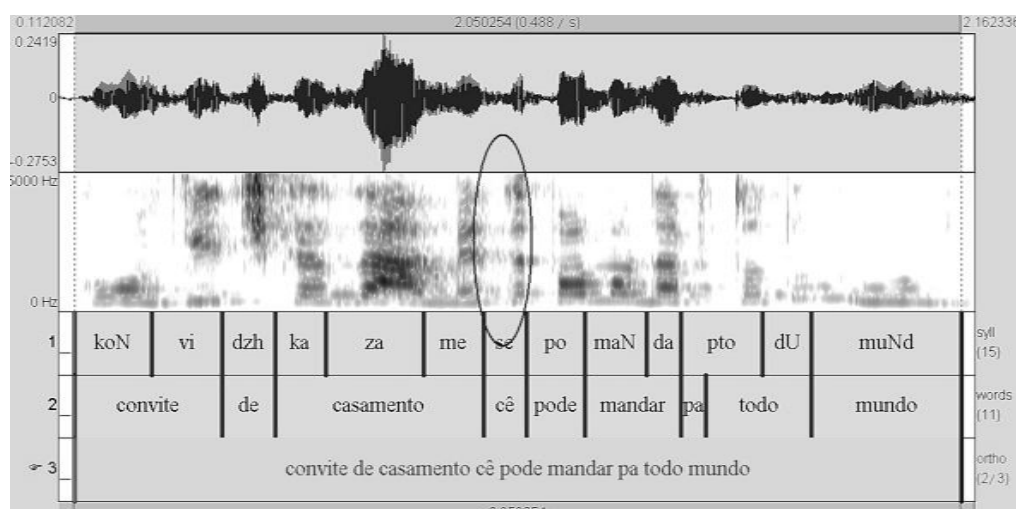


Figura 24: Tela do *Praat* com o texto (bfamcv02-335-RUT): é posta em evidência a forma *cê* pré-verbal.

O espectrograma mostra como foi feita a segmentação do enunciado. Os valores de duração extraídos foram aqueles do gráfico abaixo:

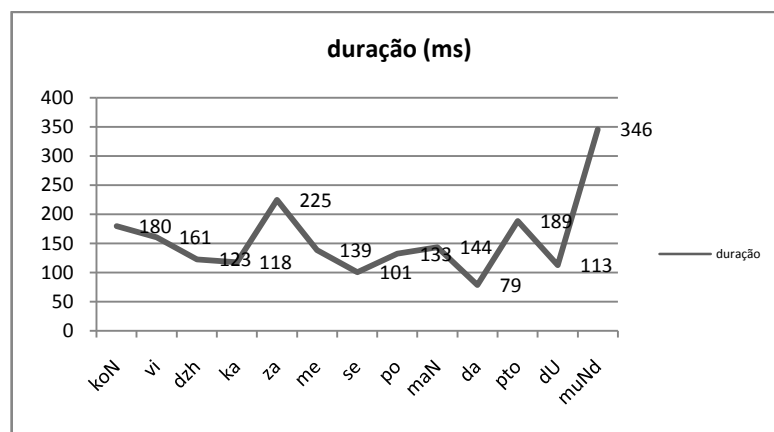


Gráfico 42: duração das sílabas fonéticas do texto (bfamcv02-335-RUT).

É evidenciado o andamento da duração bruta de cada sílaba fonética do enunciado. A duração da forma *cê* é de 101 ms. Os valores foram inseridos no *script SGDetector* (BARBOSA P. A, 2006) para normalização dos dados. Os resultados são os da tabela (2) acima: a sílaba *cê* apresentou uma duração normalizada em *z-score* de -3,44 de desvio padrão. Calculando-se a média dos valores de *z-scores* das sílabas fonéticas do enunciado, obteve-se -0,95 de desvio padrão. Portanto *cê* apresenta uma duração muito menor em relação àquela do resto do enunciado.

A unidade tonal em que *cê* está inserido é “*cê pode mandar pa todo mundo*”, observável na figura 24 acima. Os valores da duração das sílabas fonéticas permanecem os mesmos, portanto não vamos repetir esta parte. Contudo, quando se calcula a média dos valores de *z-score* das sílabas relativas à unidade tonal, o resultado é um pouco menor do que o anterior: -1,28 de desvio padrão. Mesmo assim continua maior que aquele de *cê*, de -3,44. A diferença percentual entre *cê* e a média das sílabas fonéticas do enunciado equivale a um valor 262% menor; em relação às sílabas fonéticas da unidade é 169% menor.

(58) bfamcv18-208-HER: //que *cê* falou / márcia //

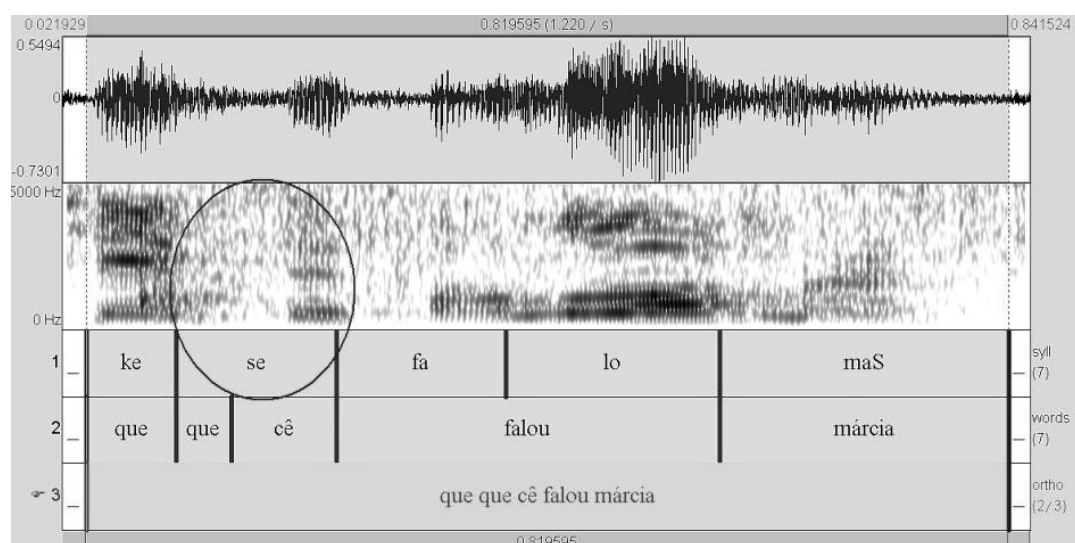


Figura 25: Tela do Praat com o texto (bfamcv18-208-HER): é posta em evidência a forma *cê* pré-verbal.

O espectro evidencia a segmentação. Os valores de duração bruta são os que seguem:

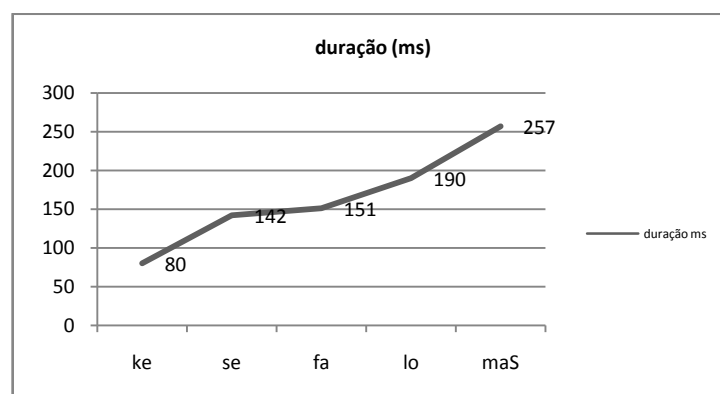


Gráfico 43: duração das sílabas fonéticas do texto (bfamcv18-208-HER).

Neste exemplo, a duração bruta da forma *cê* é de 142 ms. Seu valor normalizado é de -2,09 de desvio padrão. Comparado com a média de *z-score* das sílabas fonéticas do enunciado, que é de -1,21, obtém-se uma porcentagem de diminuição equivalente a 73%. Quando a normalização de *cê* é comparada com a média da unidade tonal a que pertence (“*que cê falou*”), a porcentagem de diminuição é muito maior, de 761%.

(59) bfamd102-45-BAL: // *cê sabe que aquelas caixinhas ali / ela / eu descobri ontem* //

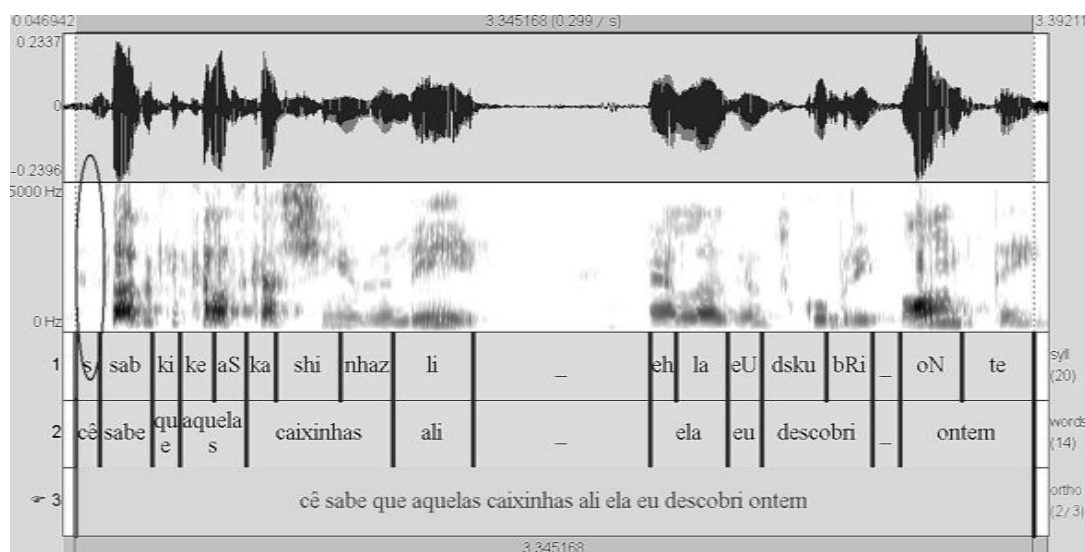


Figura 26: Tela do *Praat* com o texto (bfamd102-45-BAL): é posta em evidência a forma *cê* pré-verbal.

Este exemplo parece-nos especialmente significativo por mostrar o que está ocorrendo com a língua em uso. No espectrograma, pode-se observar que o informante BAL não chega a pronunciar a vogal [e] de *cê*, mas somente o fone [s]. Este não foi o único caso encontrado do que talvez seja uma tendência ao apagamento vocálico, reduzindo ainda mais a duração bruta do *cê* pré-verbal, que neste enunciado é de 83 ms, como mostra o gráfico abaixo.

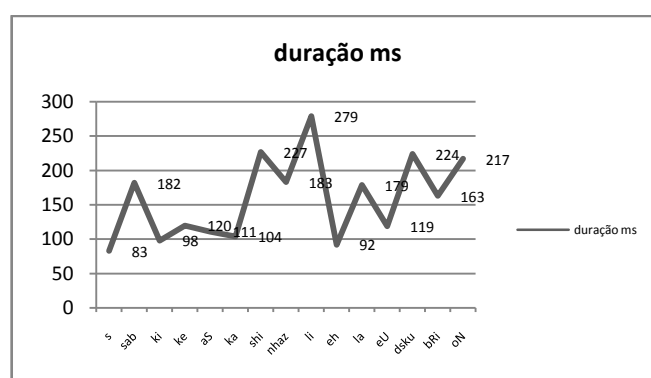


Gráfico 44: duração das sílabas fonéticas do texto (bfamd102-45-BAL).

A análise procedeu como nos outros exemplos: a duração normalizada da forma *cê* resultou ser -0,54 de desvio padrão. Comparando-a com a média de *z-score* das sílabas fonéticas do enunciado, que é de 0,75, se obtém uma porcentagem de diminuição equivalente a 172%. Até este ponto os resultados confirmaram nossa hipótese. Contudo, quando

comparamos a duração normalizada de *cê* com a média da unidade tonal a que pertence (“*cê sabe que aquelas caixinhas ali*”), a porcentagem de diferença aumentou para 7%. Isto talvez seja devido ao fato de a sílaba analisada ser composta somente pela consoante [s] havendo assim uma forma de compensação dos valores normalizados por parte do *script SGDetector* (BARBOSA P; A, 2006). Com certeza é algo que merecerá investigações futuras.

4.1.7.2. Forma *cê* em posição pós-verbal

A parte talvez mais esperada dentre os resultados do estudo resida nas próximas páginas. Como foi visto na seção de revisão da literatura, há grandes controvérsias sobre a possibilidade de *cê* assumir a posição pós-verbal, tanto mais ser objeto direto, indireto ou sujeito posposto. E foi realmente confortante poder validar nossas hipóteses com dados extraídos de um *corpus* de fala espontânea.

Exporemos em primeiro lugar os resultados de todas as situações de *cê* pós-verbais analisados para comentá-los em seguida.

TEXTO	FUNÇÃO SINTÁTICA	DURAÇÃO SIL Z/ENUNCIADO		PROP ENUNC	DURAÇÃO SIL Z /UT		PROP UT
bfamcv03-215-TON	OBJ PREP	-2,71	-1,96	-38,265306	-1,9	-2,38	-13,86554622
bfamcv04-453-BRU	SUJ	-1,85	-1,98	6,56565657	-1,85	-1,98	6,565656566
bfamcv05-240-CEL	OBJ DIR	-1,31	-2,07	36,7149758	-1,31	-1,85	29,18918919
bfamcv06-256-JOL	OBJ PREP	-0,61	-0,67	8,95522388	-0,61	-0,75	18,66666667
bfamcv06-327-FLA	OBJ PREP	-0,21	-1,83	88,5245902	-0,21	-1,83	88,52459016
bfamcv15-137-MAR	OBJ PREP	-1,95	-2,22	12,1621622	-1,95	-2,22	12,16216216
bfamcv16-219-EDE	OBJ PREP	0,95	-1,37	169,343066	0,95	-1,37	169,3430657
bfamcv17-5-KAR	OBJ PREP	1,97	-0,4	592,5	1,97	-0,53	471,6981132
bfamcv18-185-HER	OBJ PREP	-0,11	-2,18	94,9541284	-0,11	-2,18	94,95412844
bfamcv18-19-MAC	OBJ PREP	3,82	-0,27	1514,81481	3,82	0,48	695,8333333
bfamcv21-406-FLA	OBJ PREP	-0,3	-2,38	87,394958	-0,3	-2,44	87,70491803
bfamcv23-28-DIN	OBJ PREP	-1,82	1,49	-222,14765	-1,82	-2,52	27,77777778
bfamcv23-43-DIN	OBJ PREP	-0,67	0,31	-316,12903	-0,67	0,07	-1057,142857
bfamcv25-248-GLA	OBJ PREP	-1,53	-2,52	39,2857143	-1,53	-2,52	39,28571429
bfamd102-247-BAL	OBJ PREP	-0,34	-0,33	-3,030303	-0,34	-1,2	71,66666667
bfamd105-103-ANE	OBJ PREP	-0,49	1,88	-126,06383	-0,49	-1,15	57,39130435
bfamd110-102-REG	OBJ PREP	7,03	-0,02	35250	7,03	0,51	1278,431373
bfamd110-54-HEL	OBJ PREP	-0,36	-0,57	36,8421053	-0,36	-0,72	50
bfamd112-21-JMA	OBJ PREP	-0,33	-0,56	41,0714286	-0,33	-1,72	80,81395349
bfamd117-201-JUN	OBJ PREP	1,42	-0,74	291,891892	1,42	-0,62	329,0322581
bfamd122-175-HHA	OBJ PREP	-0,98	-0,724	-35,359116	-0,98	-0,724	-35,35911602
bfamd127-68-LUA	OBJ PREP	1,97	0,25	688	1,97	-0,53	471,6981132
bfamd128-177-ALV	OBJ PREP	1,32	-1,02	229,411765	1,32	-1,03	228,1553398
bfamd131-426-LIQ	OBJ PREP	3,6	2,28	57,8947368	3,6	0,14	2471,428571
bfamd133-107-JAN	OBJ PREP	-0,93	-1,13	17,699115	-0,93	-2,74	66,05839416
bfammn20-89-ALO	OBJ PREP	-1,22	0,17	-817,64706	-1,22	-2,04	40,19607843
bfammn29-6-JDL	OBJ PREP	3,52	2,85	23,5087719	3,52	2,85	23,50877193
bfammn35-233-JMA	OBJ PREP	3,59	0,64	460,9375	3,59	1,09	229,3577982
bpubcv02-266-CAR	OBJ PREP	-0,79	-1,47	46,2585034	-0,79	-1,88	57,9787234
bpubdl04-2-MUR	OBJ PREP	1,03	-1,74	159,195402	1,03	-1,74	159,1954023
bpubdl04-260-ELI	OBJ PREP	-0,73	-0,64	-14,0625	-0,73	-0,84	13,0952381

Tabela 12: resultados das análises acústicas da forma *cê* em posição pós-verbal no *corpus* C-ORAL-BRASIL.

A tabela segue o mesmo modelo da tabela (11), portanto, poupamos o leitor de explicações repetidas. Observemos que dos 31 casos examinados, 28 apresentam valores de duração normalizada de *z-score* superiores à média dos valores de *z-score* das sílabas fonéticas das unidades tonais a que pertencem. A tabela mostra que quando são comparados

os valores normalizados de *cê* pós-verbal com a média dos valores de *z-score* das sílabas fonéticas dos enunciados, os resultados variam: às vezes são superiores e às vezes inferiores. Isso confirma a hipótese de que o termo de comparação deva ser a unidade tonal para evitar a interferência de diferentes funções informacionais em uma pesquisa que procura valores motivados sintaticamente. Retirando a possibilidade de haver elementos que possam influenciar os resultados de forma não previsível, podemos observar a correlação existente entre sílabas de *cê* longas, e posição pós-verbal.

Há três casos, que examinaremos adiante, que não se encaixam na hipótese aqui apresentada. Contudo, parece-nos bastante revelador o fato de que : tenham sido encontrados em *corpus* de fala espontânea exemplos até agora negados por muitos, e supostos por poucos, como o caso de objeto direto e sujeito posposto; a maioria dos casos analisados confirmou a hipótese de que possa haver uma duração dupla, associada à marca de acento: uma curta, pré-verbal e uma longa, pós-verbal; a forma *cê* pós-verbal, por ter duração longa, seria tônica. A relação entre função e duração se confirma mais consistente do que aquela entre função e forma segmental.

Dos 28 casos que se encaixam nesta hipótese, foi encontrado um cuja função é de objeto direto, um de sujeito posposto (em oração clivada) e os outros de objeto preposicionado. Vejamos agora alguns exemplos.

(60) bfamecv05-240-CEL: *CEL: [240] chamou *cê* não / sô //

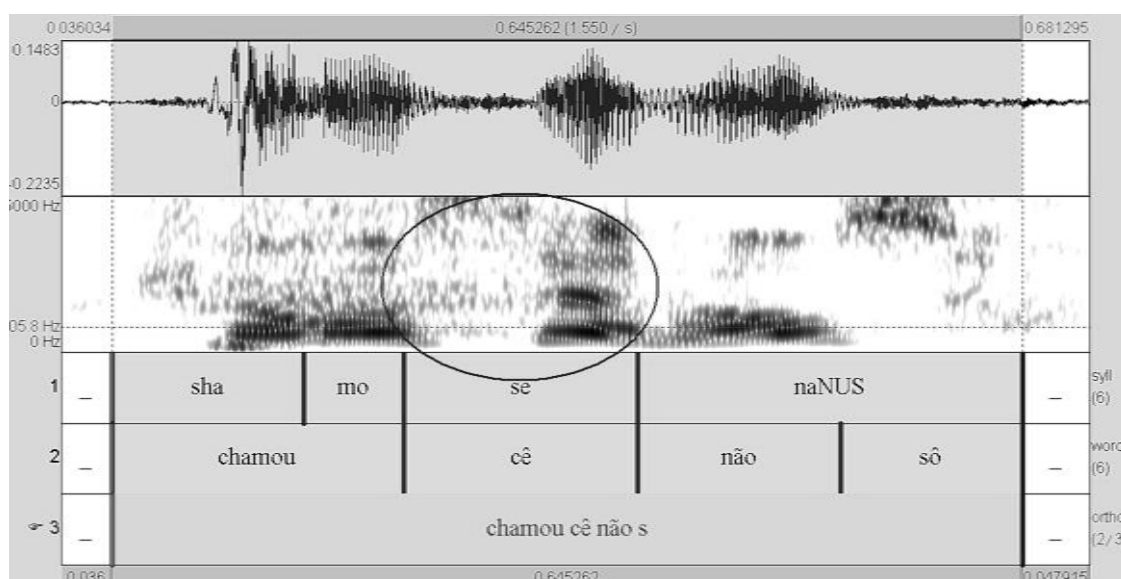


Figura 27: Tela do *Praat* com o texto (bfamecv05-240-CEL): é posto em evidência o *cê* pós-verbal.

O exemplo acima mostra um caso de *cê* pós-verbal com função de objeto direto. A análise do espectrograma pode abrir uma discussão sobre se o fone [o] final da palavra “chamou” não possa ser o fone [o] inicial da palavra “ocê”, por exemplo. Contudo, apesar de à oitiva poderem permanecer dúvidas, acreditamos que nossa segmentação respeitou o que foi pronunciado, pois não parece haver alongamento vocálico no fone [o], que é seguido imediatamente pela fricção do fone [s], visível no espectrograma.

Quanto às medidas de duração bruta são como segue:

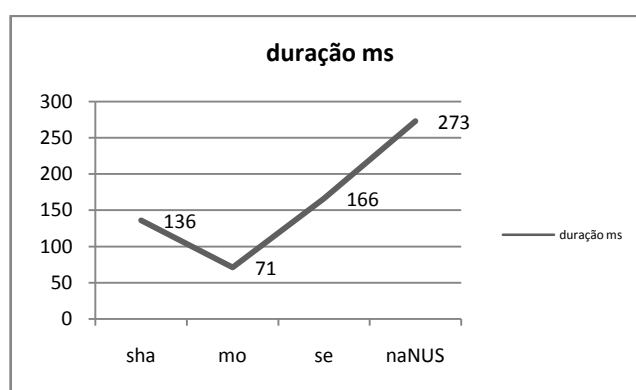


Gráfico 45: duração das sílabas fonéticas do texto (bfamcv05-240-CEL).

A duração bruta da forma *cê* é de 166 ms. Sua duração normalizada é de -1,31 de desvio padrão. Comparamos este valor com a média das durações normalizadas em *z-score* da unidade tonal, que é de -1,85. Calculamos então a proporção da diferença em relação à unidade tonal, cujo resultado é 29%. A diferença entre *cê* e a unidade tonal equivale a 29% do valor médio da unidade tonal, sendo *cê* maior que média da unidade tonal. Isto confirma que esta sílaba é mais longa do que a média de todas as sílabas fonéticas da unidade tonal a que pertence e, portanto, pode ser considerada tônica.

Observemos agora o exemplo em que *cê* ocorre como sujeito posposto. Esclarecemos que é um caso peculiar, pois se trata de uma construção clivada e, portanto, uma estrutura específica que não pode ser generalizada como caso de sujeito posposto.

(61) bfamcv04-453-BRU: *LUC: [428] *pode ir* // *BRU: [429] *pode* // *LUC: [430] *a* [1] *ea* [1] *ea* *ela tem que adivinhar* // *BRU: [431] *ela vai adivinhar* // *HEL: [432] *hhh* *barril* // [433] *alambique* // [434] *garrafa* // *BRU: [435] *uhn* // *HEL: [436] *gargalo* // [437] *tampa* // [438] *&he / tampinha* // *LUC: [439] *hhh* *<tampinha>* // *HEL: [440] *<&abri>* + [441] *&he / bico* // [442] *&he / &he / saca-rolhas* // [443] *champanhe* // [444] *hhh* *garrafa* // [445] *&garra* + [446] *rolha* // *BRU: [447] *uhn* // [448] *yes* *hhh* // [449] *yyyy* // *HEL: [450] *<ai / que foda>* // *BRU: [451] *<yyyy / né>* // *HEL: [452] *<Nossa>* // *BRU: [453] *agora é cê que faz* // *LUC: [454] *pera aí* // [455] *xá eu / voltar* *<a coisa>* // *HEL: [456] *<sério>* // *LUC: [457] *<pera aí / pera pera aí>* / *só um segundo* / *só vão esperar o* [1] *a ampulheta* / *<desneçoar>* // ¹³³

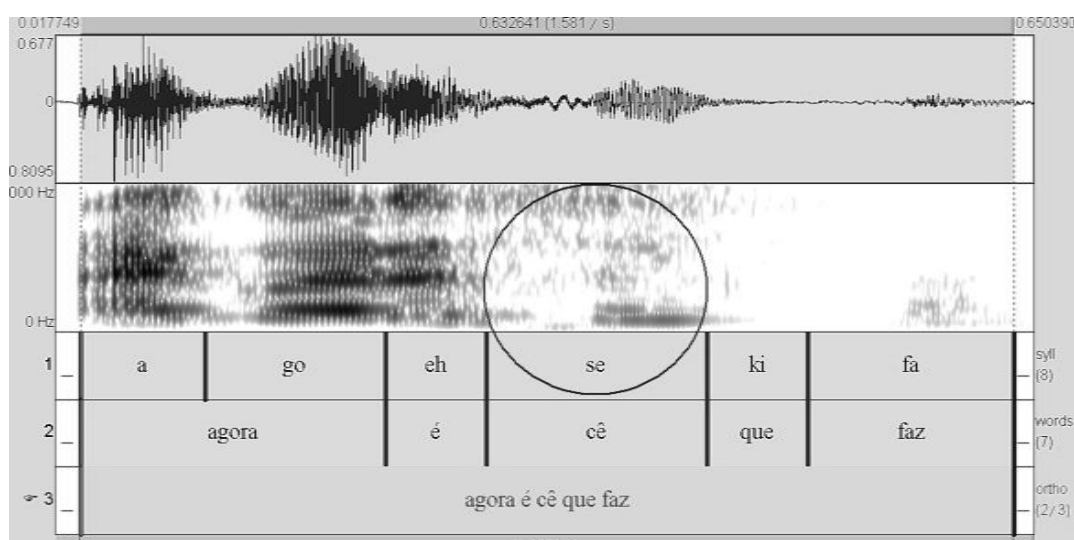


Figura 28: Tela do *Praat* com o texto (bfamcv04-453-BRU): é posto em evidência o *cê* sujeito posposto.

A duração bruta da forma *cê* é de 149 ms, como vemos pelo gráfico abaixo.

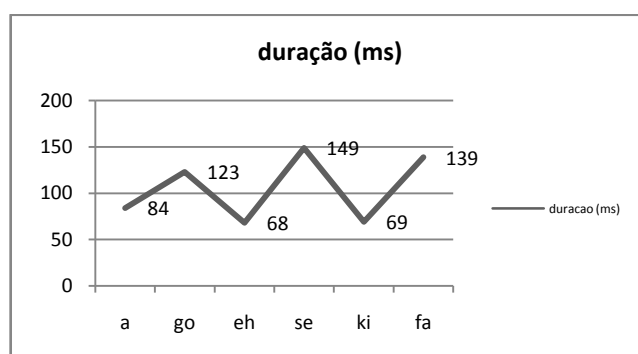


Gráfico 46: duração das sílabas fonéticas do texto (bfamcv04-453-BRU).

¹³³ Para melhor compreensão do enunciado analisado, em alguns casos, como este, citamos uma parte maior do texto.

A duração normalizada da sílaba é de -1,85 de desvio padrão. Calculando a proporção da diferença entre *cê* e sua respectiva unidade tonal, obtemos o valor de 6,56%. A sílaba *cê* é maior que a sílaba média da unidade tonal 6,56% do valor da respectiva unidade tonal.

Vejam agora um caso em que *cê* é preposicionado.

(62) bpubdl04-260-ELI: *ELI: [...] [260] *eu tava a fim de convencer minha irmã a pegar umas aulas com **cê** / pra ela aprender a andar na rua //*

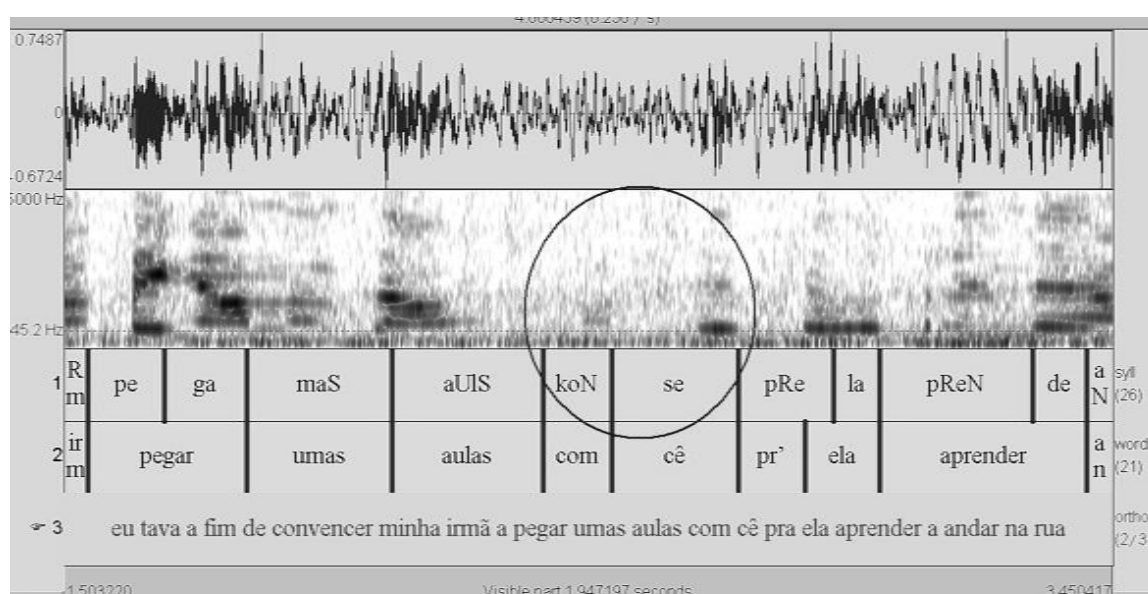


Figura 29: Tela do *Praat* com o texto (bpubdl04-260-ELI): é posto em evidência o *cê* preposicionado.

O gráfico abaixo mostra as durações brutas das sílabas fonéticas do enunciado.

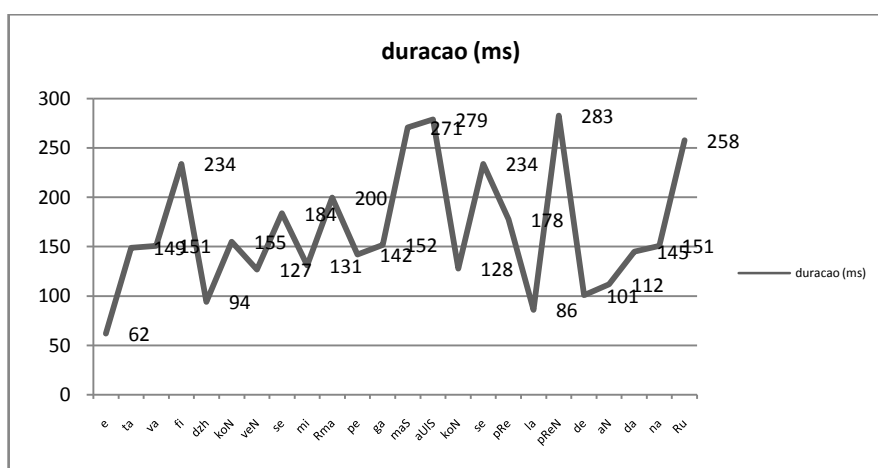


Gráfico 47: duração das sílabas fonéticas do texto (bpubdl04-260-ELI).

A duração bruta da sílaba [ˈse] é 184 ms. Seu valor normalizado é de -0,73 de desvio padrão. O cálculo foi feito a partir da unidade tonal “*tava a fim de convencer minha irmã a pegar umas aulas com cê*”. O valor da proporção da diferença em relação à unidade tonal é 13%, um valor positivo. Portanto, também neste caso podemos afirmar que a sílaba é acentuada.

(63) bfamcv23-28-DIN: *DIN: [24] *quando cê terminou a primeira vez com a Mari // [25] <foi um> término instantâneo igual eu tive // *MAR: [26] <ham ham> // [27] ham ham // *DIN: [28] eu falei p' cê / não / se ocê [2] <se eu quero que ocê mude / e passe a fazer xis coisas> / então eu também tenho que mudar / *CAB: [29] <cê corta agora / Bruno> // [30] <é / agora cê sabe / ué> // [31] entendi // *DIN: [32] / de forma / a tornar bacana a sua mudança //*

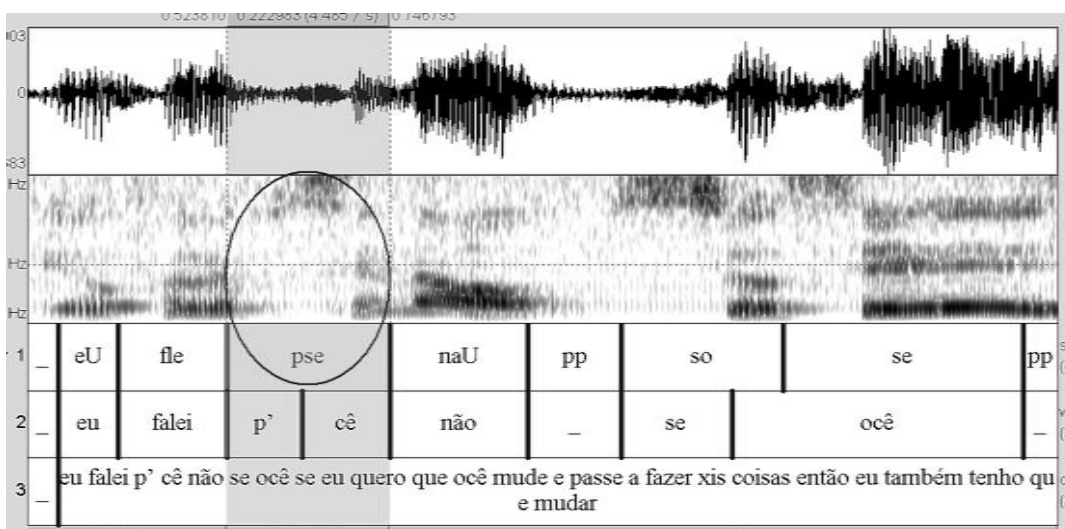


Figura 30: Tela do Praat com o texto (bfamcv23-28-DIN): é posto em evidência o *cê* preposicionado.

A duração bruta da sílaba [ˈpse] é de 223ms, como vemos pelo gráfico abaixo.

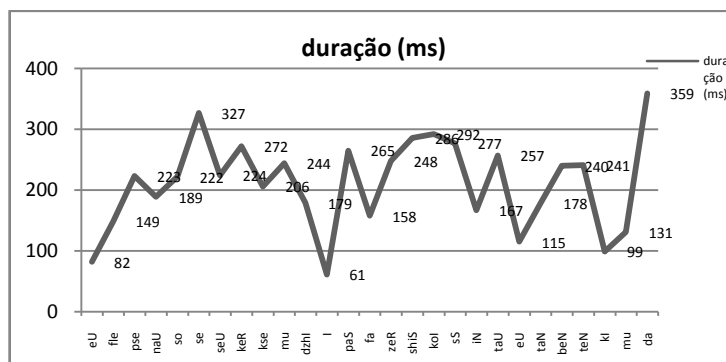


Gráfico 48: duração das sílabas fonéticas do texto (bfamcv23-28-DIN).

A duração normalizada da sílaba é de -0,67 de desvio padrão. Calculando a proporção da diferença entre *cê* e sua respectiva unidade tonal, obtemos o valor de 27,8%. A sílaba ['pse] é maior que a unidade tonal 27,8% do valor da respectiva unidade tonal.

(64)bfamdl02-247-BAL: *BEL: [240] &he / &e [/1] quando eu fui / minha mala tinha / tipo / quatorze quilos // *BEL: [241] aí na hora que eu voltei / eu fiquei toda feliz que ela tinha dez quilos / e eu achando / que eu arrumei ela melhor / e por isso ela tava pesando menos // *BEL: [242] olha o grau / da burrice // *BEL: [243] quando eu cheguei aqui / todas as minhas calças tinham ficado lá hhh // *BEL: [244] hhh todas as minhas blusas de frio tinham ficado lá // *BEL: [245] tipo / quatro quilos a menos / eu achei que é porque eu tinha arrumado melhor hhh // *BAL: [246] ô Bel // *BAL: [247] eu vou / falar o que que a Bruna falou com **cê** assim / lesada // *BEL: [248] hhh que lerdeza // *BAL: [249] muito lesada //

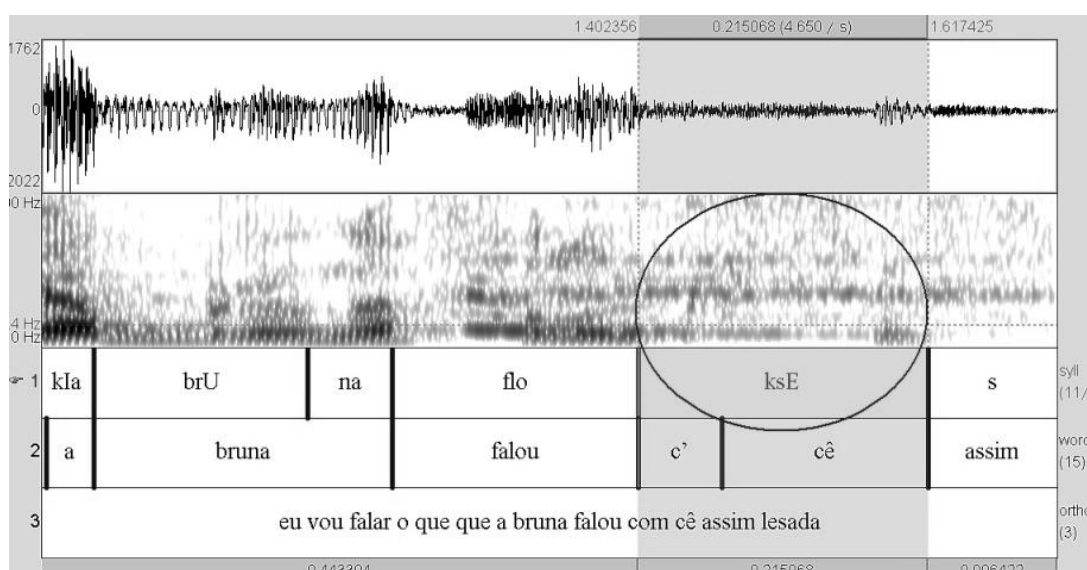


Figura 31: Tela do *Praat* com o texto (bfamdl02-247-BAL): é posto em evidência o *cê* preposicionado.

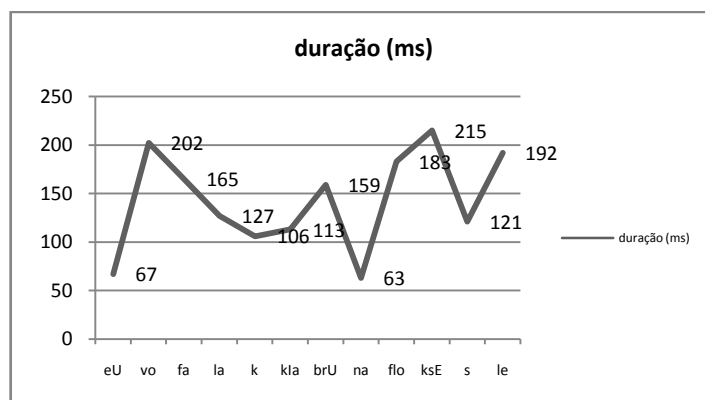


Gráfico 49: duração das sílabas fonéticas do texto (bfamdl02-247-BAL).

A duração bruta da sílaba ['kse] é 215 ms. Seu valor normalizado é de -0,34 de desvio padrão. O cálculo foi feito a partir da unidade tonal. O valor da proporção da diferença em relação à unidade tonal é 71,7%, um valor positivo.

Antes de passar à exposição dos casos que não satisfizeram nossa hipótese, vejamos um último exemplo de *cê* preposicionado.

(65) bfammn29-6-JDL: *JDL: [3] <é> // [4] <eu falei> + [5] eu falei / meu filho // [6] mas / só falo **com *cê*** isso // [7] *ocê* é zé-mandado do juiz / ou é do Juquita // [8] e' disse / é do juiz // [9] se for de Juquita / *nũ* rancava não // [10] *nũ* rancava mesmo / que nós *nũ* assina / o menino // [11] nós *nũ* aceitava // [12] mas é de juiz / sior po' arrancar / mas eu vou atrás d' *ocês* //

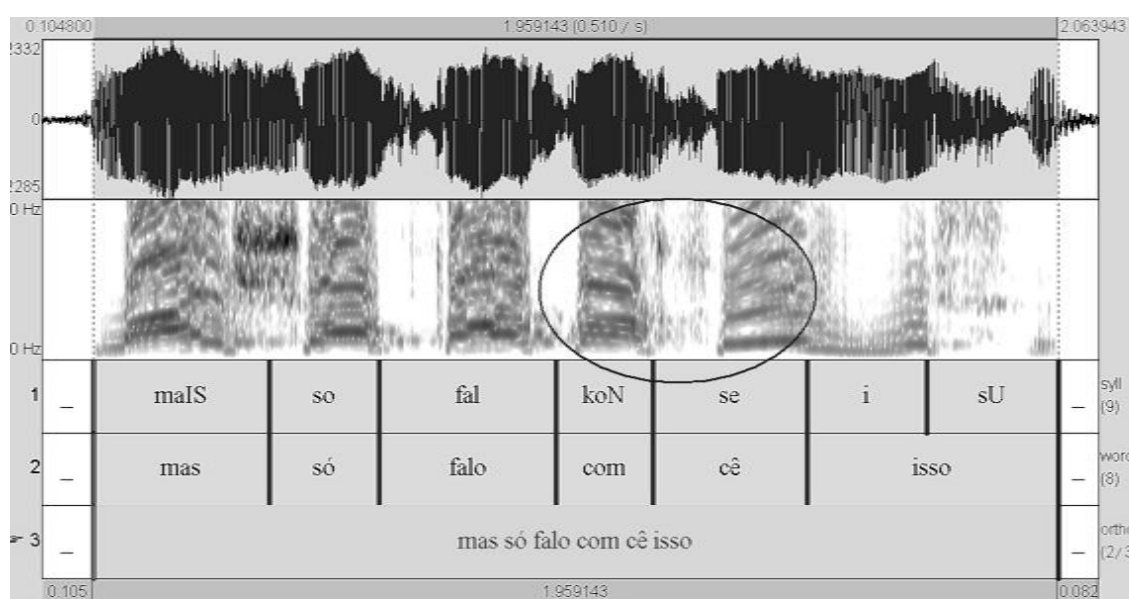


Figura 32: Tela do Praat com o texto (bfammn29-6-JDL): é posto em evidência o *cê* preposicionado.

Como é possível observar no gráfico abaixo, a duração bruta de *cê* é de 314ms.

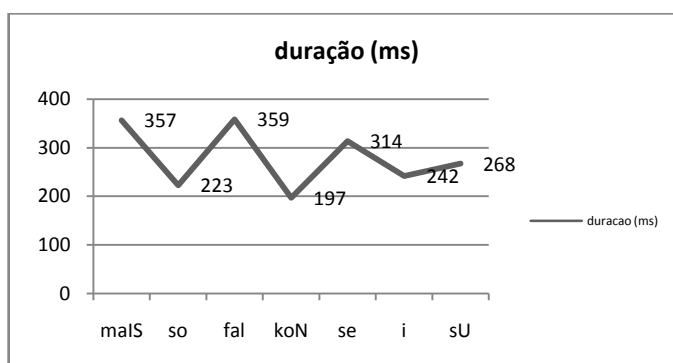


Gráfico 50: duração das sílabas fonéticas do texto (bfamnm29-6-JDL).

Sua duração normalizada é 3,52 de desvio padrão. Calculando a proporção da diferença em relação à unidade tonal obtemos 23,5%, outro valor positivo.

Os exemplos a seguir são casos em que as normalizações e, portanto, as proporções, resultaram em valores negativos, não respeitando assim a nossa hipótese.

- (66) bfamcv03-215-TON: *TON: [211] *morreu / ué // *CAR: [212] morreu o quê / <Toninho> // *TON: [213] <morreu> / essa <bola> / né // *CAR: [214] <nts> // *TON: [215] *tô te falando com **cê** que morreu / uai // *TON: [216] olha / p' **cê** ver //**

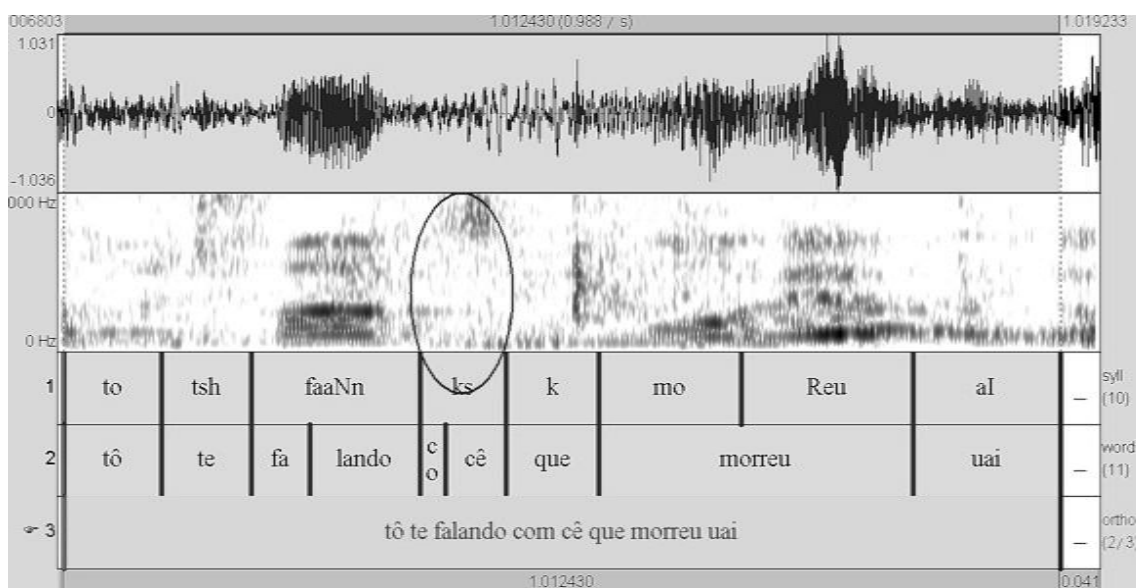


Figura 33: Tela do Praat com o texto (bfamcv03-215-TON): é posto em evidência o *cê* preposicionado.

O espectrograma mostra que o informante passou da oclusiva [k] à fricativa sibilante [s] e não pronunciou a vogal [e]. Já vimos em outro exemplo esta tendência ao apagamento vocálico, e neste caso também acreditamos que o cálculo da normalização de uma sílaba desprovida de vogal possa ter condicionado os dados. A duração bruta de [ks] é de 88ms, seu valor normalizado é de -2,71 de desvio padrão. Os cálculos da proporção da diferença em relação à unidade tonal fornecem um resultado negativo de 13,83%, significando que a sílaba [ks] é menor em relação ao valor da unidade tonal. A sílaba seria, portanto, curta, ou seja, átona.

Além do fato da sílaba carecer de um segmento, justamente aquele vocálico, pode haver outras explicações para esta idiosincrasia. O [ks] não exerce aqui uma verdadeira função de objeto preposicionado, pois, analisando por inteiro o enunciado, pode-se notar que em “tô te falando com cê” há uma retomada do pronome *te*. O [ks] seria pronunciado como se fosse de fato um clítico justamente por ser uma reduplicação. Por fim, a impressão é que [ks] se comporte acusticamente como um único segmento africado e não como dois segmentos, um oclusivo e outro fricativo.

Outro exemplo que apresentou resultados divergentes quanto à hipótese apresentada é o que segue.

(67) bfamcv23-43-DIN: *MAR: [39] <ah / cê tá querendo dizer que eu nũ> fui bacana com a Mari / naquela época // *DIN: [40] não // [41] que ela [/2] ela + [42] não // [43] cê [/1] a gente saiu / <ela não brigou com cê> em momento algum / te dando beijinho / carinho super <gente boa e ocê assim / conversando> comigo / tipo assim //

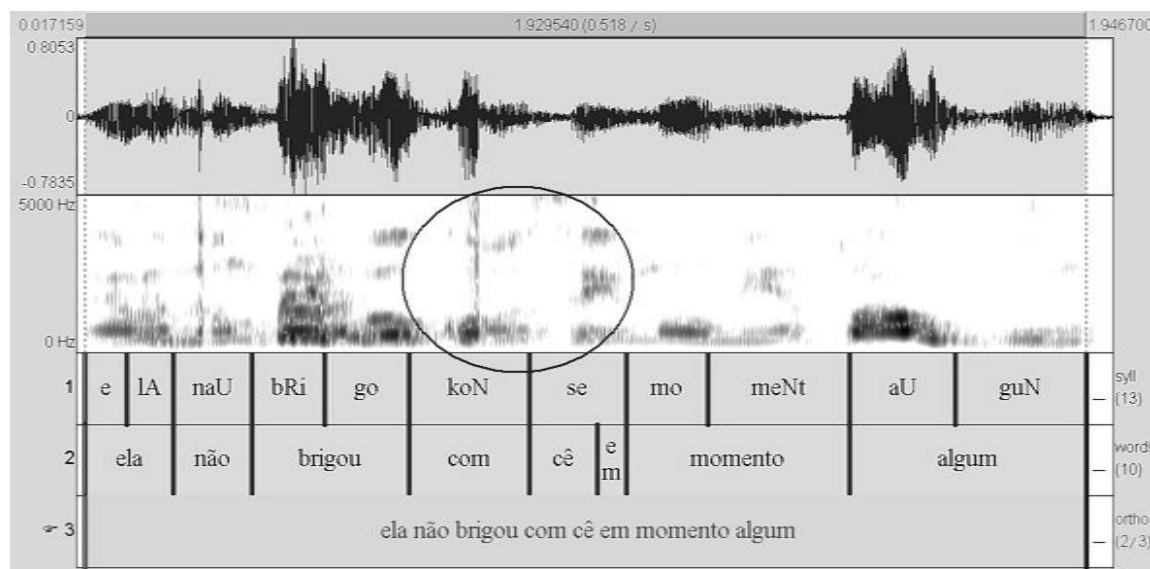


Figura 34: Tela do Praat com o texto (bfamcv23-43-DIN): para melhor visualização foi inserido somente o trecho da unidade tonal analisada.

A análise calculou que, segmentando o enunciado como mostra a figura acima, ou seja, considerando a forma pronominal como *cê* preposicionado, o valor bruto da sílaba *cê* é de 187 ms. Seu valor normalizado é de -0,67. A diferença entre *cê* e a unidade tonal equivale a 1057% do valor da unidade tonal, sendo *cê* menor que a mesma, pois o valor é negativo. Como não conseguimos entrever uma possível explicação para tal resultado, ficamos em dúvida sobre a correta segmentação do mesmo: havendo um alongamento vocálico do fone [o], foi contemplada a possibilidade de o que foi pronunciado não ser “com ocê”. Foi refeita a segmentação levando em conta esta hipótese.

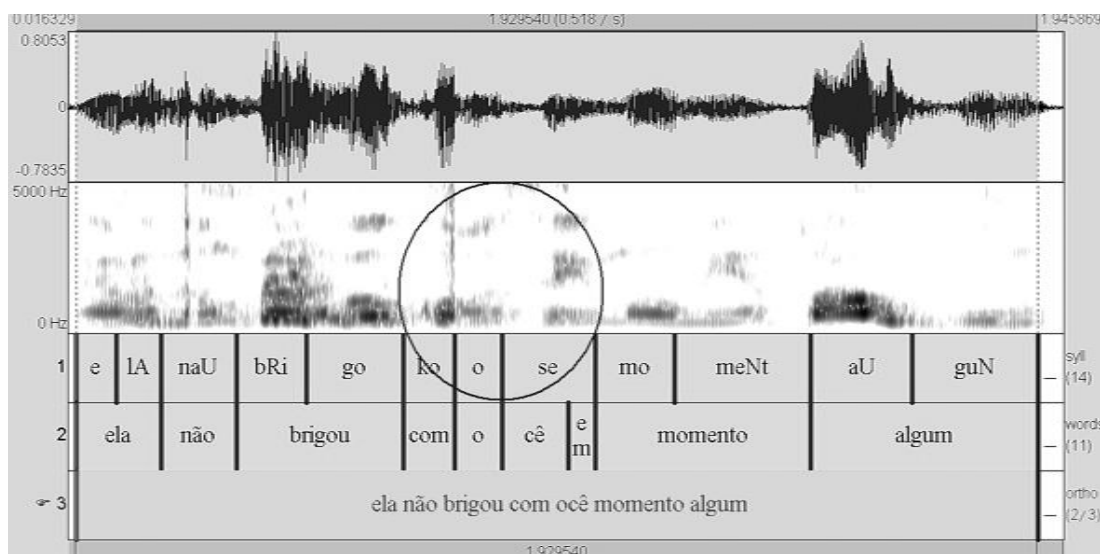


Figura 35: Tela do *Praat* com o texto (bfamcv23-43-DIN) segmentado com a forma *ocê*: para melhor visualização foi inserido somente o trecho da unidade tonal analisada.

Analisando a forma como *ocê*, a duração bruta da sílaba *cê* é de 188 ms, enquanto sua normalização fornece o resultado de -0,59 de desvio padrão. Os cálculos da proporção da diferença entre *cê* e sua unidade tonal fornecem valores mais razoáveis, mas ainda não aqueles esperados, equivalendo a 198,73% do valor da unidade tonal, sendo *cê* menor que a mesma, pois o valor é negativo.

Portanto, não temos aqui nenhuma explicação a oferecer para este resultado não esperado.

O último caso da forma *cê* pós-verbal que não se encaixou em nossa hipótese é o que segue.

(68) bfamd122-175-HHA: *HHA: [166] *então vou te dar de vinte-e-cinco* // *CCA: [167] *ah* // [168] *tava pensando que cê ia me dar de um* // *HHA: [169] *hhh é / ué* // [170] *pôr no cofre / aqui o'* // *CCA: [171] *é* // [172] *é dinheiro* // [173] *<tá valendo / né>* // *HHA: [174] *<um>* // [175] *não / &n [1] nũ vou comprar c' cê de vinte-e-cinco* // *CCA: [176] *<ah> / então tá bom* // [177] *ajudou* //

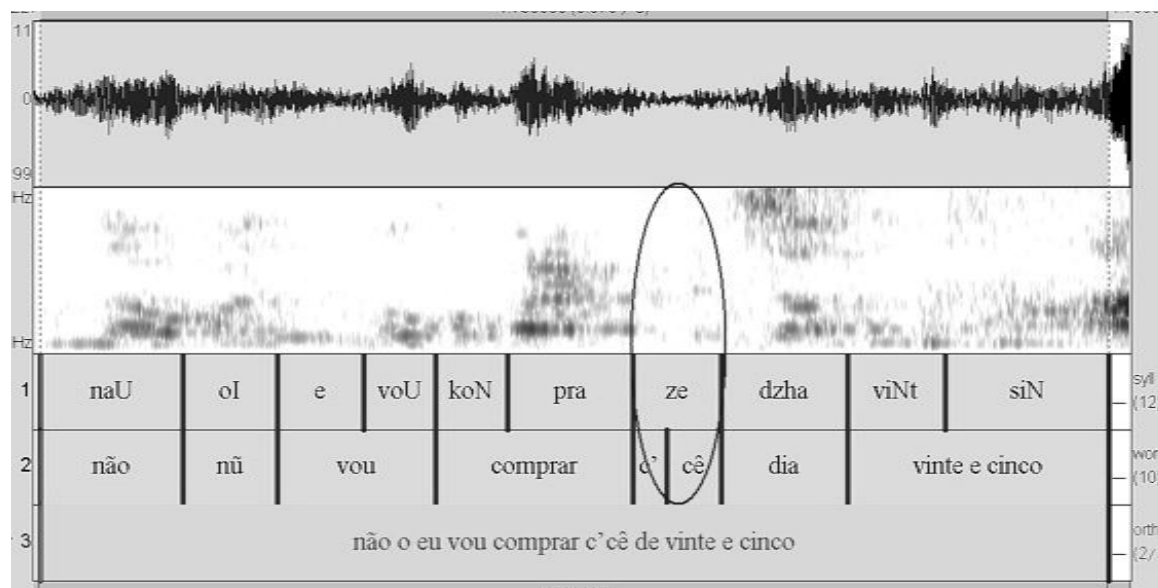


Figura 36: Tela do Praat com o texto (bfamev23-43-DIN): foi evidenciada a sílaba “cê”.

Neste enunciado também tivemos dúvidas quanto à correta segmentação. Inicialmente tentamos uma segmentação do que foi percebido à oitiva, ou seja, /kse/. Contudo, o espectro não mostra nem o silêncio precedente, nem a plosão da oclusiva velar /k/. Ao contrário, percebe-se uma vibração típica de uma fricativa vozeada. Por isso a segmentação acima. Contudo, mesmo assim, a sílaba registrou uma duração de 143ms. A duração normalizada é de -0,98 de desvio padrão. O cálculo da proporção da diferença em relação à unidade tonal resultou em 35,36% negativo. Este caso também mostrou uma incongruência em relação à nossa hipótese que não conseguimos explicar.

4.1.7.3. Forma *ocê* em posição pré-verbal

Serão agora apresentados os resultados da forma *ocê* em posição pré-verbal. A análise acústica contemplou um total de apenas seis exemplos porque, como foi referido acima, dois casos estudados entravam no rol daqueles com função informacional específica, e um (bfamev11-5) estava transcrito como *ocê*, mas após a análise do espectro percebeu-se que se tratava da forma *cê*.

A hipótese testada, mais uma vez previa que a sílaba ['se] em posição pré-verbal seria curta, portanto átona, e aquela em posição pós-verbal longa, isto é tônica.

Observemos os resultados.

TEXTO	FUNÇÃO SINTÁTICA	DURAÇÃO Z		PROP ENUNC	DURAÇÃO Z SIL /UT		PROP UT
		SIL/ENUNCIADO					
bfamcv21-43-FLA-1	SUJ	-3,1	-1,1	-181,81818	-3,1	-1,64	-89,02439024
bfamcv23-5-DIN	SUJ	-2,28	-2,11	-8,056872	-2,28	-1,94	-17,5257732
bfamcv23-7-DIN	SUJ	-2,05	-0,3	-583,33333	-2,05	-1,66	-23,4939759
bfamdl03-10-LUZ	SUJ	-1,37	0,53	-358,49057	-1,37	-0,38	-260,5263158
bfamdl03-35-LUZ	SUJ	-1,47	2,52	-158,33333	-1,47	-0,87	-68,96551724
bfammn03-56-ALO	SUJ	0,46	1,2	-61,6666667	0,46	1,2	-61,66666667

Tabela 13: resultados das análises acústicas da forma *ocê* em posição pré-verbal no *corpus* C-ORAL-BRASIL.

A tabela segue o modelo já apresentado para a forma *cê*, portanto, observemos a coluna da proporção da diferença entre os *z-scores* da sílaba e aqueles da média normalizada em relação à unidade tonal: todos os resultados são negativos, confirmando que a sílaba em questão é mais curta do que as outras. Isto significaria que *ocê* em posição pré-verbal segue o mesmo padrão de duração do *cê* pré-verbal, o que pode ser associado à sua natureza átona. Conferindo sua função morfossintática, os casos são todos sujeito.

Vejamos, portanto, alguns exemplos:

- (69) bfamcv21-43-FLA: *GUS: [39] <então> beleza // \$ [40] ah / porque só <conta os que estão dentro> do <quadrante / né> // *LUD: [41] <pera aí que eu tô tentando resolver> // *FLA: [42] <os dentro> // [43] **se ocê pegar o de dez / cê só marca o que tá dentro** // *GUS: [44] agora faz / todo o sentido // \$ [45] tudo <faz sentido> // *FLA: [46] <quando cê joga> com feijão / cada feijão vale dois / né //

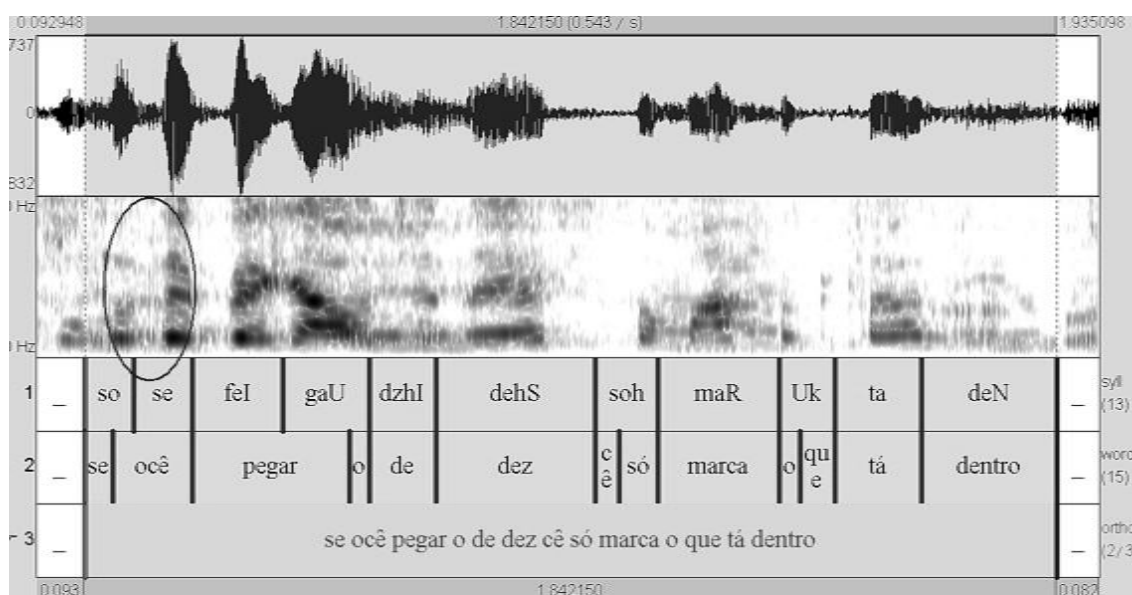


Figura 37: Tela do Praat com o texto (bfamcv21-43-FLA): foi evidenciada a sílaba fonética ['se].

A duração bruta da sílaba ['se] é de 111 ms.

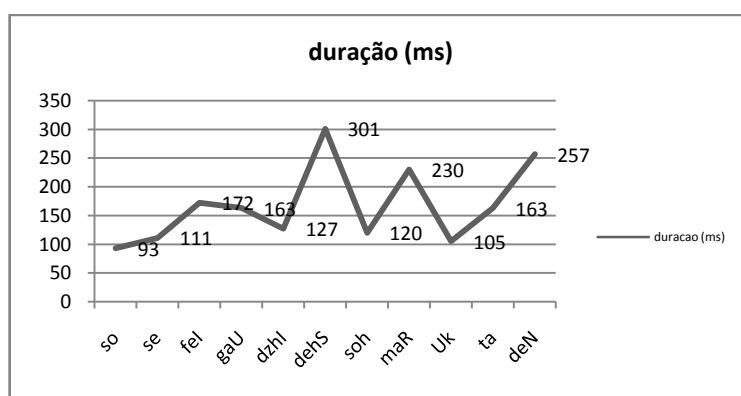


Gráfico 51: duração das sílabas fonéticas do texto (bfamcv21-43-FLA).

Sua duração normalizada é de -3,1 de desvio padrão. Calculamos então a proporção da diferença em relação à unidade tonal (“se ocê pagar o de dez”), cujo resultado é 89%: a diferença entre ['se] e a unidade tonal equivale a 89% do valor da unidade tonal, sendo ['se]

menor que a unidade tonal, visto que é um número negativo. Isto confirma que esta sílaba é mais curta do que a média de todas as sílabas fonéticas da unidade tonal a que pertence e, portanto, átona.

(70) bfamev23-7-DIN: *DIN: [7] <pois> é / ocê apareceu / várias vezes lá no diálogo //
 *MAR: [8] ham ham // *DIN: [9] ela / mas cê foi falar dos nossos problemas com o Marco e com os meninos / eu falei / <lógico> // *CAB: [10] <hum hum> // *MAR: [11] <é / são> meus amigos / pô //

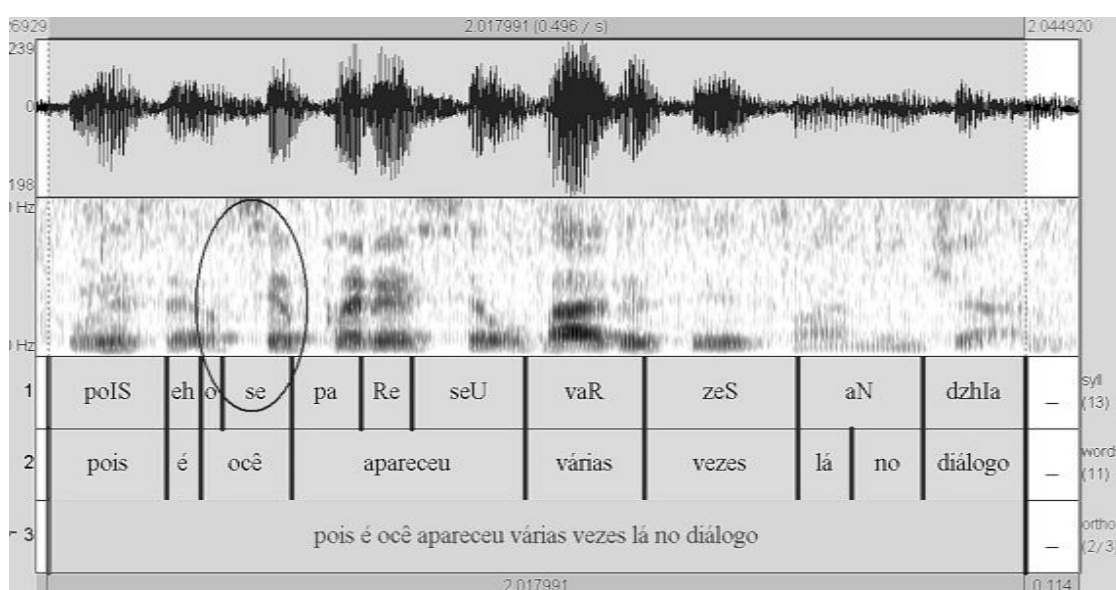


Figura 38: Tela do Praat com o texto (bfamev23-7-DIN): foi evidenciada a forma “ocê”.

No texto a duração bruta da sílaba fonética ['se] é de 143 ms.

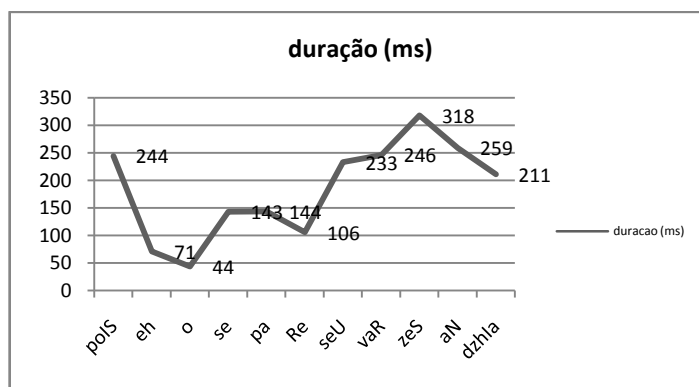


Gráfico 52: duração das sílabas fonéticas do texto (bfamev23-7-DIN).

A normalização da duração resultou em -2,05 de desvio padrão. O cálculo da proporção da diferença em relação à unidade tonal foi novamente um número negativo, -23,5%, confirmando que ['se] seria átono.

- (71) bfamdl03-35-LUZ: *LAU: [34] *quem que falou // *LUZ: [35] gente / eu / te falei / a Fabíola / eu falei se ocê nũ faria // *LUZ: [36] aí cê ainda falou / não / eu nũ vou fazer que nũ é uma / disciplina que tem a ver comigo // *LUZ: [37] eu concordei / nós conversamo tanto sobre isso / Lau hhh //*

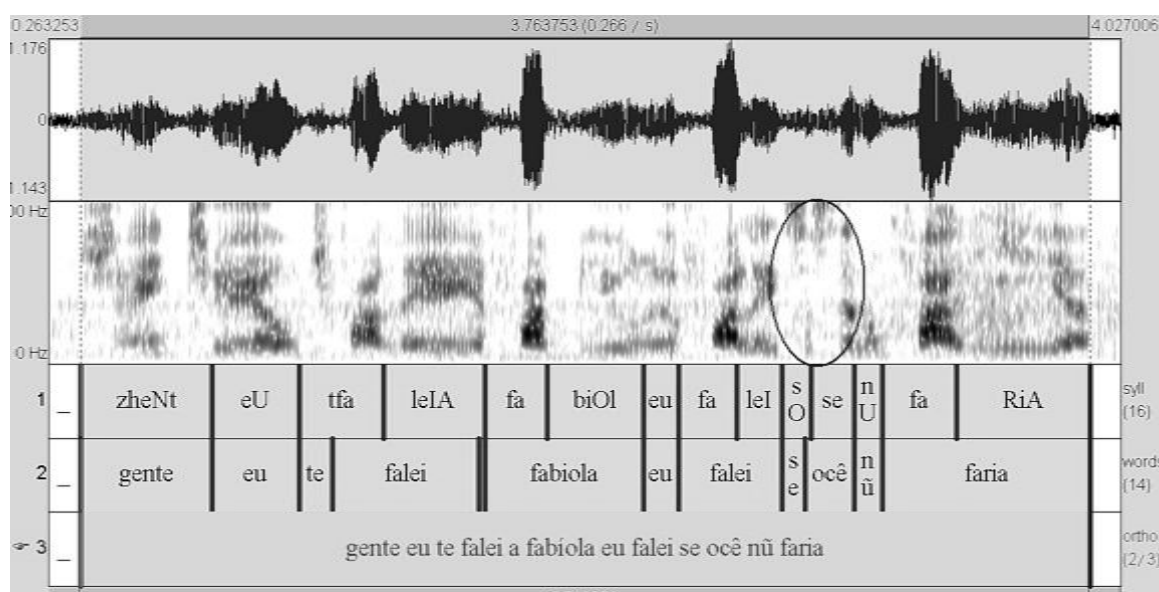


Figura 39: Tela do Praat com o texto (bfamdl03-35-LUZ): foi evidenciada a forma “ocê”.

A duração bruta de ['se] é de 161 ms.

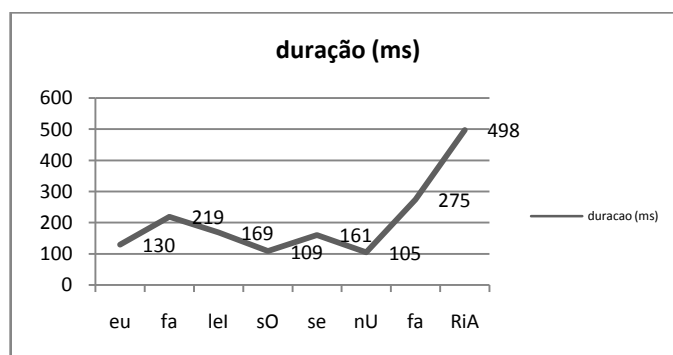


Gráfico 53: duração das sílabas fonéticas do texto (bfamdl03-35-LUZ).

A duração normalizada da sílaba é de -1,47 de desvio padrão. Foi calculada a proporção da diferença em relação à unidade tonal, cujo resultado é 69%, ou seja, a diferença

entre ['se] e a unidade tonal equivale a 69% do valor da unidade tonal, sendo ['se] menor que a unidade tonal, visto que é um número negativo. Isto confirma que esta sílaba é mais curta do que a média de todas as sílabas fonéticas da unidade tonal a que pertence e, portanto, átona.

4.1.7.4. Forma *ocê* em posição pós-verbal

Passemos agora à análise da forma *ocê* em posição pós-verbal. Foram estudados 14 exemplos deste caso. Encontramos um sujeito posposto, 12 objetos indiretos e um caso de duplo acusativo. A hipótese inicial previa que nesta posição os valores da proporção da diferença em relação à unidade a que a sílaba ['se] pertence seriam positivos, indicando sílabas longas, ou seja, tônicas. Isto se confirmou em 11 casos, enquanto três, que serão discutidos, fugiram à previsão.

Vejamos a tabela com os resultados gerais.

TEXTO	FUNÇÃO SINTÁTICA	DURAÇÃO Z		PROP ENUNC	DURAÇÃO Z SIL /UT		PROP UT
		SIL/ENUNCIADO					
bfamcv05-139-JOS	OBJ PREP	-0,44	-1,43	69,2307692	-0,44	-1,43	69,23076923
bfamcv05-450-MAR	OBJ DIR	-0,61	-1,33	54,1353383	-0,61	-2,12	71,22641509
bfamcv11-197-ONO	OBJ PREP	2,2	-0,16	1475	2,2	-2,02	-8,910891089
bfamcv11-30-ONO	OBJ PREP	1,67	-0,3	656,666667	1,67	-0,3	656,6666667
bfamcv21-317-EME	OBJ PREP	3,51	-0,06	5950	3,51	-0,37	1048,648649
bfamcv21-50-FLA	OBJ PREP	0,94	0,15	526,666667	0,94	0,14	571,4285714
bfamcv02-70-TER	SUJ	-0,38	-0,63	39,6825397	-0,38	-1,84	79,34782609
bfamcv33-249-DAN	OBJ PREP	0,03	-1,35	102,222222	0,03	-1,36	102,2058824
bfamdl19-201-MUD	OBJ DIR	-0,9	-0,1	-800	-0,9	-0,93	3,225806452
bfamdl19-73-AVI	OBJ PREP	-2,26	-1,63	-38,650307	-2,26	-2,32	2,586206897
bfamdl30-68-REN	OBJ PREP	2,01	-1,02	297,058824	2,01	-1,02	297,0588235
bfammn03-102-ALO	OBJ DIRETO (DUPLO)	-1,31	-0,76	-72,368421	-1,31	-0,87	-50,57471264
bfammn03-105-ALO	OBJ PREP	-0,97	-1,43	32,1678322	-0,97	-0,95	-2,105263158
bfammn34-142-MER	OBJ PREP	6,09	3,41	78,5923754	6,09	1,26	383,3333333

Tabela 14: resultados das análises acústicas da forma *ocê* em posição pós-verbal no *corpus* C-ORAL-BRASIL.

Os dados ressaltam os onze casos em que nossa hipótese se confirmou. Vamos agora analisar alguns exemplos.

(72) bfamd130-68-REN: *FLA: [66] *ô Rena / tinha um detergente aqui // *REN: [67] então // *REN: [68] falei c' ocê // *REN: [69] tinha um // *REN: [70] que aí a gente só trouxe um //*

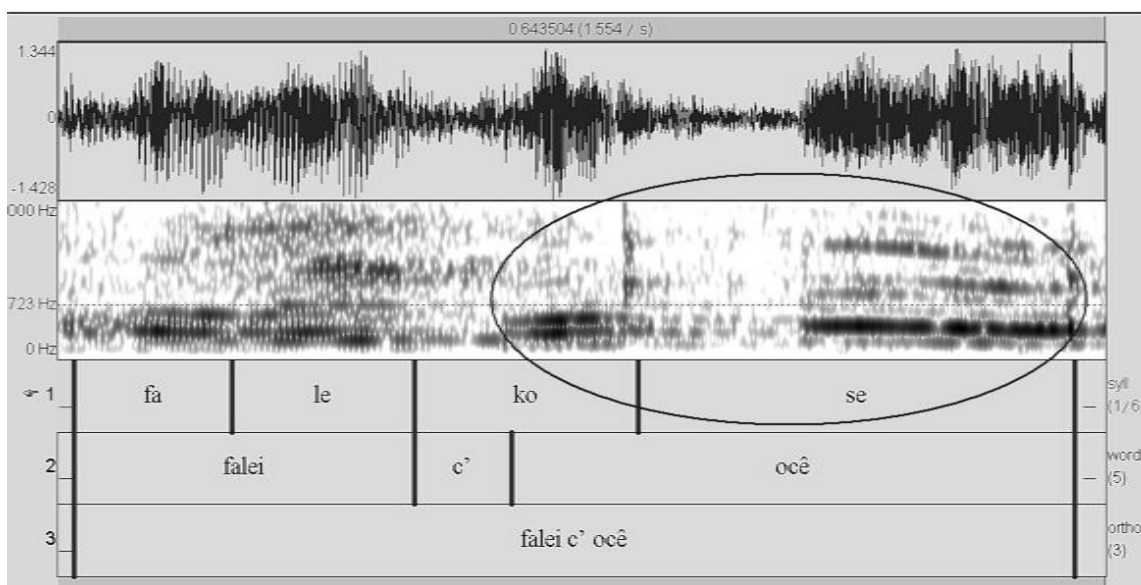


Figura 40: Tela do *Praat* com o texto (bfamd130-68-REN): foi evidenciada a forma “ocê”.

A duração bruta da sílaba ['se] é de 268 ms, a maior do enunciado, talvez também pela posição final.

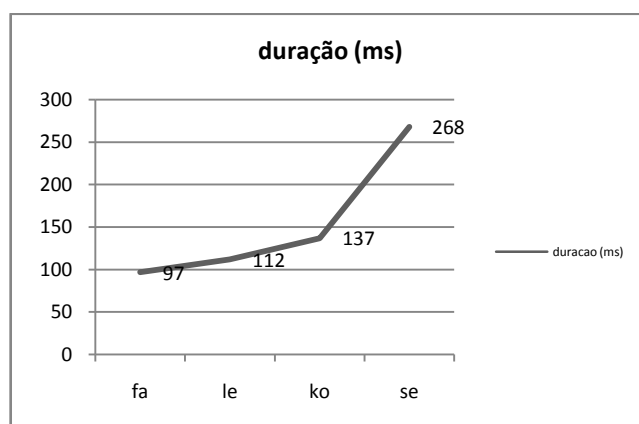


Gráfico 54: duração das sílabas fonéticas do texto (bfamd130-68-REN).

A normalização dos dados reverte o valor de ['se] para 2,01 de desvio padrão. O cálculo da proporção da diferença em relação à unidade tonal resultou em um número positivo, 297%, confirmando que ['se] seria tônico.

(73) bfamd19-73-AVI: // [70] *falou é não / porque eu deitei aí e sua mulher passou +*
 [71] *eu acordei de madrugada com ela passando a mão ni mim pa todo lado // [72] e'*
*falou / ah / então ocê é meu amigo mesmo // [73] **porque quem passou a mão n' ocê***
foi eu // [74] só pa ver se ocê tinha coragem de contar //

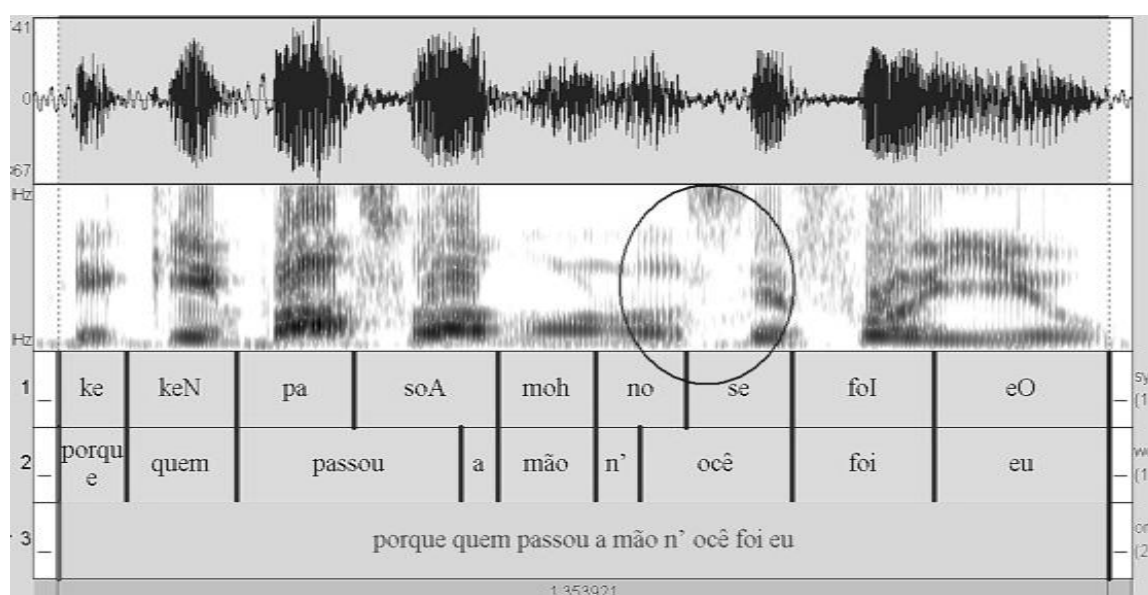


Figura 41: Tela do Praat com o texto (bfamd19-73-AVI): foi evidenciada a forma “ocê”.

A sílaba ['se] registrou uma duração bruta de 137 ms, como se vê pelo gráfico abaixo.

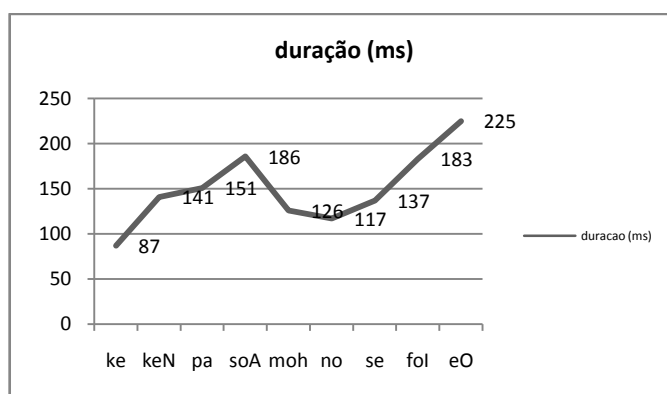


Gráfico 55: duração das sílabas fonéticas do texto (bfamd19-73-AVI).

A normalização transformou o valor de ['se] para 2,58 de desvio padrão. O cálculo da proporção da diferença em relação à unidade tonal resultou novamente em um número positivo, 2,58%, confirmando que ['se] seria tônico.

Vejam agora o exemplo em que *ocê* é sujeito posposto.

(74) bfamcv02-70-TER: *TER: [70] *conta ocê / Rute* // *TER: [71] *mas ea tem*
quarenta-e tantos // *TER: [72] *sei o tanto não* // *RUT: [73] *<pele> amor de Deus*
 // *TER: [74] *<hein>* // \$*RUT: [75] *nũ me convida pa ser &pa [1] madrinha não /*
hein // *TER: [76] *<não>* // *RUT: [77] *<já vou> avisando com antecedência* //
 *RUT: [78] *<morro de vergonha>* // *TER: [79] *<escuta só que que a> Dani tava*
<falando> // *RUT: [80] *<morro de> vergonha de <desfilar>* //

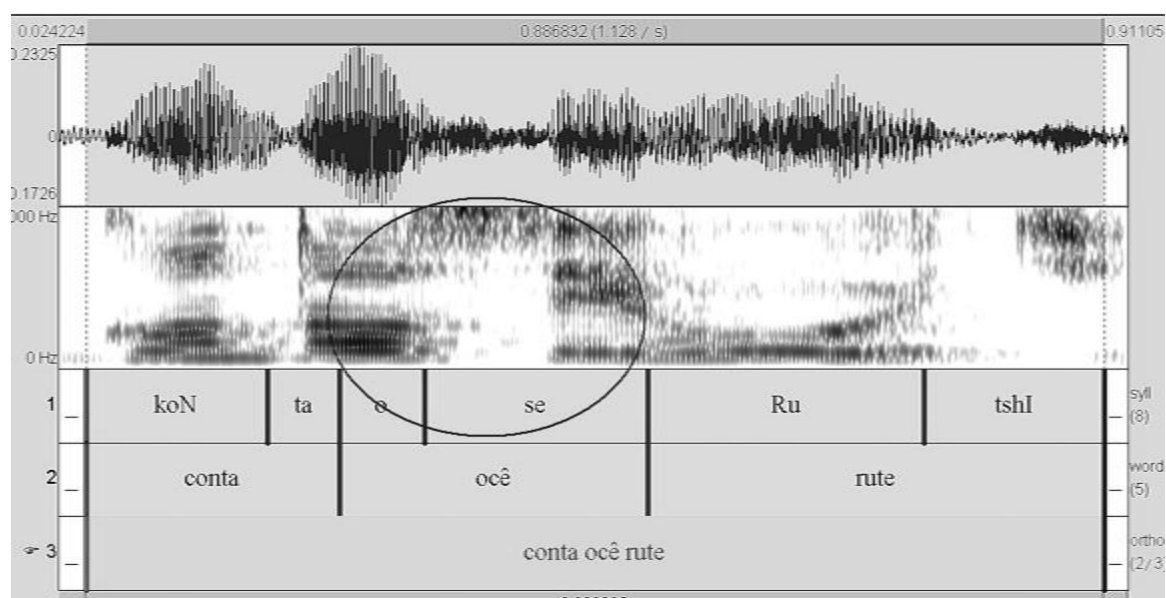


Figura 42: Tela do Praat com o texto (bfamcv02-70-TER): foi evidenciada a forma “ocê”.

A duração bruta de ['se] é de 184 ms.

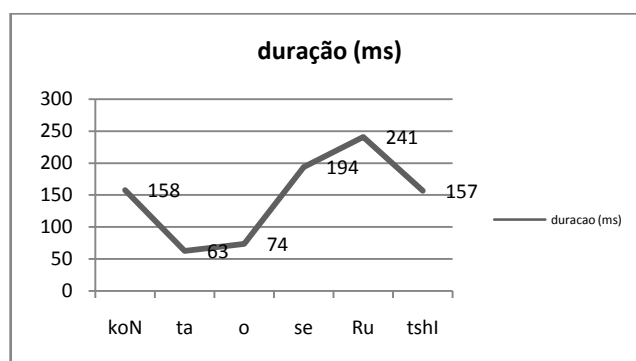


Gráfico 56: duração das sílabas fonéticas do texto (bfamcv02-70-TER).

A duração normalizada é de -0,38 de desvio padrão. Calculamos então a proporção da diferença em relação à unidade tonal cujo resultado é 79%, ou seja, a diferença entre ['se] e a unidade tonal equivale a 79% do valor da média da unidade tonal, sendo ['se] maior que a média da unidade tonal, visto que é um número positivo, confirmando que esta sílaba é mais longa do que a média de todas as sílabas fonéticas da unidade tonal a que pertence e, portanto, tônica.

Vejamos agora os exemplos que não satisfizeram a hipótese.

(75) bfamcv11-197-ONO: *ONO: [197] *já falei c' ocê topei co marido da mulher* //
 *TIT: [198] *hum hum* // *CAR: [199] *marido de quem* // *ONO: [200] *mulher que*
trabalha aqui // *TIT: [201] *marido da yyy* // *ONO: [202] *áí e' perguntou se / nós*
gostou dela / cês gostou dela // *ONO: [203] *sua mulher gostou* // *ONO: [204] *falei*
/ ah / n' é boa / trabalhadeira / e' falou assim / o' //

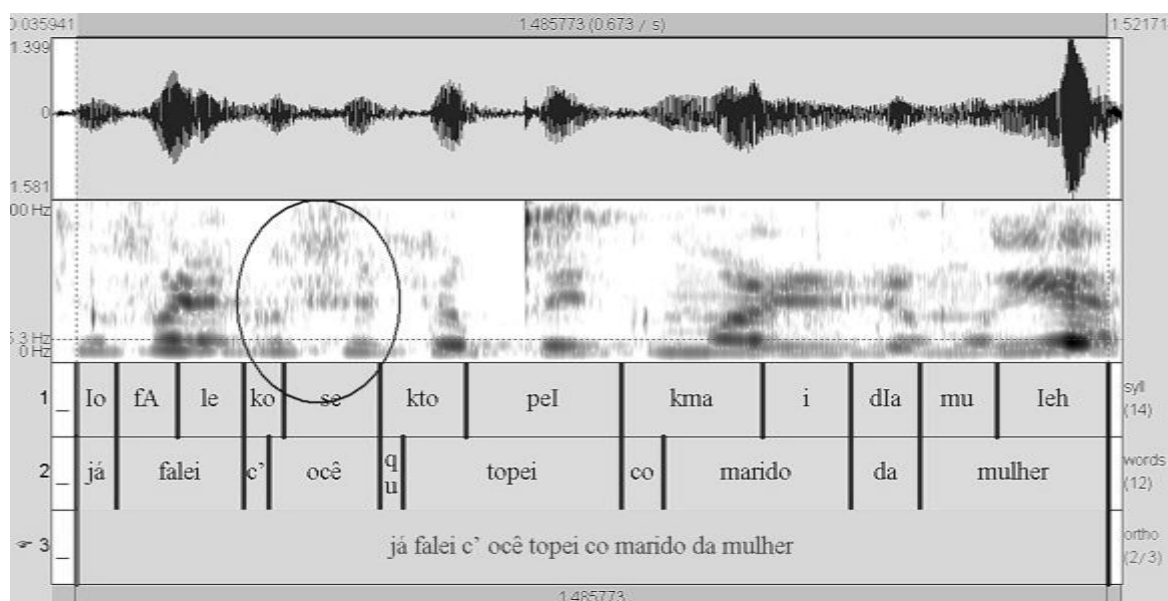


Figura 43: Tela do Praat com o texto (bfamcv11-197-ONO): foi evidenciada a forma “ocê”.

Sem passar por todos os cálculos, pode-se observar pela tabela 5 que o valor da diferença entre ['se] e a unidade tonal equivale a -8,9% do valor da média da unidade tonal, negativo. Tentamos re-segmentar o texto de outras formas, mas os resultados foram sempre negativos, portanto não conseguimos uma explicação plausível para o fenômeno.

O mesmo aconteceu com os próximos dois exemplos.

- (76) bfammn03-102-ALO: *ALO: [100] *ai / quando é três anos atrás mais ou menos / chega um filho dela / perto de mim / falou assim / ô Aloysio // *ALO: [101] minha mãe morreu // *ALO: [102] devia ocê uma conta na loja //*



Figura 44: Tela do Praat com o texto (bfammn03-102-ALO): foi evidenciada a forma “ocê”.

A duração bruta de [‘se] é de 166 ms, de fato não muito longa em relação às outras, como se vê pelo gráfico abaixo.

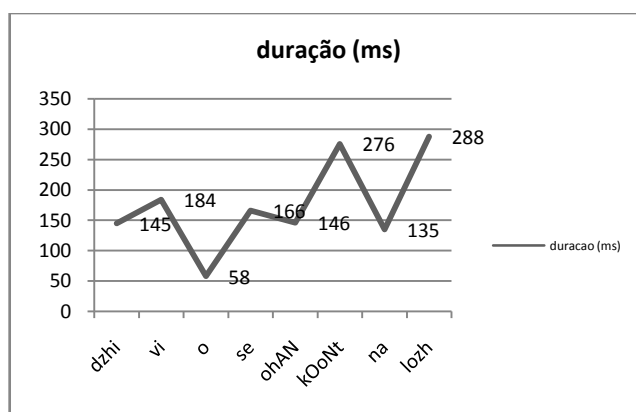


Gráfico 57: duração das sílabas fonéticas do (bfammn03-102-ALO).

Mesmo após a rodada de normalizações e os cálculos da proporção da diferença em relação à unidade tonal o resultado é um número negativo, -50,6%, indicando que [‘se] seria átono.

(77) bfammn03-105-ALO: *ALO: [103] *eu falei / ah / &Eu / I Eustáquio / larga pa lá
isso / uai // *ALO: [104] não / nũ tô nem / pensando não // \$ *ALO: [105] não / mas
eu quero só saber d' ocê // *ALO: [106] não // \$ *ALO: [108] e eu sei que ea devia /
porque / &he / foi / I foi &q / I nas véspera d' eu vim embora // *ALO: [109] foi &n
/ I em / I no / I na véspera do Natal de mil-novecentos-e sessenta-e-dois // *ALO:
[110] ela me fez uma compra lá / comprou presente pr' ocês todo / pos menino / roupa
/ e tal / e / ea nũ me pagou / também nũ cobrei / e já perdoei isso há muito tempo / nem
converso mais não //*

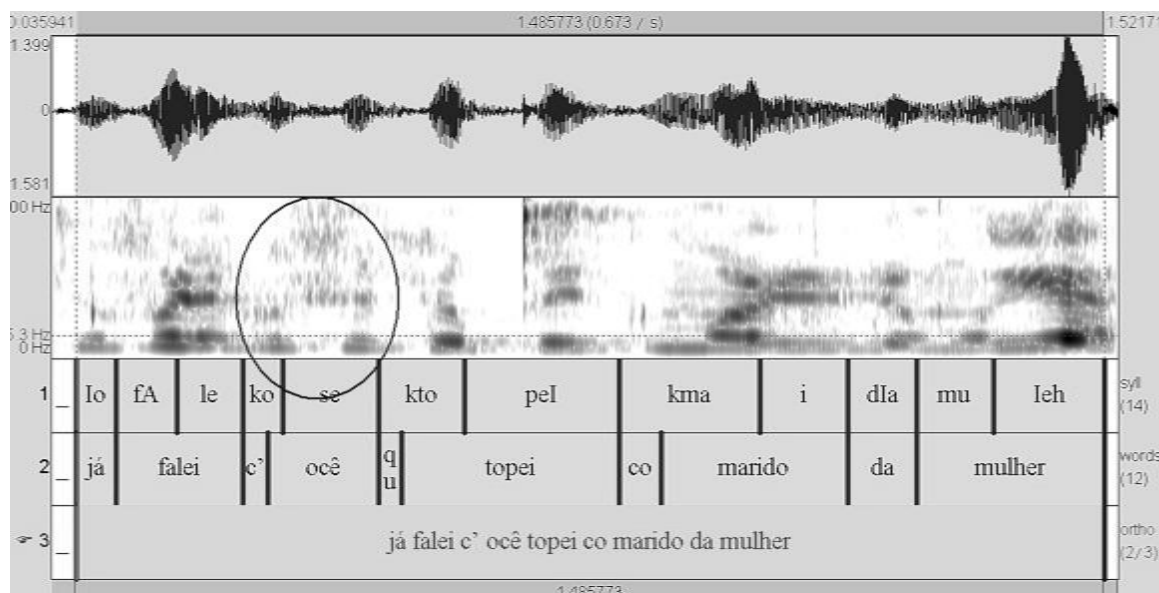


Figura 45: Tela do Praat com o texto (bfammn03-105-ALO): foi evidenciada a forma “ocê”.

Também neste caso a duração bruta não é das mais longas, como pode ser observado pelo gráfico.

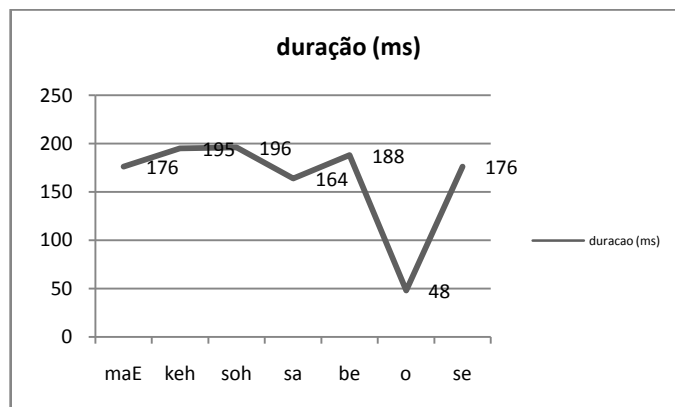


Gráfico 58: duração das sílabas fonéticas do (bfammn03-105-ALO).

A normalização dos dados e os cálculos da proporção da diferença em relação à unidade tonal retornaram também neste caso um número negativo, -2,1% de desvio padrão.

Quisemos mostrar os casos divergentes de nossa teoria porque, apesar de acreditarmos que haja uma forte correlação entre duração e tonicidade, talvez ainda tenhamos que observar outros fatores que possam influir sobre os dados.

4.1.7.5. Forma *você* em posição pré-verbal

Apresentaremos agora os resultados da forma *você* em posição pré-verbal. Os casos analisados acusticamente são 20. A tabela abaixo evidencia os valores obtidos.

TEXTO	FUNÇÃO SINTÁTICA	DURAÇÃO SIL Z/ EN		PROP ENUNC	DURAÇÃO SIL Z/ UT		PROP UT
bfamcv11-26-TIT	SUJ	-2,05	0,17	-1305,8824	-2,05	0,17	-1305,882353
bfamcv19-49-MAE	SUJ	-1,48	-1,01	-46,534653	-1,48	-1	-48
bfamcv19-79-MAE	SUJ	-2,19	-1,05	-108,57143	-2,19	-1,05	-108,5714286
bfamcv20-190-CAI	SUJ	-1,35	0,02	-6850	-1,35	0,59	-328,8135593
bfamcv20-227-ANC	SUJ	-1,68	0,13	-1392,3077	-1,68	-0,67	-150,7462687
bfamdl11-22-MAR-1	SUJ	-2,23	-1,33	-67,669173	-2,23	1,03	-316,5048544
bfamdl11-22-MAR-2	SUJ	-2,03	-1,33	-52,631579	-2,03	-1	-103
bfamdl11-70-MAR	SUJ	-1,82	-0,16	-1037,5	-1,82	-0,16	-1037,5
bfamdl14-25-CAR	SUJ	-1,68	-0,25	-572	-1,68	-0,63	-166,6666667
bfamdl33-187-HER	SUJ	-1,22	0,09	-1455,5556	-1,22	-0,14	-771,4285714
bfammn05-63-CAR	SUJ	-2,55	0,38	-771,05263	-2,55	0,38	-771,0526316
bfammn05-64-CAR	SUJ	-2,1	-0,61	-244,2623	-2,1	-0,61	-244,2622951
bfammn05-65-CAR	SUJ	-1,49	-0,015	-9833,3333	-1,49	-0,74	-101,3513514
bfammn12-104-JUN-1	SUJ	1,54	-0,23	769,565217	1,54	-0,06	2666,666667
bfammn12-104-JUN-2	SUJ	-1,79	-0,23	-678,26087	-1,79	-0,33	-442,4242424
bfammn12-144-JUN-1	SUJ	-1,2	1,9	-163,15789	-1,2	-0,59	-103,3898305
bfammn24-90-MIC	SUJ	0,4	0,82	-51,219512	0,4	1,24	-67,74193548
bfammn24-93-MIC	SUJ	0,008	3,57	-99,77591	0,008	0,56	-98,57142857
bfammn24-96-MIC	SUJ	-0,71	1,27	-155,90551	-0,71	0,19	-473,6842105
bpubcv04-308-ANT	SUJ	-0,29	0,05	-680	-0,29	0,3	-196,6666667

Tabela 15: resultado das análises acústicas da forma *você* em posição pré-verbal no *corpus* C-ORAL-BRASIL.

O modelo da tabela é o mesmo utilizado anteriormente. Observando a coluna da proporção da diferença entre os *z-scores* da sílaba e aqueles da média normalizada em relação à unidade tonal, pode-se perceber que, à exceção de um caso, que será devidamente comentado, todos os resultados são negativos.

Passemos agora a alguns exemplos.

- (78) bfamcv19-49-MAE: *PAI: [44] <hoje> eu passei perto daquea loja que abriu ali de retalho / mas tem cada / &he / <cada pano> bonito / lá hhh // *RAQ: [45] <retalho bom> // *PAI: [46] aí eu me lembrei daqueles de [2] aque' tipo de chita / aquele enfiadozinho que fazia / blusas antigamente // *MAE: [47] ali perto do EPA // *PAI: [48] é // *MAE: [49] **pois é / mas ali você acha mais é malha** // *MAE: [50] eles tão com mais é com malha //



Figura 46: tela do Praat com o texto (bfamcv19-49-MAE): foi evidenciada a forma “você”.

O valor encontrado da duração bruta da sílaba [‘se] é de 161 ms.

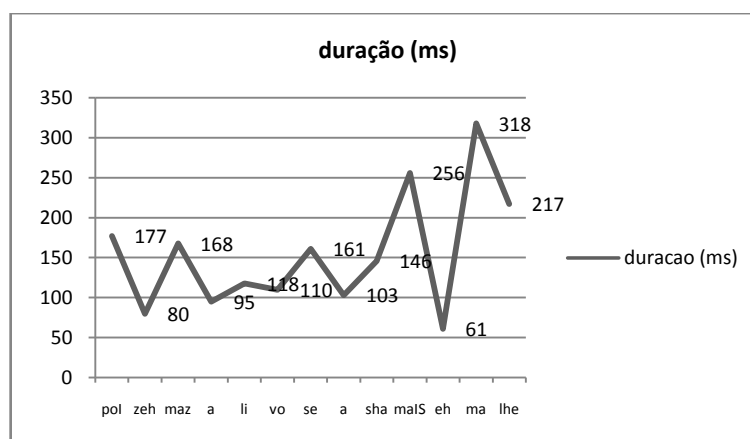


Gráfico 59: duração das sílabas fonéticas do texto (bfamcv19-49-MAE).

Após estabelecida a duração normalizada, de -1,48 de desvio padrão, foi calculada a proporção da diferença em relação à média da respectiva unidade tonal: o resultado foi 48% negativos, portanto a sílaba [se] é menor que a média das sílabas da unidade tonal.

(79) bfamd111-70-MAR: *MAR: [64] né // [65] então ele gosta // [66] aí / depois que tá prontinho ali / você / abre a massa / &n [1] eu costumo abrir ela num plástico / porque aí fica mais fácil d' ocê virar no [/1] no [/1] no seu + [67] <chama mesmo> // *TAT: [68] <pirex> // *MAR: [69] no seu pirex // [70] **do tamanho que você quiser assim** // [71] nũ pode ser tão pequenininho // [72] né // [73] porque / vai dar uma [/1] uma massa boa // [74] aí cê vai virar uma parte dessa massa / ali / vai arrumar direitim / vai vim com esse recheio / pôr ali por cima / certo //

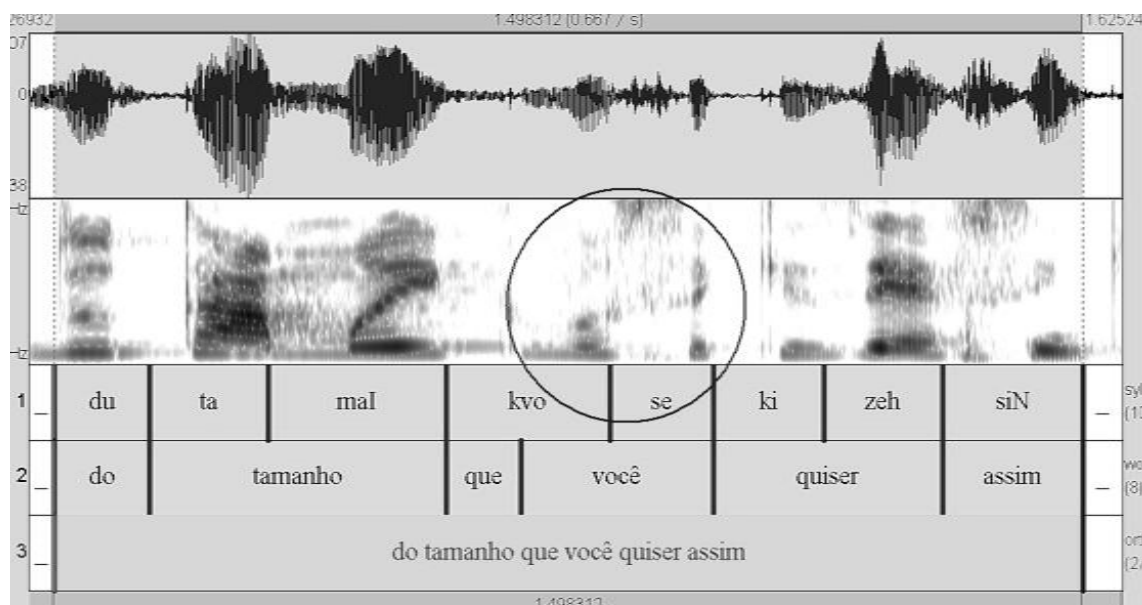


Figura 47: tela do Praat com o texto (bfamd111-70-MAR): foi evidenciada a forma “você”.

Novamente foram extraídas as durações silábicas brutas: aquela de ['se] é de 150 ms.

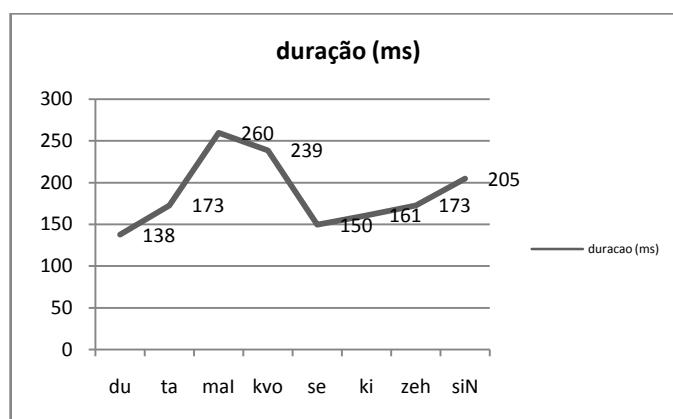


Gráfico 60: duração das sílabas fonéticas do texto (bfamd11-70-MAR).

A normalização da duração resultou em -1,82 de desvio padrão. O cálculo da proporção da diferença em relação à média da unidade tonal resultou novamente em um número negativo, -771%: ['se] seria, portanto, menor que a média das sílabas da unidade tonal.

(80) bfamd114-25-CAR: *CAR: [17] *como / eu acho que eles manobram / as edições do Big Brother // [18] porque / a gente assistindo pelo pay-per-view / *MAU: [19] é outra coisa // *CAR: [20] / é outra coisa // [21] cê vê o negócio limpo / direto / sem edição / é meio confuso / porque / &q [1] eu tô falando / cê tá falando / ele tá falando lá / <às vezes> você nũ <seleciona o> [1] o [1] o [1] o [1] quem você quer ouvir / *MAU: [22] <hum hum> // [23] <o> + [24] certo // *CAR: [25] / **e aí você ouve mais ou menos umas conversa truncada / né** // [26] agora / quando eles mandam pro ar / e a grande maioria que vota / é a pessoal que assiste &he [1] a [1] o [1] o programa <editado> / aí eles direcionam pra aquele que eles querem que <ganhe> //*

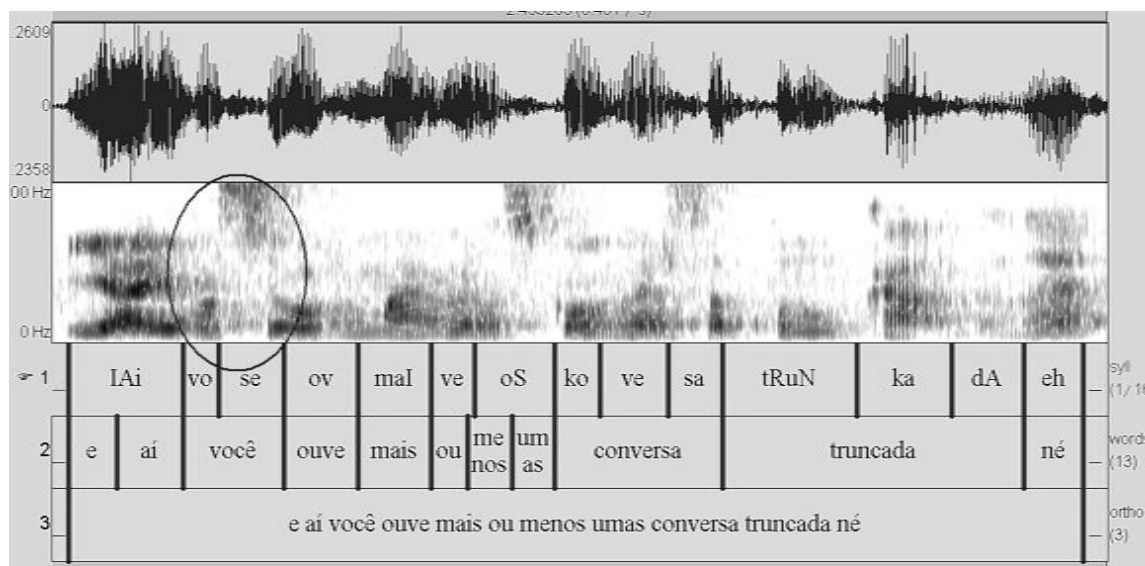


Figura 48: Tela do Praat com o texto (bfamd114-25-CAR): foi evidenciada a forma “você”.

Extraímos as medidas de duração bruta, sendo aquela de ['se] de 155 ms.

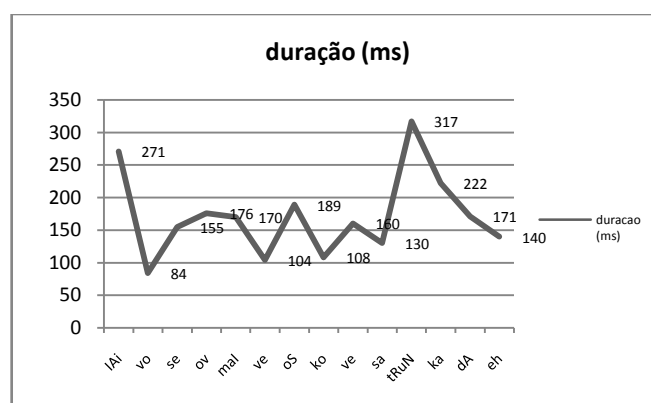


Gráfico 61: duração das sílabas fonéticas do texto (bfamd114-25-CAR).

Os cálculos feitos pelo *SGDetector* (BARBOSA P.A., 2006) resultaram em um valor de [‘se] normalizado de -1,68 de desvio padrão. Conferindo na tabela acima, pode-se observar que o cálculo da proporção da diferença em relação à média da unidade tonal é mais uma vez negativo, -167%, confirmando a atonicidade da sílaba fonética.

Antes de passar para o exemplo que não satisfaz a hipótese, vamos mostrar um caso em que a informante não pronuncia a vogal [o] em “você”, em mais um caso do que talvez seja uma tendência ao apagamento vocálico. No exemplo a forma bisilábica torna-se silábica.

(81) bfammn05-63-CAR: *CAR: [58] *não falo porque eu acho muito pesado* // [59] *mas eu falo com ela que ela + [60] quando ela tinha dois aninho de idade / ela chegou perto de mim / falou comigo assim / mamãe* // [61] *como é que / eu [1] eu / nasci da senhora* // [62] *eu falei assim / ô filha* // [63] ***você não nasceu do meu ventre*** // [64] *você nasceu do meu coração* // [65] *então você é gerada aqui* //

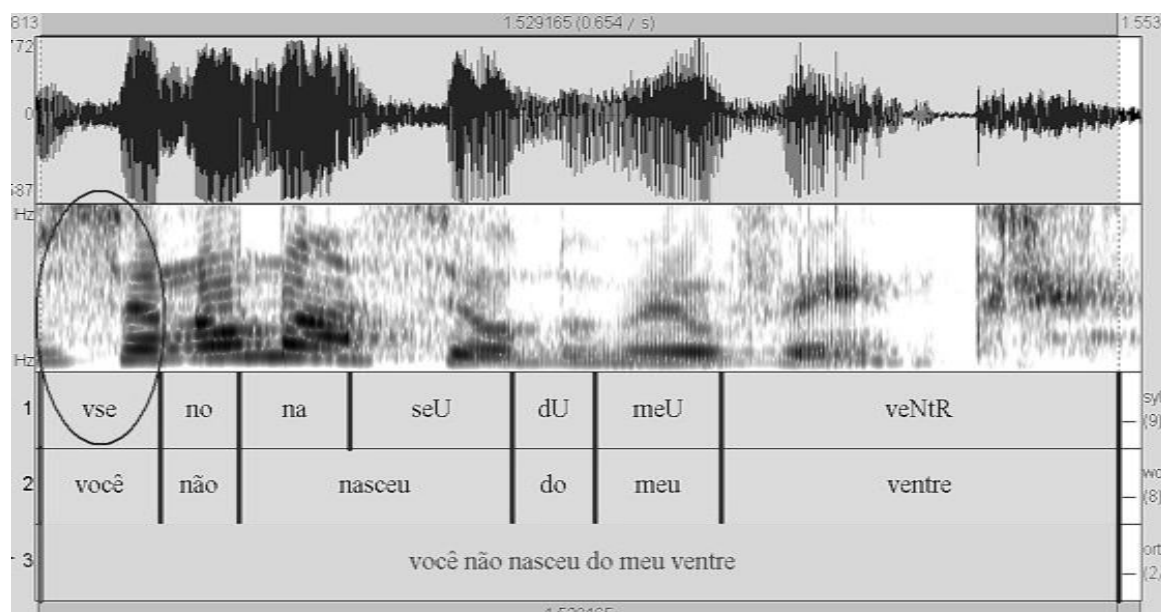


Figura 49: Tela do *Praat* com o texto (bfammn05-63-CAR): foi evidenciada a sílaba [‘vse].

Nota-se pelo espectrograma a ausência dos formantes da vogal [o] e, deste modo, a forma você torna-se [‘vse]. Sua duração bruta é de 171 ms.

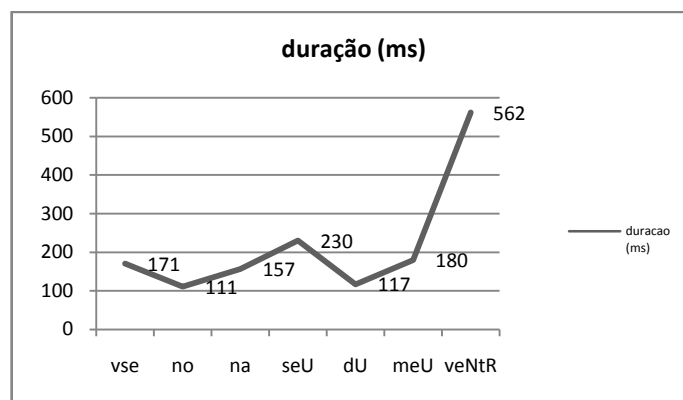


Gráfico 62: duração das sílabas fonéticas do texto (bfamnm05-63-CAR).

Os cálculos de normalização levaram em conta essa sílaba de três segmentos e o resultado foi de -2,55 de desvio padrão. A proporção da diferença em relação à média unidade tonal foi de 771% negativos, portanto a sílaba é menor do que a média das sílabas da unidade tonal.

Vejamos agora o caso cuja proporção resultou positiva.

(82) bfammn12-104-JUN1: [92] *aconteceu tudo / com ele também // [93] nós dois távamos juntos / só que ele / também / começou c' esse projeto / aí / agora / a &gen [2] tipo assim / nós já estamo muito mais amadurecido // [94] aprendemos / o que que é / no [1] no [1] no [1] no + [95] sai pinchando / fazendo vandalismo aí o' hhh // [96] inda mais agora que eu sou de maior // [97] nũ tem mamãe pa me defender mais não // [98] aí / aquele estilo lá / que ele usou / aquilo ali / o' / é o Winnie Style // [99] que é / letras mais / assim / diferentes // *MMM: [100] uhn // [101] <&s [1] meio> distorcido / <né> // *JUN: [102] <tá vendo> // [103] <é> // [104] **se você reparar bem** / você consegue enxergar um esse e um tê ali o' // *MMM: [105] <ham ham> // *JUN: [106] <repara> bem / p' cê ver //*

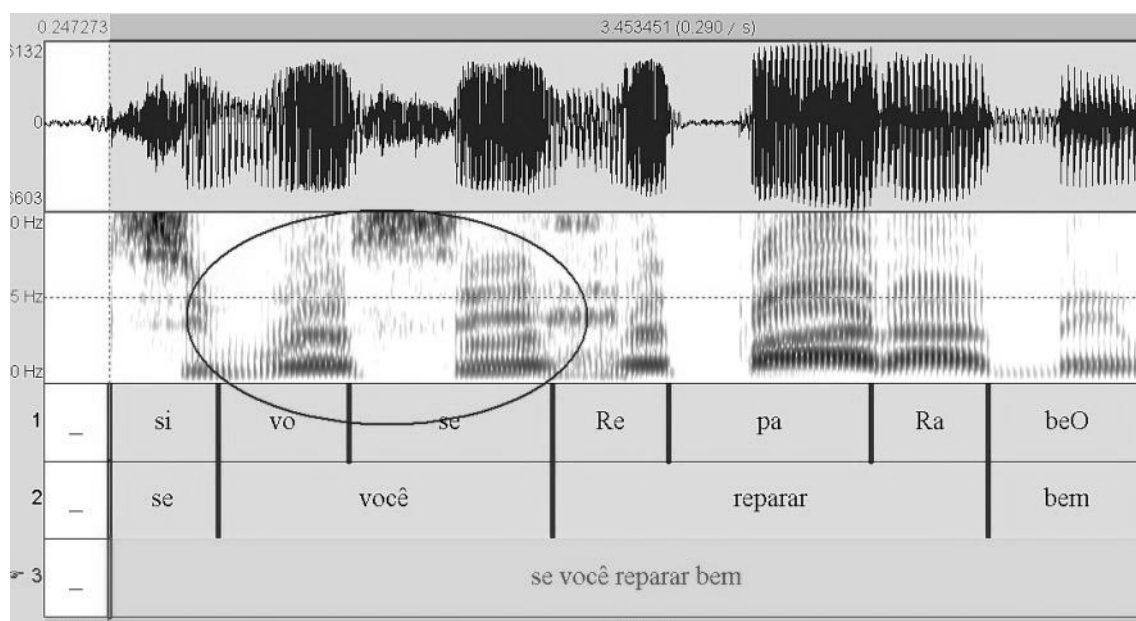


Figura 50: Tela do Praat com o texto (bfammn12-104-JUN1): foi evidenciada a forma “você”.

Foram extraídas as durações silábicas brutas: aquela de ['se] é de 253 ms, bem longa em relação às outras.

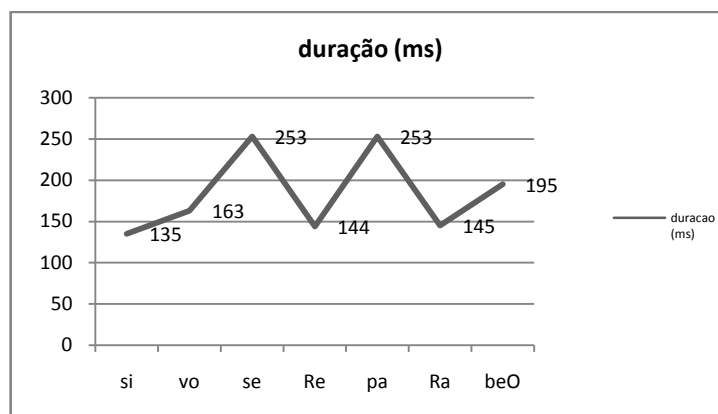


Gráfico 63: duração das sílabas fonéticas do texto (bfammn12-104-JUN1).

A normalização dos dados forneceu um valor de 1,54 de desvio padrão. A comparação com a proporção da diferença em relação à unidade tonal resultou desta vez em um número positivo, 2667%: ['se] seria, portanto, maior que a média das sílabas da unidade tonal.

Uma possível explicação para este caso é que a forma *você* poderia ter uma motivação pragmática, mas é difícil determiná-lo somente à oitiva, pois trata-se de suposições sobre o que o informante queria pôr em evidência.

4.1.7.6. Forma *você* em posição pós-verbal

Será agora apresentada a forma *você* em posição pós-verbal. A tabela abaixo mostra os resultados gerais das análises.

TEXTO	FUNÇÃO SINTÁTICA	DURAÇÃO SIL Z/ EN		PROP ENUNC	DURAÇÃO SIL Z/ UT		PROP UT
bfamcv18-191-HER	OBJ PREP	1,34	-0,8	267,5	1,34	-0,8	267,5
bfamd111-126-MAR	OBJ PREP	4,12	-1,57	362,420382	4,12	-1,57	362,4203822
bfamd133-107-JAN	OBJ PREP	0,41	-1,13	136,283186	0,41	-0,61	167,2131148
bfammn05-1-CAR	OBJ PREP	1,6	-0,15	1166,66667	1,6	-0,08	2100
bfammn07-146-NAN	OBJ PREP	3,18	-0,48	762,5	3,18	-0,47	776,5957447
bfammn16-67-SUE	OBJ PREP	8,13	-0,94	964,893617	8,13	-0,49	1759,183673
bfammn24-53-MIC	OBJ PREP	0,36	0,6	-40	0,36	0,2	80
bfammn35-295-JMA	OBJ DIR	0,23	-0,59	138,983051	0,23	-0,59	138,9830508
bpubcv04-217-ANT	OBJ PREP	0,85	-0,43	297,674419	0,85	-1	185
bpubcv08-98-LEO	OBJ PREP	-0,15	-0,47	68,0851064	-0,15	-0,86	82,55813953
bpubdl06-17-TAT	OBJ PREP	3,12	0,53	488,679245	3,12	0,53	488,6792453
bpubdl06-173-TAT	OBJ PREP	0,75	-0,6	225	0,75	-0,6	225
bpubdl07-596-JAD	OBJ PREP	-1,61	-0,9	-78,888889	-1,61	-2,05	21,46341463

Tabela 16: resultado das análises acústicas da forma *você* em posição pós-verbal no *corpus* C-ORAL-BRASIL.

Foram estudados 13 casos: do ponto de vista sintático há um exemplo de objeto direto, enquanto os outros são de objeto indireto. Quanto à análise acústica, todos os exemplos mostraram resultados da proporção da diferença em relação à unidade tonal positivos: ['se] seria, portanto, maior que a média das sílabas da unidade tonal a que pertence, confirmando nossa hipótese.

Seguem alguns casos.

(83) bfamdl11-126-MAR: *TAT: [124] *então cê tem que dividir a massa // *MAR: [125] é // [126] por isso que eu falei com você // [127] a massa divide em dois // [128] do mesmo tamanho / porque o tamanho do pirex / é que você vai &p [2] vai tampar o [1] o [1] o empadão depois //*

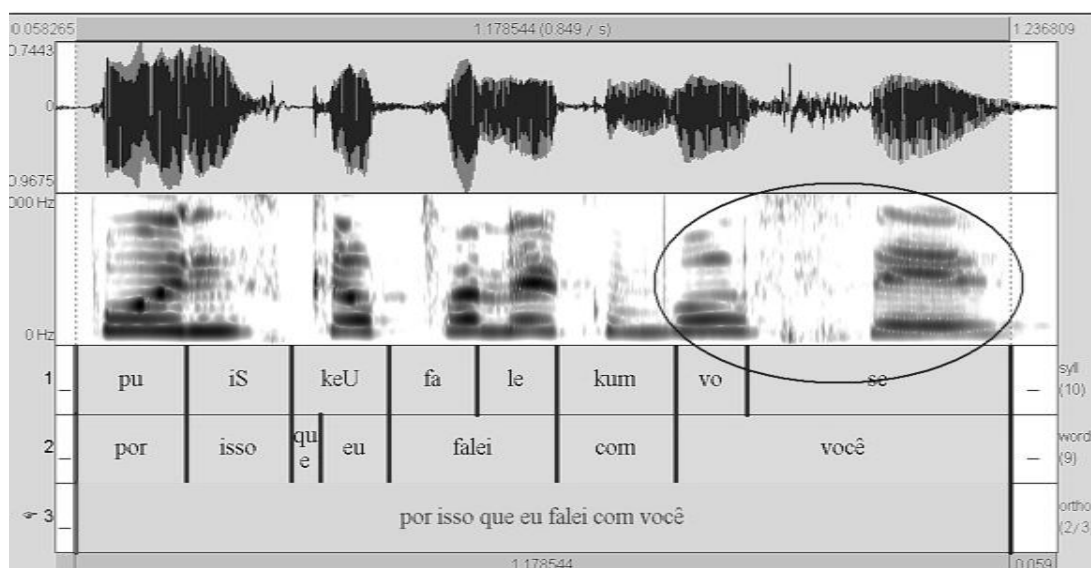


Figura 51: Tela do *Praat* com o texto (bfamdl11-126-MAR): foi evidenciada a forma “você”.

Foram extraídas as durações silábicas brutas: aquela de ['se] é de 332 ms, muito maior que a das outras sílabas fonéticas, talvez também devido à posição em final de enunciado.

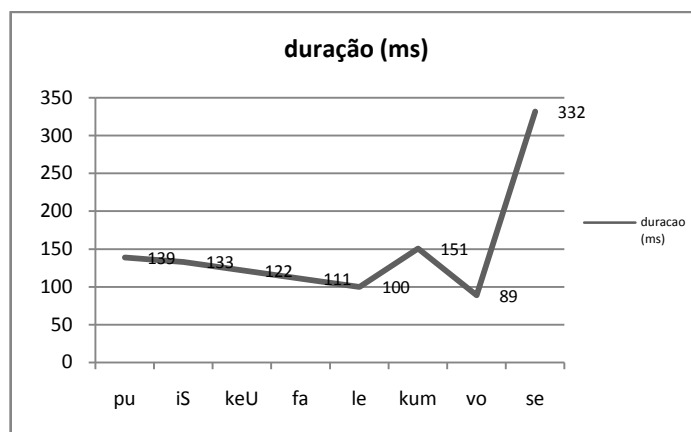


Gráfico 64: duração das sílabas fonéticas do texto (bfamd111-126-MAR).

A normalização dos dados reverteu este valor para 4,12 de desvio padrão. Calculou-se a proporção da diferença em relação à respectiva unidade tonal: o resultado foi 362%, portanto a sílaba ['se] é maior que a média das sílabas da unidade tonal.

(84) bfammn05-1-CAR: *CAR: [1] *bom / &m /1 primeiro lugar / eu vou dizer pra você é uma /1 uma história triste / Mara // [2] &um /1 pra /1 para iniciar / a /1 pra mim iniciar / a /1 a / adotar /2 a adoção da minha filha // [3] porque primeiro eu tive que /3 conteceu uma tragédia comigo que tive que perder minha filha / que é um retratinho que tá ali na estante / comigo / ela abraçada em mim / pra / eu / &he / chegar a essa conclusão pa mim pegar uma filhinha pa mim criar //*

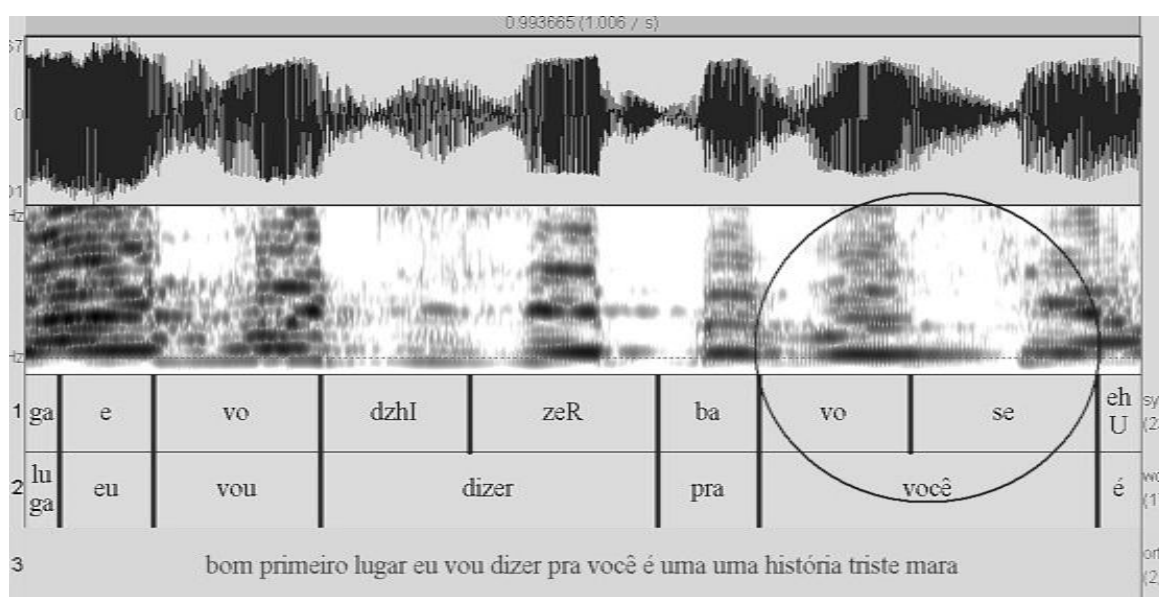


Figura 52: Tela do Praat com o texto (bfammn05-1-CAR): foi evidenciada a forma “você”.

O enunciado é bastante longo e subdividido em várias unidades tonais. O valor da duração bruta da sílaba ['se] é de 154 ms.

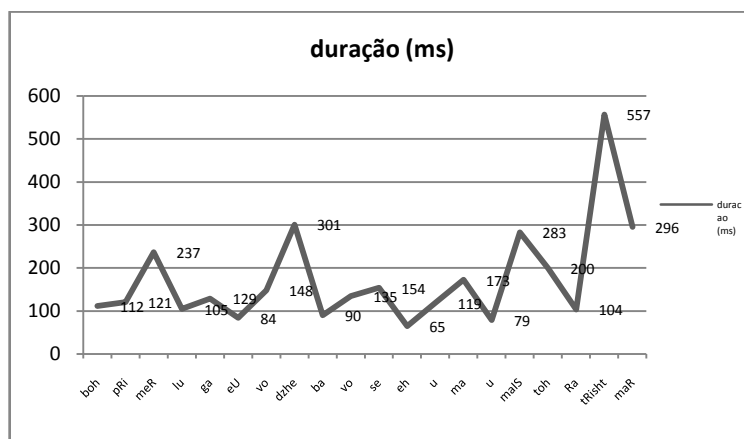


Gráfico 65: duração das sílabas fonéticas do texto (bfamnm05-1-CAR).

A operação de normalização levou ao valor de ['se] de 1,6 de desvio padrão. O cálculo da proporção da diferença em relação à respectiva unidade tonal resultou em 2100%, portanto a sílaba ['se] é maior que a média das sílabas da unidade tonal.

(85) bfammn07-146-NAN: *NAN: [133] sabe hhh // [134] oi / Edgar // [135] aí / o Edgar / ô / Nandinho / é o Edgar // [136] eu tô te ligando / porque me deu muita saudade de você // [137] sabe onde que eu tô // [138] eu falei / aonde / Edgar // [139] eu estou na bica da fazenda // [140] e eu desabei a chorar // [141] isso me emociona até agora // [142] mas eu chorava copiosamente hhh / sabe // [143] eu tô na bica aqui da fazenda // [144] é um lugar que eu &go [/] adorava ir // [145] perto do moinho // [146] eu lembrei de você aqui // [147] era o meu primo mais novo //

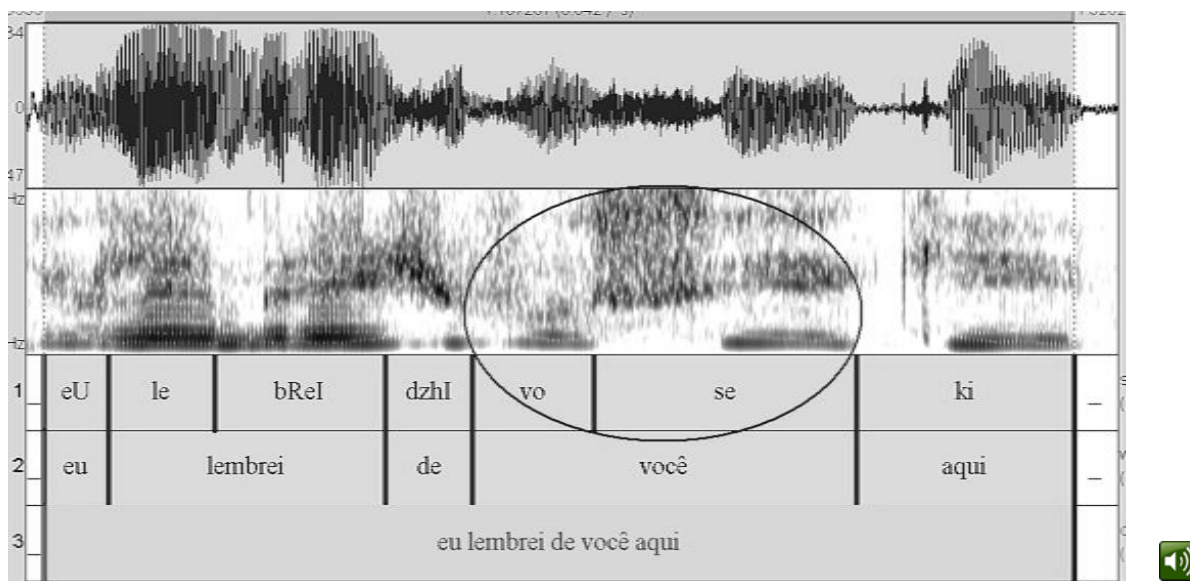


Figura 53: Tela do Praat com o texto (bfammn07-146-NAN): foi evidenciada a forma “você”.

A duração bruta da sílaba [’se] é de 303 ms, a mais longa, como se pode observar pelo gráfico abaixo, apesar de não ser em posição de final de enunciado.

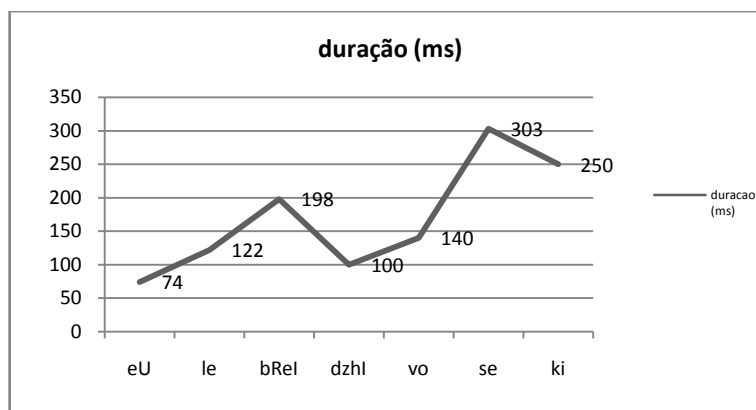


Gráfico 66: duração das sílabas fonéticas do texto (bfammn07-146-NAN).

Após a normalização dos dados esse valor foi revertido para 3,18 de desvio padrão. Calculou-se a proporção da diferença em relação à respectiva unidade tonal: o resultado foi 776%, portanto a sílaba [ˈse] é maior que a média das sílabas da unidade tonal.

4.1.8. Validação estatística

A estatística inferencial foi utilizada para corroborar inferencialmente os resultados encontrados na pesquisa¹³⁴. O objetivo era verificar se há diferença significativa na duração entre a sílaba [ˈse] em posição pré e pós-verbal. Os testes foram efetuados em relação ao enunciado e à unidade tonal.

Foram carregados no programa *R* (cf. <http://www.r-project.org/>) todos os dados analisados e calcularam-se a média, mediana e desvio-padrão das três formas. Em seguida foi efetuada a rodada dos testes de normalidade.

Em ambos os casos o teste de normalidade *Shapiro-Wilk* (cf. SHAPIRO-WILK, 1965; ROYSTON, 1982) apontou que os dados não apresentaram uma distribuição normal.

O *Shapiro-Wilk* (cf. SHAPIRO-WILK, 1965; ROYSTON, 1982), que mediu, com um nível de significância de 5%, a normalidade dos dados da duração de [ˈse] em relação ao enunciado, retornou que a distribuição não é normal. Os valores obtidos são:

p-value encontrado no *shapiro* teste pré-verbal: 3.995e-07

p-value encontrado no *shapiro* teste pós-verbal: 2.814e-12

O *p-value* no *shapiro.test* é um valor que determina ou não se a variável tem distribuição normal. Se for menor que 0.05, não há distribuição normal. Se for maior que 0.05, apresenta distribuição normal. No caso aqui apresentado ambos os *p-values* são extremamente pequenos e menores que 0.05. Logo, não são estatisticamente normais. É possível observar isso claramente também através dos gráficos de densidade.

¹³⁴Agradecemos a preciosa ajuda de Bárbara Helohá pelos cálculos estatísticos e de Marcelo Vieira pela orientação sobre os testes mais adequados a serem utilizados.

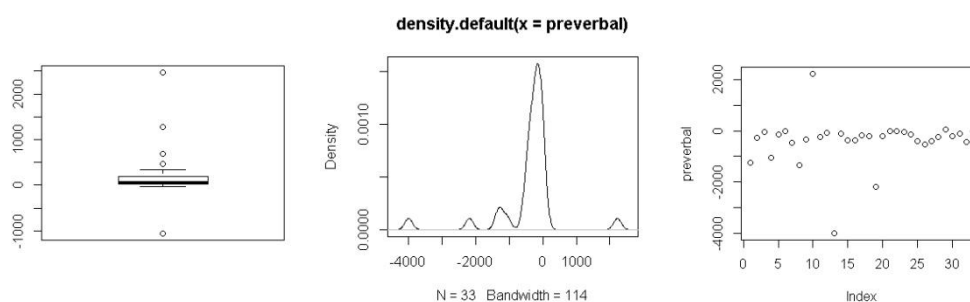


Figura 54: Resultados dos testes de normalidade de [se] em relação ao enunciado em posição pré-verbal: o primeiro é o gráfico *boxcox*, que mostra os dados fora do padrão; o segundo é de densidade e o terceiro é um gráfico de dispersão dos dados, que mostra simultaneamente duas variáveis. No caso, duração e forma pré-verbal ou pós-verbal. Nota-se que a amostra exibida não possui distribuição normal.

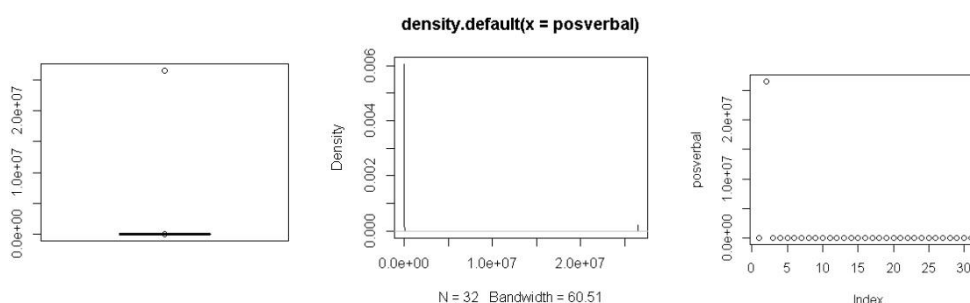


Figura 55: Resultados dos testes de normalidade de [se] em relação ao enunciado em posição pós-verbal: nota-se que a amostra exibe não possui distribuição normal.

Também o segundo teste de hipótese retornou, com um nível de significância de 5% que a duração de [se] em relação à unidade tonal não apresenta distribuição normal. Os valores obtidos são:

- *p-value* encontrado no *shapiro* teste pré-verbal: $9.543e-10$ ¹³⁵;
- *p-value* encontrado no *shapiro* teste pós-verbal: $1.165e-07$

No caso aqui apresentado ambos os *p-values* acima são extremamente pequenos e menores que 0.05. É possível observar isso também através dos gráficos de densidade.

¹³⁵ Os valores, por serem extremamente pequenos são expressos em potência: este número significa: $9,543 \times 10^{-10} = 0,0000000009543$

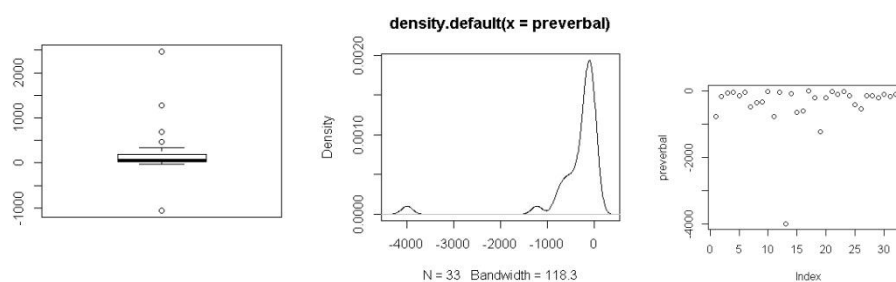


Figura 56: Resultados dos testes de normalidade de ['se] em relação à unidade tonal em posição pré-verbal: nota-se que a amostra exibe não possui distribuição normal.

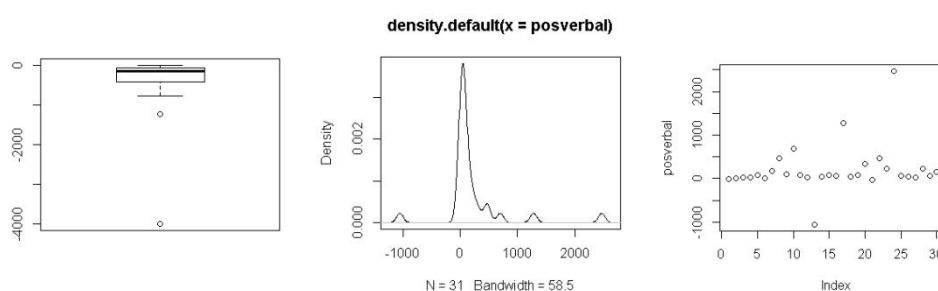


Figura 57: Resultados dos testes de normalidade de ['se] em relação à unidade tonal em posição pós-verbal: nota-se que a amostra exibe não possui distribuição normal.

Tendo em vista a distribuição não normal da amostra, optou-se por não utilizar o *teste t de Student*, mas aquele de *Wilcoxon-Mann-Whitney* (cf. SPRENT & SMEETON, 2001; CONOVER, 1999 entre outros) que verifica se duas amostragens estatísticas provêm da mesma população. O *teste t de Student* é baseado na média. Já o *Wilcoxon-Mann-Whitney* é destinado a dados sem distribuição normal e baseado na mediana, que é menos tendenciosa.

Os resultados obtidos para a comparação entre ['se], pré-verbal e pós-verbal, e o enunciado foram os seguintes:

- *p-value* encontrado no *wilcox.test* com correção para distribuição normal: 6.181e-08;
- *p-value* encontrado no *wilcox.test* sem correção para distribuição normal: 6.181e-08.

A correção torna o teste mais rigoroso. Contudo, não houve alteração no resultado. Isso significa que o resultado obtido, com e sem correção, foi o mesmo. Os valores de *p-value* 6.181e-08 são aqueles que irão determinar o resultado do teste: se forem maiores que 0.05, significa que as duas variáveis (pré e pós-verbal) são iguais. Neste caso específico, as diferenças na duração de ['se] em relação ao enunciado encontradas na estatística descritiva seriam devidas à má coleta da amostra. Se o *p-value* for menor que 0.05, significa que as duas variáveis (pré e pós-verbal) são diferentes. Os valores encontrados são bem menores que 0.05.

O teste, portanto, é estatisticamente significativo, pois os valores estão bem abaixo de 0.05. Em outras palavras, pode-se afirmar que, com um nível de significância de 5%, as durações de ['se] nas duas posições (pré e pós-verbal) são distintas estatisticamente. O teste utilizado foi unilateral à esquerda (comando "less"). Este comando testou se a variante pré-verbal era mais curta do que a pós-verbal. Visto que o *p-value* resultou abaixo de 0.05, com 5% de significância, em relação ao enunciado pode-se afirmar que a duração de ['se] pré-verbal é menor que ['se] pós-verbal.

O mesmo teste foi efetuado em relação à unidade tonal. Os resultados obtidos para a comparação entre ['se], pré-verbal e pós-verbal, e a unidade tonal foram os seguintes:

- *p-value* encontrado no *wilcox.test* com correção para distribuição normal: 1.36e-10;
- *p-value* encontrado no *wilcox.test* sem correção para distribuição normal: 1.36e-10.

A correção torna o teste mais rigoroso. Contudo, também neste caso, não houve alteração no resultado, sendo o mesmo com e sem correção.

Novamente é o *p-value* que determinará o resultado do teste: sendo maior que 0.05, significa que as duas variáveis (pré e pós-verbal) são iguais. Eventuais diferenças na duração de ['se] em relação à unidade tonal encontradas na estatística descritiva seriam devidas à má coleta da amostra. No caso de o *p-value* ser menor que 0.05, as duas variáveis (pré e pós-verbal) seriam diferentes. Os valores encontrados são bem menores que 0.05.

Novamente o teste é estatisticamente significativo, pois os valores estão bem abaixo de 0.05. Pode-se, portanto, afirmar que, com um nível de significância de 5%, as durações de ['se] em relação à Unidade Tonal, nas duas posições (pré e pós-verbal), são distintas. O teste utilizado foi unilateral à esquerda (comando "less"). Este comando testou se a variante pré-verbal era mais curta do que a pós-verbal. Como o *p-value* foi abaixo de 0.05, com 5% de significância, em relação à unidade tonal ['se] pré-verbal é menor que ['se] pós-verbal.

Por fim, queremos esclarecer que tínhamos previsto outros testes estatísticos que medissem a diferença entre as formas *cê* e *ocê/você*. Contudo, dado o tamanho reduzido da amostra, e visto que este primeiro teste confirmou nossa hipótese, decidimos aprofundar a análise estatística para um trabalho futuro, que possua um maior número de dados a serem verificados.

Nesta seção apresentamos as análises gerais e acústicas realizadas a partir dos dados obtidos no corpus C-ORAL-BRASIL. Nas próximas páginas apresentaremos as análises do vernáculo florentino, para em seguida retomarmos as discussões e considerações finais sobre ambas as partes da pesquisa.

4.2. Resultados do vernáculo florentino

Nas seções 2.4.4 e 2.4.5, foram relatados os critérios e métodos de seleção dos verbos pesquisados para a identificação do modo de expressão do sujeito no florentino. Apresentaremos agora os resultados obtidos.

4.2.1. Verbo *essere*

Foram analisadas 1.977 ocorrências da forma verbal *è*, terceira pessoa do singular do indicativo presente, no *Corpus Stammerjohann*. Destas, 1.136 resultaram serem verbo não auxiliar. Já no *Corpus de Florentinos* foram encontradas um total de 1.347 casos da forma: destes 697 são como verbo não auxiliar.

A análise dos dados partiu dos resultados fornecidos utilizando o *software AntConc* que, como visto, busca automaticamente nos *corpora* a forma pesquisada fornecendo seu relativo contexto. A tabela abaixo mostra uma visão geral do que foi encontrado. Como o total de palavras de cada *corpus* é diferente, para poder compará-los, os dados foram normalizados por 100.000: a coluna da esquerda mostra os valores absolutos enquanto aquela da direita, os valores normalizados.

Formas	CORPUS STAMMERJOHANN		CORPUS DE FLORENTINOS	
	Ocorrências absolutas	Ocorrências Normalizadas	Ocorrências absolutas	Ocorrências Normalizadas
TOTAL de ocorrências de <i>è</i> como verbo autônomo	1136	1317	697	1196
<i>L'</i>	238	276	133	228
<i>Gl'</i>	44	51	39	67
<i>La</i>	2	2	1	2
<i>E</i>	0	0	0	0
Reduplicação clítica <i>e l'</i>	2	2	0	0
Reduplicação clítica <i>e gl'</i>	2	2	0	0
Reduplicação clítica <i>la l'</i>	1	1	0	0
Sujeito \emptyset	382	443	244	419
Sujeito lexical pré-verbal	187	217	129	221
Sujeito lexical pós-verbal	118	137	93	160
Sujeito Pronome Demonstrativo	158	183	129	221
Pronome tônico	17	20	12	21

Tabela 17: Formas de expressão do sujeito encontradas nos dois *corpora* comparados em valores absolutos e normalizados.

Abaixo fornecemos ao leitor, para cada tipo de sujeito, dados extraídos dos *corpora*.

▪ Sujeito *l'* clítico:

(86) *ART: [...] // [431] / no / **l'** è / hhh Castiglion della Pescaia //



não/ ele (clít.) é Castiglion della Pescaia

(Corpus Stammerjohann, arti)¹³⁶

(87) *LUC: [392] [<] <ma bellina> / guarda // [393] vedessi come **l'** è bellina //



mas lindinha olha (você) teria que ver como ela (clít.) é lindinha

(Corpus Confronto Diacronico, fam1)

▪ Sujeito *gl'* clítico:

(88) *ART: [469] non so se **gl'** è Porto Santo Stefano //



não sei se ele (clít.) é Porto Santo Stefano

(Corpus Stammerjohann, arti)

(89) *LOR: [279] <eh> // [280] ho vinto / ora **gl'** è poco //



eh, ganhei agora ele (clít.) é pouco (tempo)

(Corpus Confronto Diacronico, ipubdl03)

▪ Sujeito *la* clítico:

(90) *CL2: [164] e poi lo sai icché c' è ? c' è la &zia [///] c' è che **l'** è tedesca // [165]

ecco // [166] **l'** è [///] **la** non è italiana //



e depois o sabes o que há? há que a tia (interrompido), há que ela (clít.) é alemã isso ela (clít.) é (interrupção) ela (clít.) não é italiana

(Corpus Stammerjohann, barb02)

¹³⁶ Os símbolos adotados nos *corpora* seguem o padrão da família C-ORAL-ROM: entre parênteses indicamos o *corpus* e a sigla do texto de proveniência do exemplo. Como já foi mostrado, a barra simples (/) indica início e final de unidade tonal; a barra dupla (//) a fronteira prosódica de final de enunciado. O símbolo (+) mostra que o enunciado foi interrompido; palavras entre colchetes (<>) indicam sobreposições de fala; os números entre colchetes quadrados [] significam o número do enunciado; três letras maiúsculas após o asterisco a sigla do informante; o símbolo ([/no]) indica *retracting* ou erro na execução do enunciado; o símbolo (&), início de palavra interrompida; (&he) é usado no caso de uma hesitação; (hhh) quando há um fenômeno paralinguístico. Para evitar a citação de enunciados demasiadamente longos, quando recortamos alguns trechos utilizamos o símbolo [...].

(91) *CLE: [...] // [42] *sì* // [43] *perché la 'un è roba mia* //



sim porque ela (clít.) não (clít.neg.) é coisa minha

(*Corpus Confronto Diacronico*, ipubdl01b)

- Reduplicação do sujeito clítico *e l'*¹³⁷:

(92) *MOG [...] // [249] *perché / lui / e l' è un ragazzo tedesco / no / Giorgio Ottmar* // ¹³⁸



porque ele (tônico) ele (clít.) ele (clít.) é um rapaz alemão não Giorgio Ottmar

(*Corpus Stammerjohann*, fami1)

- Reduplicação do sujeito *e gl'*¹³⁹:

(93) *VIT: [14] *sì sì* // [15] *e / gl' è bellino / vai* // [16] <*guarda*> //



sim sim ele (clít.) ele (clít) é bonitinho veja olha

(*Corpus Stammerjohann*, vita1)

- Reduplicação do sujeito clítico *la l'*¹⁴⁰:

(94) *GIO: [32] *sta qui / no ?* [33] *sicchè / la l' è più <l' è più / turbata> degli altri / voglio dire* //



mora aqui /não? então ela (clit.) ela (clit.) é mais ela (clit) é mais transtornada que os outros quero dizer

(*Corpus Stammerjohann*, port)

- Sujeito Ø:

(95) *BRA: [...] // [158] *lo impiccano / però / siccome Ø è di legno /*



o enforcam mas como é Ø de madeira

(*Corpus Stammerjohann*, pino)

¹³⁷ Não foram encontrados exemplos deste tipo de duplicação no *Corpus di Confronto Diacronico*, portanto citamos somente o exemplo do *Corpus Stammerjohann*.

¹³⁸ Este exemplo parece-nos especialmente interessante, pois o sujeito foi expresso quatro vezes: com o tônico (*lui*), com um SN (*Giorgio Ottmar*, sujeito posposto), e com os dois clíticos *e gl'*.

¹³⁹ Neste caso também há exemplos somente no *Corpus Stammerjohann*.

¹⁴⁰ *Idem*.

- Sujeito lexical pré-verbal:

(96) *AND: [864] *mah / senti / **il francese** è più facile / eh / dell' inglese //*

bom ouça o francês é mais fácil eh do que o inglês



(Corpus Stammerjohann, fran)

(97) *GUG: [263] **Salvo** / è di' quarantacinque //

Salvo é de quarenta e cinco (do ano de 1945)



(Corpus Confronto Diacronico, fam1)

- Sujeito lexical pós-verbal:

(98) *AMI: [150] *è contento **il gatto** / eh / che [/] che [/] che il lupo sia morto //*

está feliz o gato eh que que o lobo (sujeito) tenha morrido



(Corpus Stammerjohann, vita2)

(99) *VAL: [406] certo // [407] 'un è mica sposata / **l' Ornella** //

claro, não (clit. neg.) é nada casada a Ornella (sujeito)



(Corpus Confronto Diacronico, fam1)

- Sujeito pronominal de outro tipo:

(100) *FRA: [859] **questa** è una professoressa / sai //

esta (pronomo demonstrativo) é uma professora sabe



(Corpus Stammerjohann, fran)

(101) *NIL: [116] **quella** è la procedura che stabilisce il giudice //

aquele (pron. demonstrativo) é o procedimento que estabelece o juiz



(Corpus Confronto Diacronico, ifamnn09)

- Pronome tônico sujeito:

(102) *RAG: [686] *perché / &lu [/] lui / guarda / è in mezzo a tutti //*

Porque ele (tônico) olha está no meio de todos



(Corpus Stammerjohann, arti)

(103) *TIZ: [128] magari lei è una femmina //

talvez ela (tônico) seja uma menina



(Corpus Confronto Diacronico, ifamd108)

Para maior clareza analisaremos daqui em diante os dados normalizados, facilmente comparáveis. O *Corpus Stammerjohann* apresenta um total de 1.317 palavras (a cada 100.000) da forma *è*, sendo o verbo não auxiliar, mas possuindo valor copular, identificativo ou predicativo com base locativa. No *Corpus di Fiorentini* as ocorrências são 1196.

A diminuição das formas clíticas, quanto à expressão do sujeito e baseando-nos nos dados, aparentemente não ocorreu: o clítico *l'* foi encontrado 276 vezes no *Corpus Stammerjohann* e 228 no *Corpus di Fiorentini*. Uma redução de 17%, menos intensa do que esperávamos. No caso de *gl'* houve um aumento: das 51 ocorrências no *Corpus Stammerjohann*, foram encontradas 67 naquele *di Fiorentini*. Quanto às formas *la* e *e* não houve mudanças, mas ambos os *corpora* apresentaram um número de exemplos de *la* muito baixo, dois para cada. A forma *e* não foi encontrada nem no *Corpus Stammerjohann*, confirmando sua quase extinção no segundo pós-guerra, fato observado por vários estudiosos (cf. RENZI, 1983 entre outros).

A análise averiguou também os casos de reduplicação do sujeito clítico. Foi constatada uma queda, contudo a amostra é muito pequena. No *Corpus Stammerjohann* foram encontradas duas ocorrências para as reduplicações *e l'*, *e gl'* e uma para *la l'*, enquanto no *Corpus di Fiorentini* não houve nenhuma destas reduplicações.

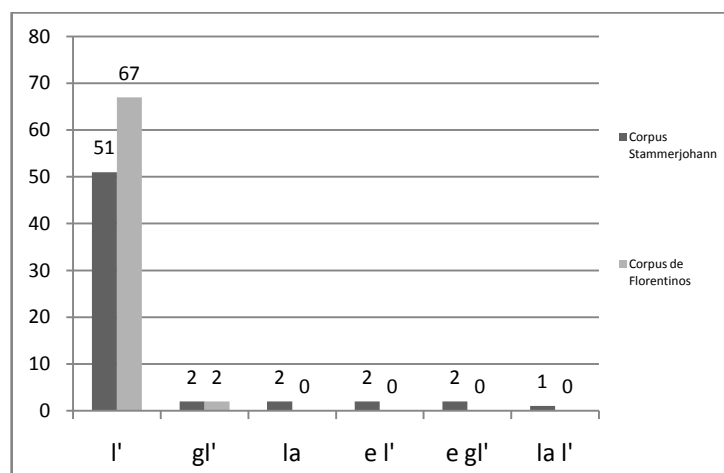


Gráfico 67: Variação no uso das formas de sujeito clítico, simples e reduplicado, nos dois *corpora*, em valores normalizados.

Foram também conferidas as outras maneiras de se expressar o sujeito. A hipótese inicial é que, além da queda dos clíticos no *corpus* mais recente, haveria uma redução geral da expressão do sujeito e um aumento da forma \emptyset . Contudo, observando os dados, passou-se de 443 ocorrências de sujeito \emptyset no *Corpus Stammerjohann* para 419 no *Corpus di Fiorentini*, uma redução, de apenas 5%, muito pequena para nos dizer algo. Quanto à expressão do sujeito lexical e pronominal, houve uma estabilidade muito consistente entre os dois *corpora* e, em alguns casos um ligeiro aumento. Entre os exemplos de sujeito lexical¹⁴¹ pré-verbais, foram encontradas 217 ocorrências no *Corpus Stammerjohann* e 221 naquele *di Fiorentini*: na prática não houve alterações.

Os sujeitos lexicais pós-verbais passaram de 137 casos do *Corpus Stammerjohann* para 160 do *Corpus di Fiorentini*. Isto significa um aumento de 17%. O caso é bastante relevante. Contudo, os sujeitos pós-verbais possuem uma função informativa que responde a um outro tipo de lógica de natureza pragmática. Isto mereceria uma análise qualitativa à parte, mas que não é objeto da presente pesquisa.

Por outro lado, os pronomes demonstrativos (*questo/a, quello/a*: este/a, aquele/a) registraram um aumento de 21%: foram encontradas 183 ocorrências no *Corpus Stammerjohann* e 221 naquele *di Fiorentini*.

Finalizando a relação dos resultados, os pronomes tônicos apresentaram uma variação mínima: sendo 20 no *Corpus Stammerjohann* e 21 no *Corpus di Fiorentini*.

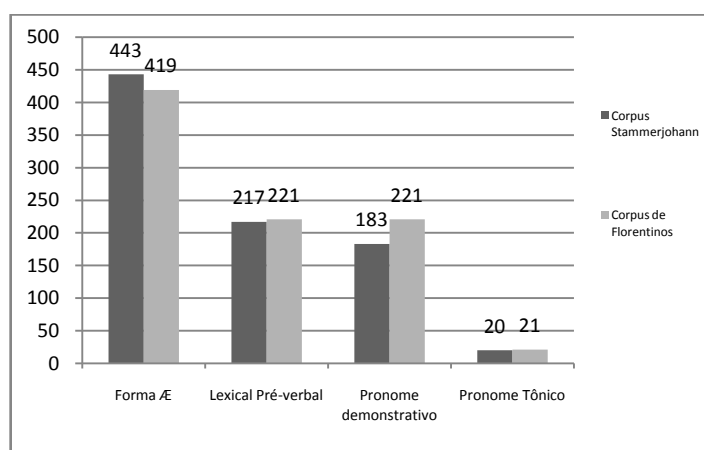


Gráfico 68: Variação no uso das formas de expressão do sujeito, nos dois *corpora*, em valores normalizados.

¹⁴¹ Para simplificar, na contagem dos sujeitos lexicais foram inseridas também as orações subjetivas (11 no *Corpus Stammerjohann* e 4 no *Corpus di Fiorentini*) e os pronomes indefinidos (3 no *Corpus Stammerjohann* e 6 no *Corpus di Fiorentini*).

Pareceria que a redução dos pronomes clíticos sujeito *e*, de forma geral, da expressão do sujeito no vernáculo florentino, não tenha sofrido a tão esperada pressão do italiano padrão.

Contudo, a imagem geral muda quando os dados são distintos com base em diferentes parâmetros e são comparados de forma mais meticulosa. Observemos primeiramente os clíticos:

Forme	<i>Corpus Stammerjohann</i>		<i>Corpus de Florentinos</i>	
	Ocorrências absolutas	Ocorrências Normalizadas	Ocorrências absolutas	Ocorrências Normalizadas
Lexical pré-verbal + <i>l'</i>	0	0	12	21
Lexical pré-verbal + <i>gl'</i>	1	1	1	1
Lexical pós-verbal + <i>l'</i>	1	1	6	10
Lexical pós-verbal + <i>gl'</i>	1	1	1	1
<i>Questo</i> + <i>l'</i>	21	24	4	7
<i>Questo</i> (posposto) + <i>l'</i>	1	1	1	1
<i>Quello</i> + <i>l'</i>	15	17	1	1
<i>Questo</i> + <i>gl'</i>	1	1	1	1
<i>Codesto</i> + <i>l'/gl'</i>	1	1	1	1
<i>Quale</i> + <i>l'</i>	0	0	1	1
<i>Chi</i> + <i>l'</i>	0	0	3	5
<i>Lei</i> + <i>l'</i>	3	3	4	7
<i>Lei/lui</i> + <i>gl'</i>	4	5	0	0
<i>Lei/lui</i> (posposto) + <i>l'</i>	1	1	3	5
<i>Lui</i> (posposto) + <i>gl'</i>	1	1	0	0
TOTAL	51	59	39	67

Tabela 18: Formas de sujeito duplicado encontradas nos dois *corpora* comparados em termos absolutos e normalizados.

Não vamos detalhar demasiadamente a tabela acima, mas evidenciar que os casos de sujeito lexical, pré-verbais e pós-verbais, com retomada pelo clítico sujeito *l'* foram os que mostraram o maior aumento no número de ocorrências. No *Corpus Stammerjohann* há um único exemplo do tipo, enquanto no *Corpus di Fiorentini* foram encontradas 31 ocorrências. Os sujeitos lexicais pós-verbais, justamente pela sua natureza de tipo pragmático, deveriam ser abordados separadamente e analisados de forma mais aprofundada, mas não são objetivo deste trabalho e foram separados dos dados computados.

Por outro lado, a situação se inverte no caso dos pronomes demonstrativos (*questo* e *quello*) com retomada com o sujeito clítico *l'*: no *Corpus Stammerjohann* há 41 ocorrências e no *di Fiorentini* somente 8.

Quanto aos pronomes tônicos retomados por clíticos sujeito, há um ligeiro aumento nos casos no *Corpus di Fiorentini*. Contudo, a amostra é muito pequena para poder formular qualquer tipo de hipótese.

O que parece ter acontecido, pelos dados aqui apresentados, é que houve uma mudança através do tempo, por parte dos falantes, no uso da forma clítica: nos anos de 1960, preferia-se a retomada clítica dos pronomes demonstrativos, enquanto mais recentemente a retomada parece acontecer com os sujeitos lexicais e, às vezes, com os tônicos.

Vamos exemplificar a questão com ocorrências dos *corpora* analisados.

- Pronome demonstrativo com clítico sujeito *l'*:

(104) *CRI: [...] / [77] <quello l' è chiaro> //
aquilo ele (clít.) é claro



(*Corpus Stammerjohann*, cust)

- Sujeito lexical com clítico sujeito *l'*:

(105) *LOR: [...] // [39] e poi / sì / la cravatta l' è poco personale / pe' un omo //
e depois sim a gravata ela (clít.) é pouco pessoal para um homem



(*Corpus Confronto Diacronico*, ifamd103)

4.2.2. Discussão

Quando consideramos o fator idade dos informantes pudemos observar resultados interessantes em cada *corpus* e nos dois, de maneira comparativa.

Como já foi dito, ambos os *corpora* seguem os critérios da família C-ORAL-ROM no que diz respeito à distribuição por idade: a faixa A indica os informantes entre 18 e 25 anos; a faixa B aqueles entre 26 e 39 anos; a faixa C inclui aqueles entre 40 e 60 anos e a faixa D aqueles acima de 60 anos. Quando a idade do informante é desconhecida, utiliza-se o símbolo X. Uma peculiaridade do *Corpus Stammerjohann* é que três dos quatro falantes da faixa A, justamente aqueles que proferem os clíticos sujeito objeto deste estudo, são crianças, um fator muito relevante. Pelo contexto se compreende que a idade deles é de aproximadamente 10

anos, já que dizem frequentar o quinto ano do primário. O outro informante da faixa A não apresenta clíticos em sua fala.

O *Corpus Stammerjohann* compõe-se pelas interações de 59 informantes enquanto o *Corpus di Fiorentini* por aquelas de 72 falantes.

A pesquisa subdividiu todas as formas de expressão do sujeito para cada informante: lexicais pré e pós-verbais, todos os pronomes tônicos e demonstrativos que ocorreram nos *corpora*, além dos clíticos sujeito e as diferentes combinações em reduplicação. Isto gerou resultados bem interessantes.

Findada a contagem, foram excluídas as formas de sujeito \emptyset e aquelas lexicais pós-verbais, motivadas pragmaticamente: desta maneira, foi possível extrair o total de todas as formas de sujeito expesso. Após essa etapa, foram cotejados os casos em que tinha sido usada uma das formas de sujeito clítico pronominal. Calculou-se então a porcentagem dos sujeitos clíticos sobre o total das formas de sujeito expesso. Os resultados estão na tabela abaixo.

	FAIXA A	FAIXA B	FAIXA C	FAIXA D
<i>CORPUS STAMMERJOHANN</i>	38,3%	17%	37,7%	-
<i>CORPUS DI FIORENTINI</i>	11,8%	15,9%	31,6%	58,1%

Tabela 19: total dos sujeitos clíticos encontrados nos dois *corpora* comparados, em porcentagem sobre os sujeitos expressos, divididos por faixas etárias.

No *Corpus Stammerjohann*, sobre o total de todas as possíveis formas de se expressar o sujeito, aquela com clíticos representam 38,3%, na faixa A. Lembramos novamente que todas essas formas são expressas por criança que, por sua vez, são filhos dos informantes da faixa C. Na faixa B os clíticos são menos utilizados: foram encontrados em somente 17% dos casos. A faixa C apresenta uma porcentagem de uso dos clíticos bem parecida com aquela da faixa A (37,7%). Não há informantes da faixa D.

Os resultados levam às seguintes reflexões: poderia ter ocorrido uma evidente redução dos clíticos sujeito na faixa B, que corresponde àqueles nascidos durante a II Guerra Mundial. A faixa A não confirmaria tal impressão pelo simples fato de representar crianças, que reproduzem o uso dos pais (informantes da faixa C), porque esse seria ainda seu único modelo. Podemos supor que, à medida que foram crescendo, tenham sido submetidos a diferentes modelos e tenham seguido ou aumentado a tendência que foi observada para a faixa B.

No *Corpus de Florentinos* parece haver um gradualismo no uso das formas clíticas sujeito: a faixa A utiliza os clíticos em apenas 11,8% dos casos, favorecendo outras formas de expressão do sujeito. Também a faixa B os utiliza pouco, em 15,9% dos casos. Na faixa C averiguamos a utilização das formas com sujeito clítico em 31,6% dos casos, um grande crescimento. Já a faixa D é aquela que mais os utiliza, em 58,1% dos casos. Vale a pena observar que a faixa D do *Corpus de Florentinos* corresponde aproximadamente à faixa B do *Corpus Stammerjohann*, isto é, à daqueles que nasceram durante a II Guerra.

Outros fatores foram levados em conta e, conseqüentemente, foram analisados, a saber, escolaridade e profissão dos informantes. No entanto, nenhum deles evidenciou diferenças no uso da expressão do sujeito tão consistentes como o fator idade.

Quando os dois *corpora* são observados de forma separada, a impressão é que houve uma redução constante no uso dos clíticos: no *Corpus Stammerjohann* a faixa B registra uma forte redução e a faixa A não entraria na contagem por ser composta por crianças, que ainda reproduzem a fala dos pais. Portanto, as faixas a serem analisadas seriam somente duas.

Por outro lado, no *Corpus de Florentinos* haveria uma gradual redução no uso dos clíticos partindo da faixa D, passando pela C e pela B e, em um nível muito mais reduzido, entre a faixa B e a faixa A. Visto desta forma, pareceria que, partindo da geração correspondente à faixa C do *Corpus Stammerjohann* em diante, a perda dos clíticos teria sido constante.

A faixa C do *Corpus Stammerjohann* corresponde de forma aproximada a pessoas nascidas entre 1925 e 1940. Não possuímos dados sobre as gerações anteriores, portanto se a análise dos dados estiver correta, poderíamos concluir que, partindo pelo menos daqueles nascidos por volta de 1940, houve uma progressiva e constante redução dos clíticos, e que tal redução poderia até ter se iniciado antes.

Contudo, observando-se com atenção os dados, a situação não é tão linear. Os falantes da faixa B do *Corpus Stammerjohann* utilizam clíticos somente em 17% dos casos, e os falantes da faixa D do *Corpus de Florentinos*, isto é, aqueles da faixa etária que no *Corpus Stammerjohann* pertenciam à faixa B, os utilizam em 58% dos casos, o que não parece plausível.

Além disso, os informantes da faixa C do *Corpus de Florentinos*, ou seja, os mais jovens entre os falantes da faixa B do *Corpus Stammerjohann*, registram uma porcentagem de uso de clíticos de 31,6%. Também isto precisa ser explicado.

A pesquisa buscou uma resposta a tais incongruências averiguando em detalhes a composição dos dois *corpora*. A hipótese era que os cortes feitos no *Corpus per il Confronto*

Diacronico para manter um conjunto de informantes totalmente florentinos pudesse ter prejudicado a comparabilidade com o *Corpus Stammerjohann* sob outros pontos de vista.

Entretanto, não nos pareceu que as diversas partes dos dois *corpora* apresentem diferenças tão significativas que justifiquem as contradições que apresentamos acima.

A análise buscou respostas então na composição dos informantes, para conferir se algum daqueles do *Corpus de Florentinos* tivesse uma presença excessiva no *corpus*.

De fato, encontramos três informantes com esta característica, dois pertencentes à faixa C e um à D:

*ELA, da faixa C, está presente em 5 textos, com 21 clíticos;

*LUC, da faixa C, está presente em 1 texto, com 18 clíticos;

*LUC, da faixa D, está presente em 3 textos, com 19 clíticos.

Ademais, *ELA e *LUC (que são individuáveis) pertencem ao limite máximo da faixa C.

A etapa seguinte foi, portanto, excluir os três informantes da contagem dos clíticos. Obteve-se um quadro bastante diferente, mas bem mais plausível. Sem estes três falantes a faixa D e a faixa C do *Corpus de Florentinos* apresentam uma porcentagem no uso de clíticos de 17%, basicamente a mesma da faixa B do *Corpus Stammerjohann*.

Se considerarmos esses dados como fidedignos, uma conclusão coerente, apesar de não conclusiva, seria que entre a faixa C e a faixa B do *Corpus Stammerjohann* já seria registrada uma forte redução no uso dos clíticos (não há como saber se iniciada em anos anteriores) e que, em seguida, esta teria prosseguido com menor intensidade. Isso poderia datar a redução pelo menos à geração nascida durante a Segunda Guerra, se não antes. Seria evidente a transformação ocorrida na língua entre a geração nascida por volta de 1925 e aquela subsequente. Não se pode afirmar nada a respeito da geração ainda anterior, mas parece-nos que o fenômeno da redução dos clíticos sujeito não tenha continuado com a mesma força.

Portanto, parece que não são os fenômenos sociais do pós-guerra, em especial dos anos de 1960 e 1970 como a escolarização de massa e a pressão da mídia, especialmente da televisão, os responsáveis pelas mudanças linguísticas. Pelo menos no caso dos clíticos sujeito do florentino a redução aconteceu antes e por motivos diferentes, a serem investigados em estudos futuros.

4.2.3. Verbos *fare*, *andare* e *venire*

A escolha pelo verbo *essere* na forma *è* poderia influenciar os resultados em relação ao clítico *e*, que não aceita a posição diante a vogal¹⁴². Decidiu-se então avaliar outros lemas verbais que pudessem ser analisados, e que possivelmente nos fornecessem a possibilidade de avaliar realmente a presença do clítico *e*.

Após uma atenta análise de todos os lemas verbais de mais alta frequência, e tendo descartado alguns deles por poderem ser utilizados em expressões fixas, optou-se pela escolha dos verbos *fare* (fazer), *andare* (ir) e *venire* (vir). Para manter uma congruência com o quanto foi feito até o momento, optou-se por manter a pesquisa na terceira pessoa singular do indicativo presente, portanto as formas pesquisadas foram *fa*, *va* e *viene*. Mesmo sendo lemas verbais de alta frequência, o número de ocorrências encontrado é muitíssimo menor em relação ao verbo *essere*. Motivo pelo qual os estamos apresentando juntos.

4.2.3.1. Análise de *fa*

A primeira varredura do *Corpus Stammerjohann* com o *software AntConc* retornou 294 ocorrências da forma *fa*. Contudo, tivemos que descartar várias delas, a saber:

- transcrições incorretas;
- orações interrompidas ou totalmente incompreensíveis (nesta etapa ouvimos todos estes casos para verificar sua efetiva ininteligibilidade);
- *fa* inserida em uma forma fixa do tipo “*x tempo fa*” (x tempo atrás);
- *fa* como parte de expressões metereológicas do tipo “*fa freddo, fa caldo*” (está calor, está frio), sem sujeito;
- *fa* na realidade apocopado: por ex. *fa’* como infinitivo ou segunda pessoa singular do imperativo;
- *fa* inserido nos comentários dos transcritores.

O total de ocorrências analisáveis caiu para 191. O mesmo procedimento foi adotado para o *Corpus de Florentinos*: os casos de *fa* varridos pelo *AntConc* caíram de 239 para 123.

	<i>Corpus Stammerjohann</i>	<i>Corpus de Florentinos</i>
Total de palavras	86.252	58.299
Ocorrências de <i>fa</i>	294	239
Ocorrências de <i>fa</i> analisáveis	191	123

Tabela 20: dados absolutos de *fa* nos dois *corpora*.

¹⁴² Agradecemos Lúcia Fulgêncio por esta observação preciosa em sede de exame de Qualificação.

Como os dois *corpora* possuem um número de palavras diferentes, foi necessário normalizar todos os valores (por 100.000).

	<i>Corpus Stammerjohann</i>	<i>Corpus de Florentinos</i>
Ocorrências de <i>fa</i>	341	410
Ocorrências analisáveis	221	211

Tabela 21: dados normalizados por 100.000.

Para cada tipo de função da forma *fa* foi completada uma tabela detalhada com os tipos possíveis de sujeito:

- sujeito \emptyset ;
- sujeito lexical pré-verbal (inclui as orações subjetivas);
- sujeito lexical pós-verbal (inclui as orações subjetivas);
- sujeito pronominal tônico, demonstrativo, interrogativo, indefinido, numeral, possessivo;
- sujeito pronominal clítico;
- *cluster* de pronomes.

Além disso:

1. em todos os casos as ocorrências foram subdivididas em assertivas, negativas, interrogativas e impessoais;
2. de todos os casos, foram assinalados os textos de proveniência, o código do informante, além de um trecho do texto.

A tabela abaixo mostra os dados obtidos nas análises em números de ocorrências absolutas e normalizadas.

Forma	<i>Corpus Stammerjohann</i>		<i>Corpus de Florentinos</i>	
	Oc. absolutas	Oc. Normal.	Oc. absolutas	Oc. Normal.
Total forma <i>fa</i>	112	230	122	209
<i>L'</i>	0	0	0	0
<i>Gl'</i>	0	0	0	0
<i>La</i>	2	2	6	10
<i>E</i>	0	0	0	0
Sujeito \emptyset	78	90	70	120
Sujeito lexical pré-verbal	14	16	14	24
Sujeito lexical pós-verbal	8	9	10	17
Sujeito pronom. de outro tipo	12 (5 tônicos)	14 (6 tônicos)	22 (7 tônicos)	38 (12 tônicos)
<i>E+ la+ mi</i> (c.ind.)	0	0	1	2
<i>La + gli</i> (c. Ind.)	1	1	0	0
<i>La + se lo</i> (reflexivo+obj. Dir.)	0	0	1	2
<i>La + me lo</i> (obj. Ind.+obj dir.)	0	0	1	2

Tabela 22: análise das formas de expressão do sujeito dos enunciados com a forma verbal *fa*.

É possível observar que, no tempo, houve um crescimento no uso do sujeito clítico *la* e em suas formas concomitantes com outros clíticos (objeto direto e indireto). Por outro lado, não foi encontrado nenhum caso do clítico sujeito *e*, nem no *Corpus Stammerjohann*, nem naquele *de Florentinos*. No caso do sujeito \emptyset o aumento no *Corpus de Florentinos* foi de 33% e de outros tipos de pronomes, de 171%. Lembramos, todavia, a exiguidade dos valores.

A tabela abaixo evidencia os dados de duplicação do sujeito clítico com outras formas de sujeito.

Formas	<i>Corpus Stammerjohann</i>		<i>Corpus de Florentinos</i>	
	Oc. absolutas	Oc. Normal.	Oc. absolutas	Oc. Normal.
Lexical pré-v + tônico + <i>la</i> + <i>gli</i> (obj. ind)	1	1	0	0
<i>La</i> + tônico	0	0	1	2
<i>La</i> + tônico (posposto)	0	0	1	2
<i>E</i> + <i>la</i> + <i>mi</i> (obj.ind)	0	0	1	2

Tabela 23: ocorrências dos sujeitos duplicados com a forma verbal *fa*.

Também neste caso pode-se observar um aumento, no tempo, da forma *la*. Contudo, os valores são muito pequenos para fornecer qualquer conjectura.

Por fim, foram verificados os falantes que utilizaram as formas clíticas: dois pertencem à faixa B, dois à faixa C e um à faixa D. Os três mais idosos são aqueles que pronunciam o maior número de formas clíticas, 6 sobre um total de 9. Mais uma vez, porém, achamos esses valores baixos e os analisaremos no conjunto das três formas verbais estudadas neste parágrafo.

4.2.3.2. Análise de *va*

Passemos agora aos resultados obtidos da análise da forma verbal *va*. A primeira varredura com o *software AntConc* retornou 278 casos, dos quais foram descartados:

- os casos de transcrições incorretas;
- os casos de orações interrompidas ou totalmente incompreensíveis (nesta etapa ouvimos todos estes casos para verificar sua efetiva ininteligibilidade);
- os casos em que *va* faz parte de uma forma fixa como *va bene*, *va beh* (está bem, tá bom);
- os casos em que *va* na realidade é a forma apocopada *va'*, segunda pessoa singular do imperativo;
- *va* é utilizado para indicar a necessidade de se fazer algo: por ex. *va detto* (deve ser dito);
- *va* é inserido nos comentários dos transcritores.

Também neste caso foi necessário normalizar todos os valores (por 100.000).

	<i>Corpus Stammerjohann</i>	<i>Corpus de Florentinos</i>
Ocorrências de <i>va</i>	278	192
Ocorrências de <i>va</i> analisáveis	180	118

Tabela 24: dados dos dois *corpora* normalizados por 100.000.

É fornecida uma tabela detalhada como aquela feita para *fa*, com todos os dados linguísticos e sociolinguísticos. Os resultados são apresentados abaixo.

Formas	<i>Corpus Stammerjohann</i>		<i>Corpus de Florentinos</i>	
	Oc. absolutas	Oc. Normal.	Oc. absolutas	Oc. Normal.
Total forma <i>va</i>	180	209	118	202
<i>L'</i>	0	0	0	0
<i>Gl'</i>	0	0	0	0
<i>La</i>	11	13	6	10
<i>E'</i>	2	2	0	0
Sujeito \emptyset	113	131	85	146
Suj. Lexical Pré-v.	30	35	10	17
Suj. Lexical Pós-v.	11	12	10	17
Suj. Pronom. Outro	12 (5 tônicos)	14 (6 tônicos)	7 (3 tônicos)	12 (5 tônicos)
<i>La + mi</i> (obj. Ind.)	1	1	0	0
<i>Me</i> (dativo) + <i>la</i>	0	0	1	2

Tabela 25: análise das formas de expressão do sujeito dos enunciados com a forma verbal *va*.

Neste caso, pode-se observar que não houve aumento, a não ser um ligeiro incremento da forma \emptyset , mas diminuição de todas as formas de expressão do sujeito. As formas de sujeito clítico registraram alguns poucos casos: a forma *e* ocorreu duas únicas vezes no *Corpus Stammerjohann*, mas os valores são novamente muito baixos. Os sujeitos lexicais pós-verbais não serão comentados porque, como foi explicado no caso da forma *è*, sua motivação é pragmática e mereceriam um estudo à parte, que não é escopo deste trabalho.

A tabela abaixo evidencia os dados de duplicação do sujeito clíticos com outras formas de sujeito.

Formas	<i>Corpus Stammerjohann</i>		<i>Corpus de Florentinos</i>	
	Oc. absolutas	Oc. Normal.	Oc. absolutas	Oc. Normal.
Lexical pré-v. + <i>la</i>	2	2	0	0
Lexical pós-v + <i>la</i>	3	3	0	0
Lexical pós-v + <i>la</i> + <i>mi</i> (dat)	1	1	0	0
<i>e</i> + <i>lui</i> (tônico)	1	1	0	0
<i>lei</i> (tônico) + <i>la</i>	0	0	1	2

Tabela 26: ocorrências dos sujeitos duplicados com a forma verbal *va*.

Pode-se observar que houve diminuição de todos os casos, à exceção de um caso de tônico reduplicado pelo *la*. Contudo, também estes dados são exíguos demais para serem comentados isoladamente e isto será feito no final da seção.

Para concluir, no *Corpus Stammerjohann*, 11 das 13 ocorrências de clíticos foram pronunciadas por falantes da faixa C, uma pela faixa B e outra pela faixa A (os informantes da faixa A apareceram em apenas um de todos os textos analisados). No *Corpus de Florentinos*, dos seis clíticos sujeito encontrados, um foi pronunciado por um falante da faixa B, três pela faixa C e dois pela faixa D.

4.2.3.3. Análise de *viene*

As varreduras iniciais dos *corpora* com o *software AntConc* retornaram 66 ocorrências para o *Corpus Stammerjohann* e 44 para o *Corpus de Florentinos*. Foram descartados todos os casos em que:

- há transcrições incorretas;
- há orações interrompidas ou totalmente incompreensíveis (nesta etapa ouvimos todos estes casos para verificar sua efetiva ininteligibilidade);
- *viene* é utilizado em substituição a *è* para indicar a forma passiva;
- *viene* está inserido nos comentários dos transcritores.

As ocorrências realmente analisáveis foram assim ainda mais reduzidas:

	<i>Corpus Stammerjohann</i>	<i>Corpus de Florentinos</i>
Total de palavras	86.252	58.299
Ocorrências de <i>viene</i>	66	44
Ocorrências analisáveis	35	20

Tabela 27: dados absolutos de *viene* nos dois *corpora*.

Os dados foram normalizados por 100.000, resultando nos valores da tabela abaixo.

	<i>Corpus Stammerjohann</i>	<i>Corpus de Florentinos</i>
Ocorrências de <i>viene</i>	77	75
Ocorrências analisáveis	41	34

Tabela 28: dados dos dois *corpora* normalizados por 100.000.

Os procedimentos para a análise e as subdivisões em tabelas seguiram aqueles adotados para as outras formas verbais. Os resultados são apresentados abaixo.

Forme	<i>Corpus Stammerjohann</i>		<i>Corpus de Florentinos</i>	
	Oc. absolutas	Oc. Normal.	Oc. absolutas	Oc. Normal.
Total forma <i>viene</i>	35	41	20	34
<i>L'</i>	0	0	0	0
<i>Gl'</i>	0	0	0	0
<i>La</i>	3	3	5	9
<i>E'</i>	1	1	0	0
Sujeito \emptyset	7	8	8	14
Suj. Lexical Pré-v.	6	7	4	7
Suj. Lexical Pós-v	16	18	3	5
Suj Pronom. Outro tipo	1	1	1	2
<i>La + ti</i> (obj. Ind.)	1	1	0	0

Tabela 29: análise das formas de expressão do sujeito dos enunciados com a forma verbal *viene*.

Pela exiguidade dos valores é difícil poder comentar os dados: o que foi registrado é um crescimento no tempo do uso do clítico *la* e um aumento também da forma \emptyset . Há um único caso do clítico *e*, no *corpus* mais antigo. Quanto às duplicações do clítico com outra forma de expressão do sujeito, há um único caso, no *Corpus Stammerjohann*, em que o tônico é duplicado pelo clítico *la*.

No *Corpus Stammerjohann* todos os clíticos são pronunciados pela faixa C, assim como no *Corpus de Florentinos*, pelas faixas de mais idosos, C e D.

4.2.4. Discussão

A análise de outros verbos, na realidade, pouco acrescentou aos resultados alcançados com a forma *è* e nos fez refletir sobre a necessidade de um trabalho futuro que contemple todos os lemas verbais, de modo a se ter um panorama mais completo da situação dos clíticos sujeito no vernáculo florentino.

Contudo, quisemos investigar mais a fundo os dados que encontramos de *fa*, *va* e *viene*, analisando as características etárias dos informantes. Foram, portanto, subdivididas as formas de expressão do sujeito para cada informante. Após a contagem foram excluídas as formas de sujeito \emptyset e aquelas de sujeito posposto que, como dissemos, é motivado pragmaticamente. Foi possível então extrair o total de todas as formas de sujeito expresso. Após essa etapa, foram cotejados os casos em que tinha sido usada uma das formas de sujeito clítico pronominal. Calculou-se então a porcentagem dos sujeitos clíticos sobre o total das formas de sujeito expresso. Os resultados estão na tabela abaixo.

	FAIXA A	FAIXA B	FAIXA C	FAIXA D
<i>CORPUS STAMMERJOHANN</i>	11,11%	22,22%	28,33%	-
<i>CORPUS DE FLORENTINOS</i>	5,5%	23,5%	20%	71,4%

Tabela 30: total dos sujeitos clíticos encontrados nos dois *corpora* comparados, em porcentagem sobre os sujeitos expressos, divididos por faixas etárias.

A tabela mostra uma situação um pouco diferente daquela de *è*. No caso do *Corpus Stammerjohann* as porcentagens dos falantes de faixa A (que lembramos serem crianças) e aqueles da faixa C (seus pais) divergem bastante. Contudo, há de se ressaltar que, com os verbos já analisados, há uma participação mínima da faixa A nos textos. Isso explicaria essa redução tão forte. Quanto ao *Corpus de Florentinos*, observa-se que os clíticos são pronunciados especialmente pela faixa D, a dos mais idosos, confirmando em parte o que foi dito anteriormente. Averiguando os informantes de tal faixa, pudemos observar que todos os clíticos pronunciados nessa faixa pertencem a um único informante, *LUC, portanto, se o retirássemos da amostra, a porcentagem se reduziria a zero. Desta forma, a situação seria bastante similar àquela que foi observada no caso de *è*.

Contudo, repetimos, os dados são muito exíguos para podermos formular hipóteses mais concretas.

5. CONCLUSÕES

5.1. Conclusões sobre as formas do pronome de segunda pessoa no PB

Este estudo teve como propósito responder a uma questão que tem sido debatida na literatura, presente em muitos trabalhos, mas cujas propostas não nos pareciam convincentes. Também por razões metodológicas que antes da realização do *corpus* C-ORAL-BRASIL não podiam ser solucionadas: a relação entre forma segmental e função dos pronomes em PB, em particular o caso das formas *cê*, *ocê* e *você*. Nem todos os estudos anteriores eram baseados na análise de dados reais, mas quando o eram se tratava ou de dados produzidos em laboratório ou de dados de bancos não consultáveis e verificáveis, e compostos sem uma arquitetura capaz de constituir um universo definido, ou de dados não analisáveis acusticamente. Como não existiam trabalhos sobre um *corpus* de fala espontânea com arquitetura e qualidade adequadas, justamente porque até muito recentemente não havia tal tipo de instrumento, a questão sobre tonicidade ou atonicidade das formas pronominais era frequentemente inferida a partir da função ou da posição, e não definida com base nos parâmetros prosódicos que a determinam. A possibilidade de utilizar o C-ORAL-BRASIL (RASO & MELLO, 2012), um *corpus* pensado e coletado para representar a fala espontânea da diatopia mineira, com um número muito grande de informantes diferentes, textos em situações comunicativas de ampla diversificação e qualidade acústica quase sempre suficiente para a análise prosódica, nos pareceu uma oportunidade para enfrentar o problema não de maneira inferencial, mas através da observação e medição de dados empíricos e ao mesmo tempo disponíveis para a comunidade científica que quisesse verificar nossas observações e conclusões sobre os pronomes *cê(s)/ocê(s)/você(s)*.

A tonicidade ou atonicidade dependem diretamente de aspectos prosódicos como a duração (em primeiro lugar), a intensidade e a frequência fundamental. As características do C-ORAL-BRASIL permitem, graças ao alinhamento texto-som e à qualidade acústica das gravações, a observação desses fenômenos prosódicos.

É aceito pela comunidade científica que, no PB, o parâmetro mais significativo na discriminação do acento é a duração silábica (cf. MASSINI, 1991; BARBOSA P. A., 2000 e 2002; FONTES, 2013): foi justamente este o critério utilizado neste estudo para correlacionar forma e acento.

Pode-se questionar porque não foram utilizados os outros parâmetros como f_0 ou intensidade. Nossa resposta é que:

1. a f0 veicula uma grande quantidade de informações, relativas a domínios de escopo muito variável, e com relevância muito maior para fatores de ênfase lexical e valor informacional ou ilocucionário do que para a marcação da tonicidade. Por isso, não teríamos como utilizá-la de forma comparativa para identificar as sílabas tônicas dentro de uma fala espontânea, onde as outras variáveis não fossem constantes;
2. a intensidade, justamente em razão da alta espontaneidade do *corpus*, certamente mostra os valores alterados por pelo menos três razões: a variação de intensidade também depende de outros fatores não controláveis dentro da fala espontânea, o ruído de fundo está presente em medida diferente em cada momento de cada gravação e o movimento da cabeça do falante pode ter gerado uma distância hora menor hora maior do microfone.

Assim sendo, nossa decisão foi de concentrarmo-nos no parâmetro da duração, como o parâmetro não somente mais relevante, mas também aquele que menos sofreria de fatores não controláveis. Não queremos afirmar categoricamente que as formas pré-verbais sejam átonas e aquelas pós-verbais tônicas. Queremos mostrar que as observações resultaram no fato de que as formas analisadas acusticamente *cê/ocê/você*, em posição pré-verbal são curtas com relação à média da unidade tonal, e em posição pós-verbal são longas. Estatisticamente não existe dúvida sobre o fato que a posição pré-verbal correlaciona muito fortemente com a duração curta e que, ao inverso, a posição pós-verbal correlaciona com a duração longa; o passo para inferir que a posição pré-verbal é átona e a pós-verbal é tônica é curto, mas não deixa de ser um passo inferencial.

Os dados, portanto, podem ser interpretados de duas formas: uma, mais fraca, afirmaria que certamente há uma correlação além de qualquer dúvida estatística entre a duração das formas e sua função e posição sintática. E que tal relação é muito mais forte do que aquela entre forma segmental e posição/função sintática. A interpretação mais forte afirmaria que a relação entre tonicidade/ atonicidade (marcadas pela duração) é decisiva para veicular a função sintática: portanto a duração marcaria se uma forma é átona, ou seja, clítica, ou se é tônica, portanto autônoma do ponto de vista fonológico.

Os testes estatísticos corroboraram nossa hipótese e foram categóricos em relacionar a duração e a função. Os casos que não corresponderam às nossas previsões são pouquíssimos e não influentes do ponto de vista estatístico. Uma vez que os elementos que agem na fala espontânea são inúmeros, a variabilidade dos casos que fogem à hipótese pode depender de fatores não controláveis. Os valores obtidos com o teste estatístico nos dizem que a correlação entre duração e posição/função sintática é fortemente significativa, além de qualquer dúvida.

Enfim, acreditamos ter mostrado pelo menos duas conclusões significativas:

1. que estudar os fenômenos linguísticos do PB através de um *corpus* de fala espontânea adequadamente preparado constitui uma perspectiva metodológica inovadora e que pode levar a entender melhor também aspectos da língua tradicionalmente muito estudados;

2. que a *vexata quaestio* das formas pronominais do PB se apresenta agora, com os dados analisados neste trabalho, de maneira muito diferente. Não nos parece exagerado supor que o PB apresente de fato dois pronomes diferentes a partir de cada forma segmental: cada forma pode ser, dependendo da posição sintática, átona ou tônica. A pergunta então é: porque existe essa alomorfa segmental, uma vez que o que parece distinguir a função são as características suprasegmentais? Esta tese não pretende responder a essa pergunta. Uma possibilidade pode ser realmente uma progressiva redução das formas bissilábicas para domínios sociolinguísticos e funções cada vez mais limitadas, deixando apenas à prosódia a marca funcional. Mas isso é apenas uma hipótese que vai além do escopo deste trabalho. Para verificá-la seria necessário dispor de *corpora* comparáveis que reflitam diacronias diversas. Talvez, do ponto de vista sociolinguístico, a análise do *corpus* formal, cuja compilação está em curso, poderá nos mostrar que em domínios diferentes o quadro muda.

5.2. Conclusões sobre os pronomes sujeito no vernáculo florentino

A parte do trabalho referente ao vernáculo florentino analisou as formas clíticas sujeito de terceira pessoa do singular *gl'*, *l'*, *la*, e em dois *corpora* comparáveis em um estudo sincrônico e diacrônico. A finalidade era averiguar o uso desses clíticos no italiano de florentinos e a hipótese era de que havia acontecido uma forte redução desse uso ao longo das quatro décadas que separam a compilação dos dois *corpora*. A disponibilidade de dois *corpora* comparáveis de fala espontânea, com texto alinhado ao som, nos permitiu fazer um estudo diacrônico sobre o uso desses pronomes.

A opinião comum entre os historiadores da língua italiana era de que havia acontecido uma forte redução do uso dos clíticos sujeito em florentino. Essa hipotética redução, como muitos outros fenômenos oriundos da pressão do italiano padrão sobre os dialetos ou os vernáculos locais, teria acontecido principalmente entre o final dos anos de 1960 e os anos de 1970, como efeito da escolarização de massa, da popularização da televisão e das grandes facilitações para a mobilidade da população.

Um exame atento dos dados, que levou em conta os informantes, sua faixa etária e a presença excessiva de alguns deles no *Corpus de Florentinos*, mostrou que, de fato, houve

uma queda consistente no uso dos clíticos. Mas, por outro lado, não confirmou a hipótese temporal dessa queda, levando a uma retrodatação significativa do fenômeno. O fenômeno aparece já em ação na geração que nasceu antes da Segunda Guerra Mundial, por volta de 1925, que mostra uma redução importante com relação à geração anterior. Não podemos, contudo, saber se o fenômeno se iniciou ainda antes, pois os informantes mais velhos do *corpus Stammerjohann* nasceram apenas uma geração antes desses. Contrariamente ao que se pensa, os anos entre o final da década de 1960 e a década de 1970 não constituem, portanto, o início do fenômeno nem um momento de aceleração dele. Nossos dados mostram que o fenômeno aparece desde a segunda geração representada no *corpus* e continua de maneira menos intensa nas gerações sucessivas.

O trabalho não conseguiu exaurir a pesquisa sobre todo o fenômeno e oferecer dados mais detalhados sobre cada pronome: inicialmente, supomos que a análise dos pronomes sujeitos do verbo *essere* conseguiria nos dar as respostas que procurávamos. Posteriormente, achamos que inserindo também os três verbos mais frequentes depois de *essere* teríamos um quadro suficiente, mas no final percebemos que será necessário, em um trabalho futuro, continuar a investigação a partir de todos os lemas verbais.

Um problema com o qual nos defrontamos foi aquele da comparabilidade dos *corpora*: os cortes que necessariamente tivemos que fazer para conseguir obter falantes totalmente florentinos acabaram por afetar em parte os resultados. Apesar disso, foram de muita valia, pois nos obrigaram a refletir sobre a importância das características sociolinguísticas dos informantes e não somente dos textos. Isso, contudo, levou a outros cortes. Enfim, a busca por uma maior comparabilidade dos *corpora* nos obrigou a desconsiderar muitos dos dados coletados.

Finalmente, a análise acústica que gostaríamos de ter implementado também nesta parte do trabalho não pôde ser feita devido à qualidade acústica do *Corpus Stammerjohann*: suficiente para uma compreensão do texto, mas muito baixa para o exame de um espectrograma.

Contudo, o trabalho trouxe duas conclusões importantes, além de permitir a elaboração da metodologia correta para aprofundamentos futuros:

1. confirmou a partir de dados empíricos a convicção entre os historiadores da língua italiana de que houve uma redução progressiva das formas clíticas de pronomes sujeito entre os florentinos;
2. não confirmou a cronologia atribuída a esse fenômeno, retrodatando-o em pelo menos 40 anos, se não mais.

5.3. Conclusões gerais

Este trabalho analisou dois problemas distintos que ao mesmo tempo permitem mostrar duas estratégias diferentes com as quais as línguas lidam com o fenômeno da cliticização dos pronomes sujeito.

A finalidade de um estudo que compara e analisa duas línguas diferentes, o PB em sua variedade mineira e o italiano na sua variedade florentina, é mostrar como estas se valem de estratégias diferentes para assinalar a relação entre forma segmental, função e tonicidade na expressão dos pronomes sujeito.

No vernáculo florentino, a função é plenamente gramaticalizada por meio de formas segmentais especializadas: os clíticos sujeito possuem formas próprias e distintivas.

No PB esta relação não está gramaticalizada através de formas identificáveis morfológica e lexicalmente, pois, como foi mostrado, a forma *cê* pode exercer função de sujeito, mas também de objeto de preposição e provavelmente também de objeto de verbo e de tópico. O mesmo acontece com as formas *ocê* e *você*.

Poderia supor que haveria restrições no uso da forma *cê* em algumas funções. Os únicos exemplos de *cê* encontrado em função de objeto direto, de sujeito posposto (em clivada) e de tópico não são certamente suficientes para afirmar que essa forma pode normalmente ser usada nessas funções. Bem diferente do caso de objeto de preposição, em que a forma *cê* é bem representada e constitui um quarto do total. Contudo, é importante observar que não encontramos nenhum caso de *você* com função de tópico e apenas um de *você* com função de objeto direto. Portanto, é muito possível que a presença de apenas um caso de *cê* como objeto direto e tópico não seja devido a restrições, mas apenas à pouca frequência do fenômeno, seja ele marcado segmentalmente com a forma *cê* ou com uma forma bissilábica. Isso significa simplesmente que para uma resposta definitiva é necessário um *corpus* maior, provavelmente muito maior.

O que podemos afirmar é que, no PB, em sua variedade mineira, o fato de que a relação entre função e tonicidade não esteja gramaticalizada em formas segmentais específicas ficou evidente através dos dados extraídos de um *corpus* de fala espontânea informal. Pôde-se observar, através dos recursos de áudio que o *corpus* possui e as análises implementadas, que a função parece ser assinalada através de traços prosódicos, e que, se existem, eventuais restrições ao uso da forma monossilábica não são absolutamente evidentes e devem ser demonstradas, enquanto a polimorfia segmental na maioria das funções é certa e a correlação entre duração e função, depois deste trabalho, parece comprovada.

Em todo caso, acreditamos que a pesquisa teve o mérito de propor um novo método de abordagem da questão do suposto processo de cliticização do *cê*. Uma abordagem que parte dos dados empíricos para tentar fornecer respostas, sem se basear em processos inferenciais que transferem considerações de um nível descritivo para outro.

Como frequentemente acontece, um trabalho abre porta para novas perguntas. Uma delas, como já mencionado, diz respeito ao tamanho de um *corpus* capaz de responder de maneira definitiva às supostas restrições de todas as funções dos pronomes. *A priori*, não era imaginável que um *corpus* que apresenta milhares de formas do pronome deixasse em aberto questões sobre algumas das funções deles. Agora sabemos que um *corpus* de 200.000 palavras ainda não é suficiente para responder a tudo (imaginemos os *corpora* de poucas dezenas de milhares de palavras usados nos melhores trabalhos feitos até então!).

Uma outra pergunta diz respeito às diferenças quantitativas das formas do singular e do plural, talvez relacionadas ao uso do pronome com valor de indefinido. Acreditamos que o *corpus* possa dar conta de responder a essa questão, mas esta tese não conseguiu enfrentar também esse aspecto.

Por fim, na pesquisa sobre o italiano, acreditamos que uma análise de todos os lemas verbais possa iluminar melhor em detalhe a evolução dos diferentes pronomes, além do quadro geral. Contudo, todo o cuidado deve ser mantido para que a comparabilidade entre os dois *corpora* seja o mais completa possível. O risco de obter resultados que são fruto de *corpora* não balanceados de maneira homogênea é alto. E isso mostra também quão discutíveis são os resultados obtidos em geral a partir de bancos de dados, se não sabemos como esses bancos de dados foram realizados e se não podemos consultá-los e verificar os resultados dos trabalhos cujos dados são extraídos a partir deles.



Consideramos que seria muito interessante poder continuar a investigação com um *corpus* que possua as mesmas características do C-ORAL-BRASIL (RASO & MELLO, 2012), mas seja maior em número de palavras. Se conseguirmos encontrar dados suficientes para confirmar nossas hipóteses sobre a correlação entre duração e tonicidade, estamos convencidos de que um *corpus* maior seria de extrema valia na análise de questões como tópico, objeto direto ou indireto.

Um outro estudo que poderia partir deste seria aquele baseado em um *corpus* com as mesmas características acima, mas coletado com uma certa distância de tempo, em um estudo diacrônico. Isso permitiria investigar se há um processo em curso, pois é o que parece estar acontecendo. Esta impressão não é demonstrável a este ponto da pesquisa: só podemos afirmar que há uma expansão no uso da forma *cê* em todas as funções e uma queda na

utilização das formas bisilábicas. Se pudéssemos comprovar isso com um maior número de dados, significaria que a forma segmental na realidade deveria ser considerada um elemento superficial: a cada forma, *cê*, *ocê* e *você*, corresponderiam dois pronomes, um átono e um tônico.

APÊNDICE A: Conteúdos do CD

O CD que acompanha este trabalho contém:

1. uma pasta com os exemplos de áudio do PB (símbolo  no texto), com a numeração dos exemplos, seguidos de seu código do C-ORAL-BRASIL (nome do texto e número do enunciado);
2. uma pasta com os exemplos de áudio do vernáculo florentino (símbolo  no texto), com a numeração dos exemplos, seguidos de seu código do *Corpus Stammerjohann* ou do *Corpus per il Confronto Diacronico* (nome do texto e número do enunciado);
3. o texto da tese em PDF;
4. uma pasta com todas as análises acústicas feitas no programa *Praat*. Dentro desta pasta encontram-se subpastas divididas entre formas e posições, além de outras subpastas com os textos, em formato .txt dos enunciados analisados, divididos em unidades tonais e os arquivos áudio dos textos analisados. Acompanha o programa *Praat*.

REFERÊNCIAS

- ABERCROMBIE, D. *Elements of general phonetics*. Edinburgh: Edinburgh University Press, 1968
- ALFIERI, G.; La Sicilia. In: F. Bruni (Ed.). *L'italiano nelle regioni*. Torino: UTET, 1992, vol I, p. 798-860.
- ALI, SAID M.; *Gramática histórica da língua portuguesa*. 5. ed. São Paulo: Melhoramentos, 1965.
- ALI, SAID M.; *Gramática secundária da língua portuguesa*. Ed. rev. e com. E. Bechara, São Paulo: Melhoramentos, 1969.
- ALINEI, M.; Dialetto: un concetto rinascimentale fiorentino. Storia e analisi. In: *Quaderni di semantica*, 1981, p. 147-173.
- AMARAL, A.; *O dialeto caipira*. São Paulo: Hucitec, 1976.
- ANDRADE, L. V. S. de. *A variação de Você, Cê e Ocê no português brasileiro falado*. Dissertação de Mestrado. Brasília: Universidade de Brasília, 2004.
- AQUILECCHIA, G.; (Ed). *Giovanni Villani, Cronica*. Con le continuazioni di Matteo e Filippo. Torino: Einaudi 1979.
- ARRIGHI, G. (Ed.). *Libro d'abaco. Dal codice 1754 (sec. XIV) della Biblioteca Statale di Lucca*. Lucca: Cassa di Risparmio di Lucca, 1973.
- ARRIGHI, G.; *Il primo abaco in volgare (1307)*. Il Cod. 2236 della Biblioteca Riccardiana di Firenze. In: *Archivio storico italiano*, 1985, p. 429-435.
- BAGNO, M. *Gramática pedagógica do português brasileira*. São Paulo: Parábola Editorial, 2011.
- BARBOSA, L. P.; *Estatuto da forma cê: clítico ou palavra?* 2005. 137 f. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.
- BARBOSA, P. A. “Syllable-timing in Brazilian Portuguese”: uma crítica a Roy Major. *DELTA*, 16:369-402, 2000.
- BARBOSA, P. A. Explaining Cross-Linguistic Rhythmic Variability via a Coupled-Oscillator Model of Rhythm Production. BEL, B. & MARLIEN, I. (eds.) In: *Proceedings of the Speech Prosody 2002 Conference*, 11-13 April, Aix-en-Provence: Laboratoire Parole et Langage, 163-166, 2002.
- BARBOSA, P. A. Measuring speech rhythm variation in a model-based framework. In: *Proceedings of Interspeech 2009 - Speech and Intelligence*. Londres: Causal Productions, 2009, p. 1527 – 1530.

BARBOSA, P. A. Measuring speech rhythm variation in a model-based framework. In: *Proceedings of Interspeech 2009 - Speech and Intelligence*. Londres: Causal Productions, 2009, p. 1527 – 1530.

BARBOSA, P. A.; Do grau de não perifericidade da vogal /a/ pós-tônica final. In; *Revista Diadorim / Revista de Estudos Linguísticos e Literários do Programa de Pós-Graduação em Letras Vernáculas da Universidade Federal do Rio de Janeiro*. Volume 12, Dezembro 2012. [<http://www.revistadiadorim.lettras.ufrj.br>]

BARBOSA, P. A.; *Incursões em torno do ritmo da fala*. Campinas. Editora Pontes. 2006, p.195– 228.

BARBOSA, P. A.; Revelar a Estrutura Rítmica de uma língua construindo máquinas falantes: pela integração entre ciência e tecnologia de fala. In: *Estudos de Prosódia*, edited by Scarpa, Ester, 21-52. Campinas: Editora da Unicamp, 1999.

BARBOSA, P. A.; Semi-automatic and automatic tools for generating prosodic descriptors for prosody research. In: *Proceedings from TRASP 2013*, Tools and Resources for the Analysis of Speech Prosody, August 30, 2013 Laboratoire Parole et Langage, Editors: Brigitte Bigi, Daniel Hirst, Aix-en-Provence, France, 2013, pp. 86-90.

BARTOLI LANGELI, A.; INFELISE, M.; Il libro manoscritto e a stampa. In: F. Bruni (Ed.). *L'italiano nelle regioni*. Torino: UTET, 1992, vol I, p. 941-972.

BARTOLI LANGELI, A.; *La scrittura dell'italiano*. Bologna: Il Mulino, 2000.

BASTO, C.; *Formas de tratamento, em português*. Revista Lusitana, XXIX, p. 182-202, 1931.

BECHARA, E.; *Moderna gramática portuguesa*. São Paulo: Companhia Editora Nacional, 1982.

BERBER SARDINHA, T. *Linguística de corpus*. Barueri: Manole, 2004.

BERRUTO G.; *Sociolinguistica dell'italiano contemporaneo*, Roma, Carocci, 1987.

BERRUTO, Gaetano, Dialect/standard convergence, mixing, and models of language contact: the case of Italy. In: Auer Peter, Hinskens Frans, Kerswill Paul (a cura di), *Dialect change. Convergence and divergence in European languages*, Cambridge, Cambridge University Press, pp. 81-97, 2005

BERRUTO, Gaetano. Linguistica del contatto e aspetti dell'italianizzazione dei dialetti: appunti di creolistica casalinga. In: *Italica et Romanica*. Festschrift für Max Pfister zum 65. Geburtstag, hrsg. von G. Holtus, J. Kramer & W. Schweickard, Tübingen, Niemeyer, 3 voll., vol. 3°, pp. 13-29, 1997

BIBER, D.; *Representativeness in Corpus Design in Literary and Linguistic Computing*, Vol. 8, n. 4, 1993. Oxford University Press, 1993.

BIBER, D.; *Variation across Speech and Writing*. Cambridge: Cambridge University Press, 1988.

BINAZZI, N.; 'Italiano' e 'dialetto' a Firenze: 50 giovani giudicano il loro lessico d'uso. In: Muljačić Žarko (a cura di), *L'italiano e le sue varietà linguistiche*, Sauerländen, Verlag für deutsch-italienische Studien, pp. 13-32, 1998

BINAZZI, N.; L'appartenenza rivelata. Lessico e tradizioni del discorso nel parlato Fiorentino. In: «*Quaderni del Dipartimento di Linguistica dell'Università di Firenze*», a. 17, pp. 137-168, 2007

BINAZZI, N.; *Le parole dei giovani fiorentini: variazione linguistica e variazione sociale*, Roma 1997

BINAZZI, N.; Una palestra fiorentina per l'italiano in movimento: «io bisogna ve ne parli». In: Giovanni Ruffino e Mari D'Agostino (a cura di), *Storia della lingua italiana e dialettologia*. Atti dell'VIII Convegno Internazionale dell'ASLI, Palermo, 29-31 ottobre 2009. Palermo, Centro di studi filologici e linguistici siciliani 2010, pp. 249-264

BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer* [Computer program]. Version 5.3.51, retrieved 2 June 2013 from <http://www.praat.org/>

BOLOGNA, C.; L'Ordine francescano e la letteratura nell'Italia pretridentina. In: Asor Rosa, Alberto (Ed.). *Letteratura Italiana. Storia e geografia*. Torino: Einaudi, 1982, vol. I, p. 729-797.

BOSTRÖM I.; *La morfosintassi dei pronomi personali: soggetti della terza persona in italiano e in fiorentino*. Contributo allo studio storico dei rapporti fra l'italiano standard e la varietà fiorentina, Volume 5 di Acta Universitatis Stockholmiensis, Stockholm, Almqvist & Wiksell International, 1972

BRACCO, C.; BRANDI, L.; CORDIN, P.; Sulla posizione soggetto in italiano e in alcuni dialetti dell'Italia centro-settentrionale. In: *Sintassi e morfologia della lingua italiana d'uso*. Teorie e applicazioni descrittive. Atti del XVII Convegno Internazionale di Studi, Urbino, 11-13 settembre 1983, SLI, Società di Linguistica Italiana.

BRANCA, V.; (Ed.). *Giovanni Boccaccio, Decameron*. Edizione critica secondo l'autografo hamiltoniano. Firenze: Accademia della Crusca, 1976.

BRANCA, V.; *Boccaccio medievale e nuovi studi sul "Decameron"*. Firenze: Sansoni, 1996.

BRANCA, V.; *Mercanti e scrittori*. Milano, 1986

BRANDI, L. & P. CORDIN. 1989. *Two Italian dialects and the null subject parameter*. In: Jaeggli, O. & K. Safir. Eds.. *The Null Subject Parameter*. Dordrecht: Kluwer, 111-142, 1989.

BRANDI, L.; CORDIN, P.; *Dialetti italiani: un confronto sul parametro del soggetto nullo*, *Rivista Di Grammatica Generativa*, , 9781556080876, Vol. 6, pp. 33-87, 1981.

BRANDI, L.; CORDIN, P.; *Subject clitics in two Italian dialects: a perspective on the Null subject Parameter*, JOURNAL OF ITALIAN LINGUISTICS, Vol. 8.2, pp. 1-24, 1983.

BRUNI, F.; *Appunti sui movimenti religiosi e il volgare italiano nel Quattro-Cinquecento*. In: *Studi linguistici italiani*, 1983, p. 3-30.

CAGLIARI, L. C. *Acento em português*. Campinas, SP: Edição do autor, 1999

CÂMARA Jr J. M.; *Dicionário de filologia e gramática*. 4º ed. Rio de Janeiro: Ozon Editor, 1974.

CÂMARA Jr J. M.; *História e estrutura da língua portuguesa*. 2. ed. Rio de Janeiro: Padrão, 1976.

CARDINALETTI A., MUNARO N. (a cura di), *Italiano, italiani regionali e dialetti*. Milano, FrancoAngeli, 2009

CARDINALETTI, A. & M. STARKE. The typology of structural deficiency: On the three grammatical classes. In: RIEMSDIJK, H. (ed.) *Clitics in the Languages of Europe, Empirical Approaches to Language Typology*. Berlin: Mouton de Gruyter, 1999.

CARDINI, R.; (Ed.). *Cristoforo Landino, Scritti critici e teorici*. Roma: Bulzoni, 1974, 2 vol.

CASAPULLO, R.; *Il Medioevo*. Bologna: Il Mulino, 1999.

CASTELLANI POLLIDORI, O. (Ed.); *Niccolò Macchiavelli e il "Dialogo intorno alla nostra lingua"*. Firenze: Olschki, 1978.

CASTELLANI, A.; *La Toscana dialettale di epoca antica*. In: *Studi Linguistici Italiani*, 1997, p. 3-46 e 219-254.

CASTILHO, A.; (Org.) *Gramática do português falado*. 2. ed., v. 3. Campinas, SP: Unicamp, 1996.

CEGALLA, D. P.; *Novíssima gramática da língua portuguesa*. 43. ed. São Paulo: Cia. Editora Nacional, 2000.

CINTRA, L. F. L.; *Sobre Formas de Tratamento na língua portuguesa: ensaios*. Lisboa: Horizonte, 1972.

CIRÍACO, L.; VITRAL, L. & REIS, C. Intensidade e duração de formas reduzidas do Português Brasileiro. In: *Revista de Estudos da Linguagem*, v. 12, n. 2, 2004.

COELHO, Maria do Socorro V.; *Uma abordagem variacionista do uso de formas de tratamento no norte de Minas*. 1999. 85 f. Dissertação (Mestrado em Linguística) - Faculdade de Letras, Belo Horizonte, 1999.

CONOVER, W.J.; *Practical Nonparametric Statistics*. 3rd ed., John Wiley&Sons, 1999

CONTINI G.; *Letteratura italiana del Quattrocento*, Sansoni, Firenze, 1976

CONTINI, G.; Per una interpretazione strutturale della cosiddetta “gorgia” toscana. In: *Boletim de Filologia*, 1960, p. 263-281.

COOK, M.; *Uma teoria de interpretação das formas de tratamento na língua portuguesa*. Hispania, n. 80, p. 451-464, set. 1997.

CRESTI E., MONEGLIA M. (a cura di), C-Oral-Rom. Integrated Reference Corpora for Spoken Romance Languages, Benjamins, Amsterdam, 2005.

CRESTI E.; Brevi note sulle principali strategie lessicali e strutturali del parlato di quattro lingue romanze (italiano, francese, portoghese e spagnolo): dati dal corpus C-ORAL-ROM.. In: J. Korzen. *Lingua cultura e intercultura: l'italiano e le altre lingue*. Atti del VIII Convegno SILFI (Copenaghen, 22-26 giugno 2004), pp. 163-176, Frederiksberg, Copenaghen: Samfundslitteratur Press, ISBN:9788759311790, 2005

CRESTI E.; *Corpus di italiano parlato*, 2 voll., vol. I Introduzione; vol. II Campioni, Accademia della Crusca, Firenze, 2000.

CRESTI E.; enunciato e frase: teoria e verifiche empiriche. In: M. Biffi, O. Calabrese, L. Salibra. Italia linguistica: discorsi di scritto e di parlato. *Nuovi studi di linguistica italiana per Giovanni Nencioni*, pp. 249-260, Siena: BIFFI M., CALABRESE O., SALIBRA L. (A CURA DI), ISBN:9788880241386, 2005

CRESTI E.; La Stanza: un'unità di costruzione testuale del parlato. In: A. Ferrari. *Atti del X° Convegno SILFI*, "Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione e giustapposizione" (Basilea 30 giugno- 3 luglio 2008), pp. 713-732, Firenze: Franco Cesati editore, ISBN:9788876673825. 2010

CRESTI E.; La testualità parlata: alcuni dati dal corpus italiano di C-ORAL-ROM nella prospettiva del parlato romanzo. In: I. Korzen. *Lingua, cultura e intercultura: l'italiano e le altre lingue* ; atti del VIII convegno SILFI, Società internazionale di linguistica e filologia italiana (Copenaghen, 22-26 giugno 2004), pp. 163-176, Frederiksberg, Copenaghen: Samfundslitteratur Press, ISBN:8759311797, 2005

CRESTI E.; Notes on lexical strategies, structural strategies and surface clause indexes in the C-ORAL-ROM spoken corpora. In: CRESTI E.; MONEGLIA M. EDS. C-ORAL-ROM. Integrated reference corpora for spoken romance languages, pp. 209-256, AMSTERDAM: Benjamins, ISBN:9781588115485, 2005

CRESTI, E, MONEGLIA M.; “Informational Patterning Theory and the Corpus-based Description of Spoken Language: The Compositionality Issue in the Topic-comment Pattern.” In *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*, ed. Massimo Moneglia and Alessandro Panunzi, 13–45. Firenze: Firenze University Press, 2010.

CRESTI, E. & GRAMIGNI, P. *Per una linguistica corpus based dell'italiano parlato: Le unità di riferimento*. In: ALBANO LEONI F., CUTUGNO F., PETTORINO M., SAVY R. (Eds.), Atti del Convegno Nazionale “Il Parlato Italiano”, CD-ROM (pp. 1–26). Napoli: M. D’Auria, 2004.

CRESTI, E. “Syntactic properties of spontaneous speech in the Language into Act Theory”. In *Spoken Corpora and Linguistic Studies*, Raso, Tommaso and Heliana Mello (eds.), 365–410, 2014

CRESTI, E. E MONEGLIA, M. (a cura di) (2004). *Campionamento Corpus Stammerjohann*. Consultabile al sito: <http://lablita.dit.unifi.it/corpora/descriptions>

CUNHA, C. F.; *Gramática da língua portuguesa*. 11.ed. Rio de Janeiro: FAE, 1985.

CUNHA, C.; CINTRA, L.; *Nova gramática do português contemporâneo*. Rio de Janeiro: Ed. Nova Fronteira, 1985.

D’ACHILLE, P.; *Sintassi del parlato e tradizione scritta della lingua italiana*. Analisi di testi dalle origini al secolo XVIII, Roma, Bonacci, 1990.

Dardano, M.; *Profilo dell’italiano contemporaneo*, In Serianni & Trifone 1993-1994, vol. 2°, pp. 343-430, 1994.

DAVIS, C.; *L’istruzione a Firenze nel tempo di Dante*. Bologna: Il Mulino, 1988.

DRUSI, R.; *La lingua “cortigiana romana”*. Venezia: Il Cardo, 1995.

DUARTE, M. E. L. *Variação e Sintaxe: Clítico Acusativo, Pronome Lexical e Categoria Vazia no Português do Brasil*. Dissertação de Mestrado, Pontifícia Universidade Católica-São Paulo, 1986.

DUARTE, M. E. L.; Do pronome nulo ao pronome pleno: a trajetória do sujeito no português do Brasil. In: ROBERTS, I. e KATO, M. (orgs.) *Português Brasileiro: uma viagem diacrônica*. Campinas: Ed. UNICAMP, 1993.

DUBOIS, J. W. et al.; *Santa Barbara corpus of spoken American English - Part 1*. Linguistic Data Consortium, 2000.

FARACO, C. A.; O tratamento você em português; uma abordagem histórica. In: *Fragmenta*, Curitiba: Ed. da UFPR, n. 13, p. 51-82, 1996.

FERREIRA NETO, W. O acento na língua portuguesa. In: ARAÚJO, G. *O acento em português: abordagens fonológicas*. São Paulo: Parábola, 2007. p. 21-36.

FERREIRA, M. *Argumentos Nulos em Português Brasileiro*. M.A. Thesis, Universidade Estadual de Campinas, 2007.

FERREIRA, M. Hyperraising and Null Subjects in Brazilian Portuguese. *MIT Working Papers in Linguistics 47: Collected Papers on Romance Syntax*, p. 57-85, 2004.

FIGUEIREDO SILVA, M. C.; *A Posição do Sujeito no Português Brasileiro*. Campinas, SP: Editora da Unicamp, 1996.

FIRENZUOLI, V. & TUCCI, I. (2003), L'unità informativa di inciso: correlati intonativi, in *La coarticolazione. Atti delle XIII giornate di studio del Gruppo di fonetica sperimentale* (Pisa, 28-30 novembre 2002), a cura di G. Marotta & N. Nocchi, Pisa, ETS, pp. 185-192.

FONTES, A. S.; Pistas Acústicas e a Percepção do acento lexical em Português Brasileiro. In *Revista Intercâmbio*, v. XXVII: 100–109, 2013. São Paulo: LAEL/PUCSP. ISSN 2237-759x.

GALLI DE' PARATESI, Nora. *Lingua toscana in bocca ambrosiana. Tendenze verso l'italiano standard: un'inchiesta sociolinguistica*, Bologna, il Mulino, 1984

GALVES, C.; ABAURRE, M. B. M.; Os clíticos no português brasileiro: elementos para uma abordagem sintático-fonológica. In: *Gramática do português falado*. Campinas: Ed. UNICAMP, 1996.

GENSINI, S. (Ed.); *La Toscana nel secolo XIV. Caratteri di una civiltà regionale*. Pisa: Pacini, 1988.

GHERARDI P.; *Opera matematica. Libro di ragioni–Liber habaci*. Lucca: Pacini Fazzi 1987.

GIANNELLI, L.; Toscana. In: *Profilo dei Dialetti Italiani*, a cura di Alberto Zamboni, Consiglio Nazionale delle ricerche, Istituto di fonetica e dialettologia, Pacini Editore, Ospitaletto, 2000.

GIANNELLI, Luciano, SAVOIA, Leonardo. L'indebolimento consonantico in Toscana III. *Rivista italiana di dialettologia* 2/4: 23-48/38-101, 1978 e 1979-1980

GIANNELLI, Luciano. Toscana. In: *Profilo dei Dialetti Italiani*, a cura di Alberto Zamboni, Consiglio Nazionale delle ricerche, Istituto di fonetica e dialettologia, Pacini Editore, Ospitaletto, 2000

GRAYSON, C.; *Leon Battista Alberti. La prima grammatica della lingua volgare*. Bologna: Commissione per i testi di lingua, 1964.

GRENDLER, P.; *Schooling in Renaissance Italy: Literacy and Learning, 1300-1600*. Baltimore: Johns Hopkins University Press, 1991.

GUY, G. & ZILLES, A. *Sociolinguística Quantitativa, instrumental de análise*. São Paulo. Parábola Editorial, 2007.

HARY, B. H. (ed.). *Corpus Linguistics and Modern Hebrew: Towards the Compilation of The Corpus of Spoken Israeli Hebrew (CoSIH)*. Tel Aviv: Tel Aviv University, The Chaim Rosenberg School of Jewish Studies, 2003.

HAYWARD, K.; *Experimental Phonetics*. Harlow, England: Longman, 2000.

HEUVEL, H. van den et al. *Validation of spoken language resources: an overview of basic aspects*. In: *Language Resources & Evaluation*, n. 42, p. 41-73, 2008.

HEUVEL, H. van den et al.; SLR Validation: present state of affairs and prospects. In: *Proceedings of the 2nd International Conference on Language Resources and Evaluation*

(LREC 2000). Paris: ELRA, p. 435-440. 2000. Disponível em: <<http://www.lrec-conf.org/proceedings/lrec2000/pdf/39.pdf>> Acesso em: 20 dez. 2012.

HOPPER, Paul J.; TRAUGOTT, Elizabeth C. *Grammaticalization*. Cambridge: Cambridge University Press, 1993.

HORA, Demerval da. Teoria da variação: uma retrospectiva. In: _____ (Org.). *Diversidade lingüística no Brasil*. João Pessoa, PB: Idéia, 1997.

ILARI, R. et al. Os pronomes pessoais do português falado: roteiro para a análise. In: CASTILHO, Ataliba T. de; BASÍLIO, Margarida. *Gramática do português falado*. vol. IV. Campinas: Ed. UNICAMP, 1996.

JEZEK, E.; Interazioni tra aspetto e diatesi nei verbi pronominali italiani, In *Studi di grammatica italiana* 22, pp. 239-281, 2004

KATO, M.; Strong and weak pronominals in the null subject parameter. In: *Probus*, 11, pp.1-37, 1999.

LABOV, W. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press, 1972.

LADEFOGED, P. *A course in phonetics*. 2nd ed. Fort Worth: Harcourt Brace College, 1982

LADEFOGED, P. *Elements of Acoustic Phonetics*. Chicago: University Press, 1996

LAVER, J.; *Principles of Phonetics*. Cambridge: Cambridge University Press, 1994.

LEONE, F., *I pronomi personali di terza persona*. L'evoluzione di un microsistema nell'italiano di fine millennio, Roma, Carocci, 2003

LIBRANDI, R. L'italiano nella comunicazione della Chiesa e nella diffusione della cultura religiosa. In: L. Serianni; P. Trifone (Eds.). *Storia della lingua italiana*. Torino: Einaudi, 1992-1994, vol. I, p. 335-381.

LO CASCIO, V., *Strutture pronominali e verbali italiane*, Bologna, Zanichelli, 1970

LOPES, Célia R. S. e DUARTE, M. Eugênia L.; De Vossa Mercê a você: análise da pronominalização de nominais em peças brasileiras e portuguesas setecentistas e oitocentistas. In: BRANDÃO, Silvia F. & MOTA, M. Antônia (Org.). *Análise contrastiva de variedades do Português: primeiros estudos*. RJ: In-fólio, 2003, p.61-76.

LOWRY, M.; *The World of Aldus Manutius. Business and Scholarship in Renaissance Venice*. Oxford: Blackwell, 1979.

LUCCHI, P.; Leggere, scrivere e abbaco: l'istruzione elementare agli inizi dell'età moderna. In: *Scienze, credenze occulte, livelli di cultura*. Convegno Internazionale di Studi. Firenze: Olschki, 1982.

LUZ, M. dos Santos; *Fórmulas de tratamento do português*. Revista Portuguesa de Filologia. Coimbra, v. II. T. I, II, p. 256-363, 1956.

MACWHINNEY B.; *Il progetto CHILDES - Strumenti per l'analisi del linguaggio*, edizione italiana a cura di Elena Pizzuto e Umberto Bortolini, Edizioni del Cerro, Pisa, 1997.

MACWHINNEY B.; *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.

MAJOR, R. *Stress and rhythm in Brazilian Portuguese*. *Language*, v. 61, n. 2, p. 259-282

MANACORDA, G.; *Storia della scuola in Italia. Il Medioevo*. Firenze: Le lettere, 1914, 2 vol.

MANNI, P.; Dal toscano all'italiano letterario. In: Serianni, Luca; Trifone, Pietro. *Storia della lingua italiana. Scritto e parlato*. Torino: Einaudi, 1992-1994, vol. II, p. 321-342.

MANNI, P.; *Il Trecento toscano*. Bologna: Il Mulino, 2003.

MANNI, P.; Ricerche sui tratti fonetici e morfologici del fiorentino quattrocentesco. In: *Studi di grammatica italiana*, 1979, p. 67-155.

MANNI, P.; Toscana. In: L. Serianni; P. Trifone (Eds.). *Storia della lingua italiana*. Torino: Einaudi, 1992-1994, vol. III, p. 294-328.

MARAZZINI, C.; Le teorie. In: L. Serianni; P. Trifone (Eds.). *Storia della lingua italiana*. Torino: Einaudi, 1992-1994, vol. I, p. 231-334.

MARTIN, P. WinPitch Corpus: A text-to-speech analysis and alignment tool. In: CRESTI, E; MONEGLIA, M. (Org.). *C-ORAL-ROM: integrated reference corpora for spoken Romance languages*. Amsterdam: John Benjamins, 2005, p. 40-51.

MASSINI, G. *A duração no estudo do acento e do ritmo em português*. [Dissertação de mestrado]. Universidade Estadual de Campinas, 1991

McENERY T., WILSON A. (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press

McENERY T., XIAO R., TONO Y. 2006. *Corpus-based language studies: an advances resource book*. London; New York: Routledge.

MELLO H. *et al.* Capítulo 4. Transcrição e segmentação prosódica do corpus C-ORAL-BRASIL: Critérios de implementação e validação. In: RASO T.; MELLO H. *C-ORAL-BRASIL I. Corpus de referência do português brasileiro falado*. I edição. Belo Horizonte, Editora UFMG, 2012, p. 125-176.

MELLO H.; Os corpora orais e o C-ORAL-BRASIL. In: RASO, T.; MELLO, H. (Org.). *C-ORAL-BRASIL-I: Corpus de referência do português brasileiro falado informal*. 1ed.: , 2012, v. 1, p. 31-54.

MELLO H.; RASO T. Para a transcrição da fala espontânea: o caso do C-ORAL-BRASIL. In: *Revista Portuguesa de Humanidades*, v. 13, n. 1, p. 153-178, 2009.

MELLO, H.; RASO, T. Illocution, modality, attitude: different names for different categories. In: MELLO, H.; PANUNZI, A.; RASO, T. (Ed.). *Illocution, modality, attitude, information patterning and speech annotation*. Firenze: FUP, 2010.

MENÓN, Odete P. S. *O sistema pronominal do português do Brasil*. Letras, Curitiba: Ed. da UFPR, n. 44, p. 91-106, 1995.

METTOUCHI, A. & CHANARD, C. From Fieldwork to Annotated Corpora: The CorpAfroAs project. In: *Cahiers de Faits de Langues* n°2. 2010, 255-265.

METTOUCHI, A.; CAUBET, D.; VANHOVE, M.; TOSCO, M.; COMRIE, B.; IZRE'EL S. "CORPAFROAS. A corpus for spoken Afroasiatic languages: morphosyntactic and prosodic analysis". In: CAMSEMUD 2007, F.M. Fales & G.F. Grassi (eds), Padova: SARGON. 2010, pp. 177-180.

MEYER C.F. *English Corpus Linguistics*. An introduction. 2004. Cambridge: Cambridge University Press

MITTMANN, M. *O C-ORAL-BRASIL e o estudo da fala informal: um novo olhar sobre o Tópico no Português Brasileiro*. Tese de Doutorado em Linguística, Faculdade de Letras, UFMG, 2012.

MONEGLIA M., BINAZZI N., CELLA R., SCARANO A., PANUNZI A., FABBRI M.; L'incidenza del lessico fiorentino nella lingua d'uso a Firenze. Un confronto tra il corpus Stammerjohann del 1965 e un corpus di parlato contemporaneo. In E. Cresti (a c. di) *Prospettive nello studio del lessico Italiano*, Atti del IX Congresso internazionale della Società di Linguistica e Filologia Italiana. Firenze, Firenze University Press, pp. 99-106, 2008.

MONEGLIA M.; RASO T.; MITTMANN, M. M.; RIBEIRO MELLO, H. "Challenging the Perceptual Relevance of Prosodic Breaks in Multilingual Spontaneous Speech Corpora: C-ORAL-BRASIL / C-ORAL-ROM." In: *Prosodic Prominence Perceptual and Automatic Identification - Speech Prosody 2010 Satellite Workshop*. Chicago: Université de Neuchâtel, 2010.

MONEGLIA, M. AND RASO T.; "Appendix". In *Spoken Corpora and Linguistic Studies*, Raso, Tommaso and Heliana Mello (eds.), 468-495, 2014

MONEGLIA, M. et al.; Challenging the perceptual relevance of prosodic breaks in multilingualspontaneous speech corpora: C-ORAL-BRASIL/C-ORAL-ROM. Prosodic prominence perceptual and automatic identification. In: *Proceedings of the Speech Prosody International Conference Satellite Workshop*. Chicago: Université de Neuchâtel, 2010. Disponível em: <<http://speechprosody2010.illinois.edu/papers/102010.pdf>> Acesso em: 20 dez. 2012.

MONEGLIA, M., SCARANO, A.; Il Corpus Stammerjohann. Il primo corpus di italiano parlato, in rete nella base dati di LABLITA. In M. Pettorino (a c. di) *Atti del convegno internazionale "La comunicazione parlata"*, Tomo III. Napoli, Liguori, pp. 1650-1685, 2008.

MONEGLIA, M.; Units of Analysis of Spontaneous Speech and Speech Variation in a Cross-linguistic Perspective. In Y. Kswaguchi et al. (a cura di), *Usage Based Linguistics Informatics "The Second International Conference on Linguistics Informatics"*. Amsterdam: Benjamins, pp. 153-179, 2006.

MONEGLIA, M.; Units of Analysis of Spontaneous Speech and Speech Variation in a Cross-linguistic Perspective. In: KAWAGUCHI, Y.; ZAIMA, S.; TAKAGAKI, T. In: *Spoken Language Corpus and Linguistics Informatics*, pp. 153-180, Amsterdam: Benjamins Publishing Company, ISBN: 9789027233172, 2006.

MORAES, J. A. Intonation in Brazilian Portuguese. In: HIRST, D.; DI CRISTO, A. MORGANA, S.; PIOTTI, M.; PRADA, M. (Eds). *"Prose della volgar lingua" di Pietro Bembo*. Milano: Cisalpino, 2000.

MOURA NEVES, M. H. Os pronomes. In. Ilari, R; Neves, M. H. M. *Gramática do português culto falado no Brasil*. Volume III: classes de palavras e construções. Campinas: Editora da Unicamp, 2008.

MOURA NEVES, M. H.; *Gramática de usos do português*. São Paulo: UNESP, 2000.

NASCENTES, A.; O tratamento de "você" no Brasil. In: *Letras*. Curitiba/PR: Ed. UFPR, v.6, n.05, p. 114-122, 1956.

NASCIMENTO, I. B.; Interpolação de constituintes entre "(vo)cê + verbo". In: *Estudos Linguísticos*, 39 (2), 2010.

NENCIONI, G. Parlato-parlato, parlato-scritto, parlato-recitato . In: NENCIONI, G. (Org.). *Di scritto e di parlato*. Bologna: Zanichelli, 1983. p. 126-179

NOCENTINI, A., Sulla diatesi del verbo italiano, *Studi di grammatica italiana* 6, pp. 451-471, 1977

NOETHER, G.E.; DUEKER, M.; *Introduction to Statistics: The Nonparametric*. Springer, 1990.

NUNES, J. Sintaxe Gerativa Comparada: Português Brasileiro e Português Europeu. Notas de aula, 1º semestre de 2007, Universidade de São Paulo, 2007.

OTHERO, G. A. Revisitando o status do pronome cê no português brasileiro. In: *Revista de Estudos da Linguagem*, v. 21, n. 1, 2013.

PALERMO, M. *Soggetto*, in *Enciclopedia dell'italiano* (a c. di Raffaele Simone), vol. II, Roma, Istituto della Enciclopedia Italiana, 2011, pp. 1377-83.

PALERMO, M., TRIFONE, P., *Grammatica italiana di base*, Bologna, Zanichelli, 2007 (I ed. 2000).

PALERMO, M.; Fenomeni di standardizzazione a Roma nel primo Cinquecento. In: *Contributi di Filologia dell'Italia Mediana*, 1991, p. 23-52.

PALERMO, M.; Sull'evoluzione del fiorentino Tre-Quattrocentesco. In: *Nuovi Annali della Facoltà di Magistero dell'Università di Messina*, 1992, p. 131-156.

PALERMO, M.; *L'espressione del pronome personale soggetto nella storia dell'italiano*. Roma: Bulzoni, 1997.

PANUNZI, A.; *La variazione semantica del verbo essere nell'italiano parlato. Uno studio su corpus*. Firenze University Press, Firenze, 2010

PAREDES SILVA, V. L.; (1998) Variação e Funcionalidade no Uso de Pronomes de 2ª Pessoa do Singular no Português Carioca. In: *Revista de Estudos da Linguagem*, v.7, nº 2, pp.121-138.

PATOTA, G.; I percorsi grammaticali. In: L. Serianni; P. Trifone (Eds.). *Storia della lingua italiana*. Torino: Einaudi, 1992-1994, vol. I, p. 93-138.

PATOTA, G.; *Lingua e linguistica in Leon Battista Alberti*. Roma: Bulzoni, 1999.

PERES, E. P. *O uso de você, ocê e cê em Belo Horizonte: um estudo em tempo aparente e em tempo real*. Tese de Doutorado. Universidade Federal de Minas Gerais, Belo Horizonte, 2006.

PERINI, M. A. *Gramática do português brasileiro*. São Paulo: Parábola, 2010.

PETERSEN, C. A tripartição pronominal e o estatuto das proformas cê, ocê e você. In: *DELTA* vol. 24, n. 2, 2008

PETRUCCI, L.; Il problema delle Origini e i più antichi testi italiani. In: L. Serianni; P. Trifone (Eds.). *Storia della lingua italiana*. Torino: Einaudi, 1992-1994, vol. III, p. 5-74.

POGGI SALANI, T.; La Toscana. In: F. Bruni (Ed.). *L'italiano nelle regioni*. Torino: UTET, 1992, vol I, p. 412-461.

POLIZIANO A.; Epistola proemiale alla Raccolta Aragonese (1477) In: G. Contini, *Letteratura italiana del Quattrocento*, Sansoni, Firenze, 1976

PONTES, E. *O tópico no português do Brasil*. Campinas: Pontes, 1987.

RAMOS, J. M. ; SOUZA, Elizete ; VITRAL, L. T.; A propósito dos pronomes nominativos no PB. Apresentação de Trabalho/Comunicação: PHPB- Londrina, maio de 2007.

RAMOS, J. O uso das formas você, ocê e cê no Dialeto Mineiro. In: HORA, D. (org.) *Diversidade Lingüística no Brasil*. João Pessoa: Ideia, 1997.

RAND, D. & SANKOFF, D.; *GoldVarb: A variable rule application for the Macintosh*. Montreal, Canada: Centre de recherches mathématiques, Université de Montréal, 1990.

RASO T.; MELLO H.; *C-ORAL-BRASIL I. Corpus de referência do português brasileiro falado*. I edição. Belo Horizonte, Editora UFMG, 2012.

RASO, T.; MITTMANN, M. M.; Validação estatística dos critérios de segmentação da fala espontânea no corpus C-ORAL-BRASIL. In: *Revista de Estudos da Linguagem*, v. 17, n. 2, p. 73-91, 2009. Disponível em: http://relin.lettras.ufmg.br/revista/upload/17-2_04.pdf

RASO, T.; O corpus C-ORAL-BRASIL. In T. Raso & H. Mello (eds), *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal*. Belo Horizonte: UFMG, 55-90, 2012

RASO, T.; O corpus C-ORAL-BRASIL. In: RASO T.; MELLO H. (Org.). In: *C-ORAL-BRASIL I. Corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora UFMG, 2012, p. 55-90.

RENZI L. e VANELLI L.; I pronomi soggetto in alcune varietà romanze. In: *Scritti linguistici in onore di G.S. Pellegrini*, Pisa, Pacini, 1983.

RENZI L. Fiorentino e italiano: storia dei pronomi personali soggetto, In: *Italia linguistica. Idee, storia, strutture*, a cura di F. Albano Leoni, D. Gambarara, F. Lo Piparo, R. Simone, Bologna, il Mulino, pp. 223-239. 1983.

RENZI, L.; “I pronomi soggetto in due varietà substandard: fiorentino e français avancé”. In: *Zeitschrift für romanische Philologie*. Ed. by Polzin-Haumann, Claudia / Schweickard, Wolfgang, Volume 108, Issue 1-2 (Jan 1992), De Gruyter, 1992

RENZI, L.; “Fiorentino e italiano: storia dei pronomi personali soggetto”. In: *Italia linguistica. Idee, storia, strutture*, a cura di F. Albano Leoni et al., Bologna, il Mulino, pp. 223-239, 1983

RENZI, L.; SALVI, G.; CARDINALETTI, A. (Orgs.). *Grande grammatica italiana di consultazione*. Nuova edizione. Bologna: Il Mulino, 2001. v. 1.

RICCA, Davide. *Italianizzazione dei dialetti*. Consultabile al sito: <http://www.treccani.it/enciclopedia/italianizzazione-dei-dialetti/> acesso em 2 de dezembro de 2013

RIZZI, L. On the status of subject clitics in Romance. In: JAEGGLI, O. ; CORVALAN, C. S. (orgs.). *Studies in Romance linguistics*. Dordrecht: Foris, 1986.

ROBINSON, J. S.; LAWRENCE, H. R.; TAGLIAMONTE, S. A. *GOLDVARB 2001: a*

ROBINSON, J. S.; LAWRENCE, H. R.; TAGLIAMONTE, S. A.; *GOLDVARB 2001: a multivariate analysis application for Windows*. Heslington: University of York, 2001.

ROMANO, R. (Ed.); *Leon Battista Alberti. I libri della famiglia*. Torino: Einaudi, 1994.

ROYSTON P.; An extension of Shapiro and Wilk's W test for normality to large samples. In: *Applied Statistics*, 31, 115–124, 1982

SABATINI, F.; L'italiano dell'uso medio: una realtà tra le varietà linguistiche italiane, In: *Gesprochenes Italienisch in Geschichte und Gegenwart*, a cura di G. Holtus & E. Radtke, Tübingen, Gunter Narr, pp. 154-184, 1985.

SABATINI, F.; Volgare civile e volgare cancelleresco nella Napoli angioina. In: Trovato, Paolo (Ed.). *Lingue e culture dell'Italia meridionale (1200-1600)*. Roma: Bonacci, 1993.

SALLES, M.; *Pronomes de tratamento do interlocutor no português brasileiro: um estudo de pragmática histórica*. 2001. - 246 f. Tese (Doutorado em Linguística) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2001.

SANGA, G. (Ed.). *Koinè in Italia dalle origini al Cinquecento*. Bergamo: Lubrina, 1990. Sauerländer, Verlag für deutsch-italienische Studien, pp. 13-32, 1998

SCARANO, A. (a cura di) (2005). *Corpus di confronto diacronico LABLITA*. Consultabile al sito: <http://lablita.dit.unifi.it/corpora/descriptions/corpusconfronto/>

SCARANO, A., MONEGLIA, M.; Il Corpus Stammerjohann. Il primo corpus di italiano parlato, in rete nella base dati di LABLITA. In M. Pettorino (a cura di), *Atti del Congresso Internazionale "La comunicazione Parlata"*, Napoli 23-25 Febbraio 2006.

SCARANO, A., SIGNORINI, S.; Corpus linguistics and diachronic variability. A study on Italian spoken language corpora from the 1960s until nowadays. In: J. Kabatek, C. D. Pusch, W. Raible (eds) *Korpora und diachrone Sprachwissenschaft / Romance Corpus Linguistics II: Corpora and Diachronic Linguistics*. Tübingen: Narr, pp. 191-202, 2005.

SCHIEL, F. et al. *The validation of speech corpora*. 2004. Disponível em: <<http://www.phonetik.uni-muenchen.de/forschung/BITS/TP2/Cookbook/>>. Acesso em: 20 dez. 2012.

SERIANNI, L. & TRIFONE, P. (a cura di) *Storia della lingua italiana*, Torino, Einaudi, 3 voll., vol 1° (I luoghi della codificazione) e vol. 2° (Scritto e parlato), 1993-1994.

SERIANNI, L.; *Il secondo Ottocento*. Bologna: Il Mulino, 1990.

SGRILLI, P.; L'espansione del Toscano nel Trecento. In: Gensini, Sergio. *La Toscana nel secolo XIV. Caratteri di una civiltà regionale*. Pisa: Pacini, 1988.

SHAPIRO, S. S.; WILK, M. B.; "An analysis of variance test for normality (complete samples)". *Biometrika* 52 (3-4): 591-611. doi:10.1093/biomet/52.3-4.591. JSTOR 2333709. MR 205384. p. 593, 1965

SIGNORINI, S., TUCCI, I.; Il restauro e l' archiviazione elettronica del primo corpus di italiano parlato: il corpus Stammerjohann. De Dominicis, L. Mori, M. Stefani (a cura di) (2005) *Costituzione, Gestione e restauro di corpora vocali*, Atti delle XIV Giornate del GFS, Collana degli atti dell'associazione italiana di acustica. Viterbo, 4-6 dicembre 2003. Roma: Esagrafica, pp. 119-126.

SILVA, Maria Cristina F.; *A posição sujeito no português brasileiro: frases finitas e infinitivas*. Campinas: Ed. da UNICAMP, 1996.

SIMONE, R., *Stabilità e instabilità nei caratteri originali dell'italiano*, In: Sobrero 1993, pp. 41-100.

- SOBRERO, Alberto A. *Italianization of the dialects*, in *The dialects of Italy*, edited by M. Maiden & M. Parry, London - New York, Routledge, pp. 412-418, 1997
- SPRENT, P.; SMEETON, N.C.; *Applied Nonparametric Statistical Methods*. Chapman Hall/CRC, 2001
- STAMMERJOHANN, H.; Strukturen der Rede. Beobachtungen an der Umgangssprache von Florenz. In: *Studi di Filologia Italiana*, XXVIII, 295-397, 1970.
- TAVONI, M.; *Latino, Grammatica, Volgare. Storia di una questione umanistica*. Padova: Antenore, 1984.
- TESI, R.; *Storia dell'italiano. La formazione della lingua comune dalle origini al Rinascimento*. Bari: Laterza, 2001.
- TRIFONE, P. Roma e il Lazio. In: F. Bruni (Ed.). *L'italiano nelle regioni*. Torino: UTET, 1992, vol I, p. 540-593.
- TROUVAIN, J. *Tempo Variation in Speech Production: Implications for Speech Synthesis*. Saarbrücken: Institute of Phonetics, Saarland University, 2004. (PHONUS 8)
- TROVATO, P.; *Il primo Cinquecento*. Bologna: Il Mulino, 1995.
- TUCCI, I. "Obiter dictum": la funzione informativa delle unità parentetiche. In: *Atti del Convegno Internazionale G.S.C.P. "La comunicazione parlata"*. Napoli, 23-25.02 2009
- TUCCI, I. L'inciso: caratteristiche morfosintattiche e intonative in un corpus di riferimento. In: LEONI, F. A. et al. (Orgs.). *Atti del Convegno "Il parlato italiano"*. Napoli, 13-15.02 2003. Napoli: M. D'Auria, 2004, p. 1-14
- ULIVI, E.; Le scuole d'abaco e l'insegnamento della matematica a Firenze nei secoli XIII-XVI. In: Freguglia, Paolo; Pellegrini, Luigi; Paciocco, Roberto. *Scienze matematiche e insegnamento in epoca medioevale*. Napoli: Edizioni scientifiche italiane, 2000.
- VALE, H. P. *A unidade informacional de Parentético no português do Brasil: uma análise baseada em corpus*. 2010. 149 f. (Mestrado em Estudos Linguísticos) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2009
- VIANELLO, V.; Editori ed intellettuali a Venezia nel Cinquecento. In: *Quaderni veneti*, 1992, p. 175-191.
- VIEGAS, M. C.; *Alçamento de vogais médias pretônicas: uma abordagem sociolingüística*. 1987. Dissertação (Mestrado em Lingüística) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 1987.
- VITALE, M.; *La questione della lingua*. Palermo: Palumbo, 1978.
- VITRAL, L. A forma cê e a noção de gramaticalização. In: *Revista de Estudos da Linguagem*, Ano 5 (4), 1996.

VITRAL, L., RAMOS, J. *Réplica a Petersen* (2008). A tripartição pronominal e o estatuto das proformas Cê, Ocê e Você. In: *DELTA* vol. 24, n. 2, 2008.

VITRAL, L.; RAMOS, J. *Gramaticalização: uma abordagem formal*. Belo Horizonte: Faculdade de Letras FALE/UFMG, 2006.

ZWICKY, A. M.; Clitics and particles. In: *Language: Journal of the Linguistic Society of America*. vol. 61, n.02. USA: Baltimore, 1985.