UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE LINGUÍSTICA:
PROGRAMA DE PÓS-GRADUAÇÃO EM SEMIÓTICA E LINGUÍSTICA GERAL

IVANETE BELÉM DO NASCIMENTO

**O uso variável do pronome de segunda pessoa
você(s)/cê(s) na cidade de São Paulo**

São Paulo
2011

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE LINGUÍSTICA:
PROGRAMA DE PÓS-GRADUAÇÃO EM SEMIÓTICA E LINGUÍSTICA GERAL

IVANETE BELÉM DO NASCIMENTO

O uso variável do pronome de segunda pessoa *você(s)/cê(s)* na cidade de São Paulo

Dissertação apresentada ao Programa de Pós-Graduação em Semiótica e Linguística Geral do Departamento de Linguística da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, para a obtenção do título de mestre em Letras.

Área de Concentração: Semiótica e Linguística Geral

Orientador: Prof. Dr. Ronald Beline Mendes

Exemplar Revisado.

De acordo: _____.

São Paulo
2011

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Catálogo na Publicação
Faculdade de Filosofia, Letras e Ciências Humanas

NASCIMENTO, Ivanete Belém do.

O uso variável do pronome de segunda pessoa *você(s)/cé(s)* na cidade de São Paulo / Ivanete Belém do Nascimento; Orientador: Prof. Dr. Ronald Beline Mendes – São Paulo, 2011.

Dissertação (Mestrado – Programa de Pós-Graduação em Semiótica e Linguística Geral. Área de Concentração: Semiótica e Linguística Geral) – Faculdade de Filosofia, Letras e Ciências Humanas.

1. Linguística 2. Sociolinguística 3. Teoria da Variação e da Mudança 4. Pronomes

IVANETE BELÉM DO NASCIMENTO

O uso variável do pronome de segunda pessoa *você(s)/cê(s)* na cidade de São Paulo

Dissertação apresentada ao Programa de Pós-Graduação em Semiótica e Linguística Geral do Departamento de Linguística da Universidade de São Paulo, para obtenção do título de Mestre em Linguística.
(versão revisada)

Área de Concentração: Semiótica e Linguística Geral

Aprovado em: 17 / 02 / 2011

BANCA EXAMINADORA

Prof. Dr. Ronald Beline Mendes

Instituição: FFLCH – USP

Presidente

Profa. Dra. Maria Marta Pereira Scherre

Instituição: Universidade Federal do Espírito Santo (UFES) e Universidade de Brasília (UnB)

Titular

Prof. Dr. James Walker

Instituição: York University (Canadá)

Titular

*Por que Deus permite
que as mães vão-se embora?
Mãe não tem limite,
é tempo sem hora,
luz que não apaga
quando sopra o vento
e a chuva desaba [...]*

(“Para sempre” – Carlos Drummond de Andrade)

À minha querida mãe, Fátima Belém...
onde quer que você se encontre...

Agradecimentos

Ao meu orientador, o Prof. Dr. Ronald Beline Mendes, pela orientação cuidadosa e minuciosa desde minha Iniciação Científica, pelo incentivo a prosseguir com os estudos acadêmicos diante das adversidades da vida, pela atenção e paciência ao longo desses anos.

Ao CNPq, agência que financiou este estudo.

À Profa. Dra. Marli Quadros Leite, que gentilmente permitiu o acesso às gravações do *corpus* do *NURC-SP*, cedendo-me parte de seu tempo.

Às Profas. Dras. Marta Scherre e Ângela Rodrigues, por disponibilizarem materiais bibliográficos importantes para o desenvolvimento deste estudo.

Ao Prof. Dr. Gregory Guy e ao Prof. Dr. Marcos Lopes, pelos valiosos comentários e sugestões, quando do exame de qualificação deste mestrado.

Aos professores Anthony Julius Naro, Marta Scherre e Scott Shwenter por sugestões à minha pesquisa, em ocasião de palestras na USP em 2009. Em especial, agradeço ao Prof. Dr. James Walker e à Profa. Dra. Marta Scherre, cujos comentários foram de grande importância para a versão final desta dissertação.

Aos funcionários do Departamento de Linguística, Érica, Robson, Ben Hur e Renata, por ajudas que seria impossível enumerar.

Ao Prof. Dr. Waldir Beividas e à Profa. Dra. Evani Viotti, coordenadores, durante o período do meu mestrado, do excelente Programa de Pós-Graduação em Semiótica e Linguística Geral da Universidade de São Paulo; e à Profa. Dra. Ana Scher, então chefe do Departamento de Linguística desta instituição.

Aos queridos colegas do GESOL (Grupo de Estudos e Pesquisa em Sociolinguística): Livia Oushiro, Melissa Cruz, Alexandre Sobreiro, Camila Barbosa, Mariane Biler, Larissa Soriano, Dayane Almeida, Rafael Rocha, Isabella Franchine e aos ex-integrantes Hernâni Macedo e Rafael Coelho. Agradeço, sobretudo, à Livia Oushiro, pela companhia nos congressos, pela prontidão nas leituras atentas de meus trabalhos, pelo auxílio no armazenamento e na

codificação dos dados e pela digitalização das entrevistas mais antigas do *corpus* do GESOL.

A todos os professores que contribuíram para minha formação desde o ensino fundamental até o mestrado. Todos foram importantes, mas gostaria de destacar, sobretudo, os professores Ronald Beline, Marli Quadros e Evani Viotti (“responsáveis” por despertar meu interesse nos estudos linguísticos); e aos professores Marcos Lopes, Ana Scher, Paulo Chagas, Margarida Petter, Esmeralda Negrão e Ataliba de Castilho (grandes exemplos de pesquisadores e docentes, a quem devo muito do pouco que adquiri).

Ao linguista e amigo Maurício Souza, pela leitura desta dissertação e por suas valiosas sugestões.

Aos amigos Luciana Silva, Patrícia Kondo, Melissa Yoshioka, Melissa Cruz, Livia Oushiro, Sueli Ramos, Fabiana Moretti, Eneida Leal, Délcia Ono, Leda Cisotto, Gláucia Fernandes e Maria José Baraldi, o meu muito obrigada pela amizade sincera e pela presença em minha vida.

Às queridas amigas Fabiana Maria, Pâmella Piera e Fátima Almeida Baia e aos meus tios Santo Belém e Marisa pelo imenso apoio no momento mais difícil de minha vida: a partida de minha mãe, no início de 2010. Agradeço também às amigas de minha mãe Rosa, Cleonice (Cléo) e Domingas. Sem a presença de vocês, certamente eu não teria conseguido chegar até aqui...

À ex-professora e amiga, Maria Cláudia (Claudinha), e aos amigos Marcelo Elias, Maurício Souza, Cláudia Brandão, Sandra Casemiro, Ana Lúcia e Daniel Peres, também pelo apoio e pelo carinho.

Ao Kiko, pelo companheirismo... um serzinho muito especial, que me prova a cada dia que “ser humano” não necessariamente se expressa apenas por “seres humanos”.

À minha querida mãe, Fátima Belém, pessoa mais importante de minha vida, luz que iluminou meu caminho e que certamente continuará a me guiar de algum lugar deste universo. A você, mãe, que partiu tão precoce e inesperadamente, deixando um imenso vazio, uma profunda saudade...

RESUMO

Esta dissertação trata da variação no uso do pronome de segunda pessoa (singular e plural) na cidade de São Paulo. Duas formas alternativas são empregadas nessa comunidade de fala: a variante plena *você(s)* e a forma foneticamente reduzida *cê(s)*. A pesquisa é desenvolvida de acordo com os pressupostos teórico-metodológicos da Sociolinguística Variacionista. Foram desenvolvidas análises sincrônica e diacrônica com dados extraídos de duas amostras (*NURC-SP-1970* e *GESOL-SP-2000*).

Os resultados evidenciam um equilíbrio na distribuição das duas variantes nessa comunidade de fala. Embora se verifique uma típica mudança em progresso (tempo aparente) na década de 1970, a análise dos anos 2000 revela um caso de variação estável na comunidade de fala.

Nos anos 2000, a variante inovadora, a forma reduzida *cê*, tende a ser favorecida pelos informantes mais jovens e pelas mulheres acima de 50 anos de idade. *Cê* é desfavorecida entre os indivíduos de uma faixa etária intermediária (entre 35 e 45 anos) – o que pode estar relacionado a questões de monitoramento da fala e ao mercado linguístico (Paiva & Duarte, 2003). Adicionalmente, *cê* tende a ser evitado entre os indivíduos mais escolarizados; e é favorecido em interações cujos informantes são familiares ou amigos.

De um ponto de vista linguístico, o emprego das variantes é correlacionado pelo *Princípio do Contorno Obrigatório* e por questões morfosintáticas, semânticas e discursivo-cognitivas, com especial atenção à referência do pronome. Os resultados corroboram a hipótese de cliticização da variante reduzida, mas fornecem um contra-argumento à correlação direta entre “erosão fonética” e “abstratização semântica”, defendida na literatura sobre gramaticalização.

Palavras-chave: Sociolinguística Variacionista, pronomes de segunda pessoa, referenciação, gramaticalização, *Princípio do Contorno Obrigatório*.

ABSTRACT

This master thesis analyzes the variation in the use of the second person pronoun (singular and plural) in the city of São Paulo. Two alternative forms are employed: *você* 'you', and a phonetically reduced form *cê* 'you'. The research is developed according to the theoretical and methodological framework of Variationist Sociolinguistics. Both synchronic and diachronic multivariate analyses are pursued, with data extracted from two samples (*NURC-SP-1970* and *GESOL-SP-2000*).

The results show a balance in the distribution of the two variants in the speech community. Although it was observed a typical change in progress (apparent time) in the 1970s, the analysis of the 2000s data reveals a case of stable variation in the speech community.

In the 2000s, the innovative, phonetically reduced variant tends to be favored by younger people and women over 50 years old. *Cê* is disfavored among individuals between 35 and 45 years – which can be related to issues of speech monitoring and the linguistic market (Paiva & Duarte, 2003). In addition, *cê* tends to be avoided by those whose level of education is higher, and is favored in conversations between informants who are friends or relatives.

From a linguistic perspective, the use of variants is correlated by the *Obligatory Contour Principle (OCP)* and by morphosyntactic, semantic and discursive-cognitive factors, with special attention to the reference of the pronoun. Results confirm the hypothesis of cliticization of the reduced variant, but reveal a counterargument for the direct correlation between "phonetic erosion" and "semantic abstraction", which has been claimed in the literature on grammaticalization.

Keywords: Variationist Sociolinguistics, second person pronouns, reference, grammaticalization, *Obligatory Contour Principle (OCP)*.

Lista de tabelas

	Pág.
Tabela 1: Formas de tratamento em peças de teatro nos séculos XVIII e XIX (Lopes & Duarte, 2003).....	13
Tabela 2: <i>Você, ocê e cê</i> conforme a referência e a Faixa etária em Belo Horizonte (Ramos, 1997).....	23
Tabela 3: Recursos de indeterminação do sujeito em Belo Horizonte (Alves, 1998).....	26
Tabela 4: Emprego de <i>Cê</i> em relação à referência e à faixa etária em BH em 2002 (Peres, 2006).....	27
Tabela 5: Emprego de <i>Cê</i> em relação à faixa etária do informante em BH 1982-2002 (Peres, 2006).....	28
Tabela 6: Variantes empregadas nas amostras do NURC-SP-1970 e GESOL-SP-2000.....	82
Tabela 7: Distribuição geral de <i>CÊ</i> conforme a Função/Estrutura Sintática e o contexto.....	89
Tabela 8: Envelope de variação.....	93
Tabela 9: Duração (D) e Intensidade (I) das Formas <i>se, cê</i> e <i>Zé</i> (Vital, 2006a).....	98
Tabela 10: Duração de <i>cê</i> (pronome) e <i>Cê</i> (nome) – adaptada de Ciriaco et al., 2004).....	98
Tabela 11: Duração de <i>cê</i> (pronome), <i>CÊ</i> (nome) e <i>se</i> (CIRÍACO et al., 2004).....	99
Tabela 12: Referenciações do pronome <i>você/cê</i> e traços semântico-referenciais	117
Tabela 13: Pronomes de 2ª pessoa empregados pelos paulistanos (GESOL-SP-2000).....	130
Tabela 14: Emprego de <i>CÊ</i> conforme a função e a estrutura sintática.....	131
Tabela 15: <i>Você</i> e <i>Cê</i> em contextos variáveis – informantes e documentadores (GESOL-SP-2000).....	132
Tabela 16: Distribuição de <i>você/ocê/cê</i> em diferentes comunidades de fala: BH, DF e SP.....	133

Tabela 17: Correlação de fatores sociais ao emprego de <i>você/cê</i> (Infs.- GESOL-SP-2000).....	134
Tabela 18: Emprego de <i>CÊ</i> conforme a Faixa etária do Informante e seu Sexo/Gênero (Inf-GESOL-SP-2000).....	136
Tabela 19: Emprego de <i>você/cê</i> e relação entre os interlocutores (Informantes-GESOL-SP-2000).....	138
Tabela 20: Distribuição dos informantes conforme sua relação com o interlocutor (GESOL-SP- 2000).....	139
Tabela 21: Emprego de <i>você/cê</i> conforme a relação do INF com o DOC (GESOL-SP-2000).....	140
Tabela 22: Uso de <i>CÊ</i> específico e relação entre os interlocutores (Infs e Docs – GESOL-SP-2000).....	141
Tabela 23: Uso de <i>CÊ</i> com referência específica de acordo com o grau de intimidade entre os interlocutores (Infs e Docs – GESOL-SP-2000).....	142
Tabela 24: <i>Você/Cê</i> em relação ao sexo/gênero dos interlocutores (Infs – GESOL-SP-2000).....	142
Tabela 25: Correlações de natureza linguística ao emprego de <i>CÊ</i> (Informantes – anos 2000).....	143
Tabela 26: Emprego de <i>CÊ</i> conforme a atuação de Fatores linguísticos II (Infs.-GESOL-2000).....	144
Tabela 27: <i>Você/Cê</i> de acordo com a sua referência (específica x genérica) – (Infs – GESOL-SP-2000).....	145
Tabela 28: Emprego de <i>CÊ</i> conforme o tipo de discurso e a referência do pronome (Infs-GESOL-SP-2000).....	146
Tabela 29: Referência do pronome em discursos não-reportados (oposição binária) (Infs.-GESOL-2000).....	147
Tabela 30: Correlação da referência ao emprego de <i>você/cê</i> (3 fatores) – (Infs – GESOL-SP-2000)	147
Tabela 31: Emprego de <i>você/cê</i> de acordo com sua referenciação no discurso (Infs.-GESOL-SP-2000) – (Análise expandida da referência I).....	148
Tabela 32: Emprego de <i>você/cê</i> de acordo com sua referenciação no discurso II (Infs.-GESOL-SP-2000).....	149

Tabela 33: Emprego de <i>CÊ</i> conforme a sua referenciação e a presença de constituintes /SE/ adjacentes ao pronome (Infs.-GESOL-SP-2000).....	150
Tabela 34: Emprego de <i>você/cê</i> de acordo com sua referenciação no discurso (Infs.-GESOL-SP-2000) – (Análise expandida da referência II).....	151
Tabela 35: Usos genéricos de <i>CÊ</i> em relação à sua referenciação no discurso (Inf.-GESOL-SP-2000).....	153
Tabela 36: Diferentes abordagens da referência do <i>você/cê</i> (Infs – GESOL-SP-2000).....	154
Tabela 37: Emprego de <i>você</i> e <i>cê</i> conforme a referência do pronome e o Sexo/Gênero dos informantes (GESOL-SP-2000).....	157
Tabela 38: Emprego de <i>CÊ</i> de acordo com a referência (específica x genérica) e a Faixa etária – (Inf.-GESOL-SP-2000).....	158
Tabela 39: Uso de <i>você</i> , <i>ocê</i> e <i>cê</i> de acordo com a referência e a faixa etária em BH (Ramos, 1997).....	159
Tabela 40: Uso de <i>CÊ</i> de acordo com a referência e o nível de escolaridade (Infs.-GESOL-SP-2000).....	159
Tabela 41: Uso de <i>Você/Cê</i> específico conforme o sexo, a idade, a escolaridade, o discurso e a flexão do pronome (Infs e Docs – GESOL-SP-2000).....	160
Tabela 42: Uso de <i>CÊ</i> com referência específica de acordo com o papel do falante na entrevista – (Inf-Doc – GESOL-SP-2000).....	161
Tabela 43: Uso de <i>você</i> e <i>cê</i> quanto ao Paralelismo das variantes no discurso – (Inf.-GESOL-SP-2000).....	162
Tabela 44: Emprego de <i>você/cê</i> e paralelismo das variantes – Análise concisa (Inf.-GESOL-SP-2000).....	164
Tabela 45: Distribuição de <i>você</i> e <i>cê</i> conforme o paralelismo de formas reduzidas (Infs – GESOL-SP-2000).....	164
Tabela 46: Uso de <i>você</i> e <i>cê</i> conforme a presença de formas reduzidas adjacentes ao pronome – análise concisa (Infs – GESOL-SP-2000).....	165
Tabela 47: Emprego de <i>você/cê</i> conforme a frequência do verbo (Infs – GESOL-SP-2000).....	166

Tabela 48: Uso de <i>Você</i> e <i>cê</i> conforme o contexto seguinte ao pronome (Infs – GESOL-SP-2000).....	166
Tabela 49: Uso de <i>você/cê</i> e Tonicidade da sílaba seguinte ao pronome (Inf – GESOL-SP-2000).....	167
Tabela 50: <i>Você/cê</i> conforme o contexto antecedente ao pronome (Infs – GESOL-SP-2000).....	169
Tabela 51: Uso de <i>você</i> e <i>cê</i> conforme a Tonicidade da sílaba antecedente ao pronome – (Infs – GESOL-SP-2000).....	169
Tabela 52: Distribuição de <i>você</i> e <i>cê</i> quanto à presença de constituintes /SE/ adjacentes ao pronome – (Infs – GESOL-SP-2000).....	170
Tabela 53: Uso de <i>você/cê</i> relativamente à presença de /SE/ adjacente (Inf – GESOL-SP-2000).....	170
Tabela 54: Distribuição de <i>você</i> e <i>cê</i> conforme o tipo de constituinte interveniente entre o pronome e o verbo (Inf – GESOL-SP-2000).....	171
Tabela 55: Uso de <i>Você</i> e <i>cê</i> de acordo com a mobilidade do constituinte interpolado (Infs – GESOL-SP-2000).....	172
Tabela 56: Uso de <i>Você /cê</i> e extensão fônica do constituinte interveniente (Infs- GESOL-SP-2000).....	173
Tabela 57: Uso de <i>você/cê</i> e coalescência do pronome ao verbo (Infs – GESOL-SP-2000).....	173
Tabela 58: Distribuição das variantes <i>você</i> e <i>cê</i> de acordo com o tipo de oração (Infs – GESOL-SP-2000).....	174
Tabela 59: Emprego de <i>Você</i> e <i>cê</i> conforme o tipo de oração (Infs – GESOL-SP-2000).....	174
Tabela 60: Uso de <i>Você</i> e <i>cê</i> em orações declarativas e interrogativas (Infs – GESOL-SP-2000).....	175
Tabela 61: Emprego do pronome de segunda pessoa conforme a referência do pronome e o tipo de oração – (Infs – GESOL-SP-2000).....	175
Tabela 62: Uso de <i>CÊ</i> conforme a referência do pronome e o tipo de oração (Infs e Docs: GESOL-SP-2000).....	176
Tabela 63: Distribuição Geral das variantes em 1970 e nos anos 2000.....	177
Tabela 64: Correlação de variáveis sociais ao uso de <i>você</i> e <i>cê</i> (1970-2000).....	178

Tabela 65: Distribuição de <i>você</i> e <i>cê</i> quanto à sua referenciação (NURC-SP-1970).....	180
Tabela 66: Emprego de <i>CÊ</i> em relação à sua referência (específica x genérica) – (1970-2000).....	182
Tabela 67: Emprego da variante <i>CÊ</i> na década de 1970 e nos anos 2000 em relação à referência do pronome.....	183
Tabela 68: Efeito do Paralelismo das variantes no uso de <i>CÊ</i> em 1970 e nos anos 2000.....	183
Tabela 69: Paralelismo de formas reduzidas adjacentes e emprego de <i>CÊ</i> – (1970-2000).....	184
Tabela 70: Emprego de <i>CÊ</i> relativamente à frequência do verbo que acompanha o pronome (1970-2000).....	184
Tabela 71: Emprego de <i>CÊ</i> conforme a Tonicidade da sílaba seguinte (1970-2000).....	185
Tabela 72: Emprego de <i>CÊ</i> conforme a Tonicidade da sílaba antecedente (1970-2000).....	186
Tabela 73: Emprego de <i>CÊ</i> e presença de constituintes /SE/ adjacentes ao pronome (1970-2000).....	187
Tabela 74: Emprego da variante reduzida do pronome quanto à sua flexão (1970-2000).....	187
Tabela 75: Coalescência do pronome ao verbo em 1970 e nos anos 2000.....	188
Tabela 76: Emprego de <i>Cê</i> conforme o tipo de oração (1970-2000).....	189

Lista de Quadros

	Pág.
<i>Quadro 1:</i> <i>Você e Tu</i> no PB e no PE atualmente (Ramos, 1998).....	14
<i>Quadro 2:</i> Enfraquecimento do paradigma de flexões verbais no PB.....	15
<i>Quadro 3:</i> Mudanças em cadeia no PB.....	16
<i>Quadro 4:</i> A referência do <i>você/cê</i> e seus subtipos.....	70
<i>Quadro 5:</i> Amostra GESOL-SP-2000 (documentadores paulistanos).....	78
<i>Quadro 6:</i> Amostra GESOL-SP-2000 (Informantes).....	79
<i>Quadro 7:</i> Amostra NURC-SP-1970.....	81
<i>Quadro 8:</i> Variável e variantes do pronome de segunda pessoa.....	83
<i>Quadro 9:</i> Restrições sintáticas ao emprego de <i>CÊ</i> (exemplos de Vitral, 1996).....	84
<i>Quadro 10:</i> Restrições ao emprego de <i>CÊ</i> em orações clivadas.....	87
<i>Quadro 11:</i> Grupos de fatores extralinguísticos.....	97
<i>Quadro 12:</i> Graus de indeterminação da referência do pronome <i>você/cê</i>	108
<i>Quadro 13:</i> <i>Você/cê</i> com referênciação P1.....	110
<i>Quadro 14:</i> <i>Você/cê</i> com referênciação P4-exclusiva.....	111
<i>Quadro 15:</i> <i>Você/cê</i> com referênciação P4 ‘inclusiva’	113
<i>Quadro 16:</i> <i>Você/cê</i> com referênciação P3/P6.....	114
<i>Quadro 17:</i> <i>Você/cê</i> com referênciação Híbrida.....	114
<i>Quadro 18:</i> Usos expletivos de <i>você/cê</i>	115
<i>Quadro 19:</i> Grupos de fatores linguísticos: Fonético-Fonológicos, Morfossintáticos.....	127
<i>Quadro 20:</i> Grupos de fatores linguísticos: Variáveis Semânticas e discursivo-cognitivas.....	128
<i>Quadro 21:</i> Uso de <i>você/cê</i> conforme sua referênciação no discurso.....	152

Lista de Gráficos

	Pág.
Gráfico 1: Emprego de CÊ conforme a faixa etária do informante (GESOL-SP-2000).....	135
Gráfico 2: Emprego de CÊ conforme o Sexo / Gênero e a Faixa Etária do informante (GESOL-SP-2000).....	137
Gráfico 3: Emprego de CÊ de acordo com sua referenciação no discurso (Infs-GESOL-SP-2000).....	148
Gráfico 4: Paralelismo das variantes pronominais (Paralelismo I).....	165
Gráfico 5: Paralelismo de formas reduzidas adjacentes (Paralelismo II).....	165
Gráfico 6: Emprego de CÊ conforme a faixa etária do informante (1970-2000)....	179
Gráfico 7: Emprego de CÊ entre as mulheres conforme a faixa etária do informante (1970-2000).....	180
Gráfico 8: Emprego de CÊ entre os homens conforme a faixa etária do informante (1970-2000).....	180
Gráfico 9: Emprego de CÊ conforme a referência do pronome (1970-2000).....	182
Gráfico 10: Emprego de CÊ Específico conforme a Idade (1970-2000).....	183
Gráfico 11: Emprego de CÊ Genérico conforme a Idade (1970-2000).....	183
Gráfico 12: Tonicidade da sílaba antecedente e emprego de CÊ (1970-2000).....	186
Gráfico 13: Interpolação de constituintes entre a variante reduzida CÊ e o verbo (1970-2000).....	188

Sumário

Agradecimentos.....	vi
Resumo.....	viii
Abstract.....	ix
Lista de Tabelas.....	x
Lista de Quadros.....	xv
Lista de Gráficos.....	xvi

INTRODUÇÃO.....	1
------------------------	----------

CAPÍTULO I - Pronomes de Segunda Pessoa no Português Brasileiro.....	9
---	----------

Introdução.....	9
1.1. Breve percurso histórico.....	9
1.2. Mudanças afins.....	14
1.3. Subsistemas pronominais de 2ª pessoa no PB.....	17

CAPÍTULO II - Pressupostos Teórico-Metodológicos.....	32
--	-----------

Introdução.....	32
2.1. Teoria da Variação e da Mudança Linguística.....	32
2.2. Gramaticalização.....	48
2.3. Concepções discursivo-pragmáticas para a referência.....	58
2.4. O Princípio do Contorno Obrigatório.....	70
2.5. Princípios discursivos e cognitivos.....	72

CAPÍTULO III - Análise Qualitativa.....	76
Introdução.....	76
3.1. As amostras de dados.....	76
3.2. Variantes e o envelope de variação.....	81
3.3. Grupos de Fatores.....	94
 CAPÍTULO IV - Análise Quantitativa.....	 129
Introdução.....	129
4.1. Análise sincrônica – Amostra GESOL-SP-2000.....	129
4.2. Análise diacrônica – dos anos 1970 aos anos 2000.....	177
 CONCLUSÃO.....	 190
 REFERÊNCIAS BIBLIOGRÁFICAS.....	 200
 ANEXO 1.....	 213
ANEXO 2.....	217

INTRODUÇÃO

Os pronomes de segunda pessoa, ao lado das formas de tratamento, têm sido objeto de intensa investigação sociolinguística nas últimas décadas. De uma maneira geral, tais estudos podem ser divididos em quatro linhas: (i) trabalhos que versam sobre a evolução das formas de tratamento no português e sobre a origem do tratamento *Vossa Mercê*; (ii) estudos que tratam da variação entre os pronomes *você* e *tu*; (iii) pesquisas que opõem uma forma de tratamento pronominal (*você/ocê/cê/tu*) a um tratamento mais formal (*o senhor /a senhora*); e (iv) trabalhos que focalizam a variação entre a forma plena do pronome *você* e suas variantes reduzidas *ocê/cê* em comunidades nas quais o *tu* não é uma forma empregada pelos falantes. É neste último enfoque que esta pesquisa se insere.

Entre os trabalhos que tratam da variação no uso dos pronomes de segunda pessoa no português brasileiro (doravante PB), a maioria concentra-se em comunidades de fala localizadas em Minas Gerais (Ramos, 1997, 2000; Alves, 1998; Coelho, 1999¹; Peres, 2006; Gonçalves, 2008) e em Brasília (Andrade, 2004; Lucca, 2005; Dias, 2007; Queiroz-Andrade, 2010). Além desses, foram desenvolvidas pesquisas no Rio de Janeiro (Paredes Silva, 2003); em Santos (Modesto, 2006); em estados da região sul (Loregian-Penkal, 2004; Bolívar, 2008), norte (Martins, 2010) e nordeste do país (Figueiredo, 2007; apud. Scherre, 2010-2013). Não se tem notícia, todavia, de estudos que focalizem esse fenômeno variável na cidade de São Paulo².

No presente estudo, desenvolve-se uma análise variacionista desse pronome pessoal na capital paulistana. Nessa variedade do PB, duas variantes são majoritariamente empregadas pelos falantes: uma forma pronominal plena *você(s)* e uma variante foneticamente reduzida *cê(s)*³. Constitui objeto dessa investigação,

¹ Apud. Gonçalves (2008).

² A única exceção é Corradello (1997), porém a autora não desenvolve um estudo variacionista.

³ O pronome de segunda pessoa do singular *tu* não é uma variante nessa comunidade de fala; a forma *ocê* é bastante rara, sobretudo com a função sintática de sujeito. Não foi objetivo deste estudo investigar o emprego do tratamento *o senhor/a senhora* nem uma variante zero (sujeitos nulos) – que, em princípio, também poderiam ser considerados formas alternativas das variantes em análise – v. capítulo III.

portanto, o pronome de segunda pessoa do singular e o pronome de segunda pessoa do plural.

Na maioria dos trabalhos sobre a variação *você/ocê/cê*, nem sempre foi explicitado se os dados de segunda pessoa do plural (*vo)cês* também eram incluídos na pesquisa⁴. Neste estudo, investigam-se ambas as formas, singular e plural, ainda que esta última ocorra em número bastante reduzido. Para facilitar a referência às variantes, serão empregados, na maioria das vezes, apenas os símbolos das formas singulares *você* e *cê*. Quando as análises focalizarem apenas um desses empregos (singular ou plural), isto será explicado.

Para o desenvolvimento deste estudo, adotam-se os pressupostos teórico-metodológicos da Sociolinguística Variacionista (Labov, 1972, 1994, 2001) – perspectiva teórica que concebe a variação como parte integrante do próprio sistema linguístico (Labov, 2006 [1966]; Weinreich et al., 1968; Guy, Ms. inédito) e que pressupõe que ela é regulada por fatores tanto internos (linguísticos), quanto externos (extralinguísticos).

Desenvolveu-se uma análise sincrônica e uma análise diacrônica em tempo real de curta duração (estudo do tipo *tendência* – Labov, 1994; Chambers, 2002; Bailey, 2002), mediante a comparação de duas amostras de língua falada, separadas por um intervalo de aproximadamente trinta anos. A primeira compreende 18 entrevistas no formato diálogo entre informante e documentador (DID), que integram o *corpus* do Projeto Norma Urbana Culta (NURC) da cidade de São Paulo (Castilho & Preti, 1986-1990).

A segunda amostra compreende 36 entrevistas sociolinguísticas gravadas entre 2003 e 2009, que fazem parte de um *corpus* de língua falada na cidade de São Paulo, organizado pelos integrantes do Grupo de Estudos e Pesquisa em Sociolinguística (GESOL), coordenado pelo Prof. Dr. Ronald Beline Mendes

⁴ Geralmente, os estudos davam a entender tratarem somente da segunda pessoa do singular, embora apresentassem exemplos com o pronome de segunda pessoa do plural (Ramos, 1997; Andrade, 2004, pp. 49-60).

(Departamento de Linguística – Universidade de São Paulo)⁵. Empregam-se aqui as denominações *NURC-SP-1970* e *GESOL-SP-2000*, para se fazer referência a estas duas amostras, respectivamente.

De uma perspectiva diacrônica, *você/cê* resulta da gramaticalização da expressão *vossa mercê* (“vosso favor”/“vossa graça”) – um item lexical que adquire um *status* gramatical, tornando-se uma forma de tratamento e, posteriormente, um pronome pessoal (cf. Faraco, 1996 e outros). Na continuidade desse processo, a redução *você* > *cê* tem sido analisada como correspondente à etapa de cliticização desse pronome (Vital, 1996, 2006b; Ramos, 1997) – conforme se prevê na escala *item lexical* > *item gramatical* > *clítico* > *afixo flexional* (Hopper & Traugott, 1993). Adicionalmente, uma série de restrições sintáticas (quase) categóricas em relação ao uso da variante reduzida do pronome (e.g., isolado, objeto direto, tópico, clivado, etc.) soma-se à hipótese anterior, indicando o caráter morfológico diverso de *cê* – seja ele um clítico (Vital, 1996), uma forma fraca (Petersen, 2008) ou um item em processo de cliticização (Vital, 2006b). Aqui, adota-se essa última posição.

Desse modo, dois objetivos gerais norteiam essa pesquisa:

- (i) investigar esse fato variável em uma comunidade de fala que ainda não constituiu objeto de estudo das pesquisas já desenvolvidas sobre o assunto, como é o caso da cidade de São Paulo; e
- (ii) verificar como se dá o encaixamento social e linguístico de um mesmo fenômeno variável em distintas comunidades de fala; ou seja, analisar se os mesmos fatores sociais e linguísticos que controlam a variação entre *você*, *ocê* e *cê* em Minas Gerais, por exemplo, atuam de modo semelhante em relação à variação *você/cê* na cidade de São Paulo, visto que se trata de um mesmo processo de gramaticalização.

⁵ Para que as duas amostras fossem “compatíveis”, no estudo em tempo real, utilizou-se parte da amostra *GESOL-SP-2000* (18 em vez de 36 entrevistas sociolinguísticas), cujos informantes têm nível superior de escolaridade, tal como no *corpus* do *NURC*. A ausência de uma amostra que retrate a língua falada por paulistanos menos escolarizados na década de 1970 impede um estudo diacrônico deste perfil sociolinguístico.

Introdução

Uma primeira contribuição desta pesquisa, portanto, é a de colaborar com estudos variacionistas e dialetológicos sobre o emprego dos pronomes pessoais de segunda pessoa no Brasil. A este respeito, por exemplo, um mapeamento do emprego desses pronomes no PB tem sido desenvolvido atualmente pela Profa. Dra. Maria Marta Pereira Scherre (UFES/UnB), em um projeto de pesquisa em parceria com o Prof. Dr. Anthony Julius Naro – Scherre (2010-2013, inédito).

Outra contribuição do presente estudo diz respeito à relação entre a Sociolinguística e a Teoria da Gramaticalização. A este respeito, a unidirecionalidade – considerada por vários estudiosos como inerente aos processos de gramaticalização – também pode sofrer influência de condicionamentos sociais (Zilles, 2004), que poderiam acelerar, frear ou mesmo retardar o avanço de um caso de gramaticalização.

No português europeu (PE), por exemplo, o processo de gramaticalização desse pronome seguiu outros rumos relativamente àqueles verificados no PB, pois parece ter se “estacionado” no item pleno *você*, cujo emprego restringe-se a determinados contextos e, muitas vezes, apresenta conotações negativas (cf. Paiva Boléo, 1946; apud. Ramos, 1998; Hammermüllerm, 1993; apud. Gonçalves, 2008: 88). O *tu* é a forma mais favorecida quando há maior grau de intimidade entre os interlocutores portugueses.

No PB, o pronome *tu* continua a ser uma forma produtiva em determinadas comunidades de fala, e seu emprego pode se correlacionar também ao grau de intimidade entre os interlocutores. Em determinadas comunidades de fala brasileiras, contudo, o *tu* inexistente como uma forma variante – como é o caso da cidade de São Paulo⁶.

A fim de demonstrar a sistematicidade da variação e seu encaixamento social e linguístico na comunidade paulistana, foram analisadas seis variáveis extralinguísticas: três de natureza social (sexo/gênero; faixa etária; e escolaridade) e três interacionais (sexo/gênero dos interlocutores; grau de intimidade entre os

⁶ A distribuição distinta entre os pronomes *você/cê* e *tu* no território brasileiro pode estar associada a questões históricas e às estruturas sócio-econômicas de cada região – v. Ramos (1998) para uma discussão da necessidade de estudos nesse campo.

interlocutores; e “papel” do falante na entrevista sociolinguística – documentador ou informante).

No plano linguístico, as variáveis (grupos de fatores) que condensam as hipóteses aventadas organizam-se em quatro grandes linhas:

(i) A atuação do *Princípio do Contorno Obrigatório (OCP)* como uma regra variável:

Foram investigados três grupos de fatores de natureza fonético-fonológica: tonicidade da sílaba seguinte ao pronome; tonicidade da sílaba antecedente ao pronome; e presença de constituintes semelhantes à variante reduzida *cê*. Parte-se do pressuposto de que *cê* – como uma forma intermediária entre elementos tônicos e átonos (cf. Ciriaco et al., 2004; Vitral, 2006a) – poderia ser preferencialmente empregada quando ao pronome se seguir uma sílaba tônica, evitando-se um “choque acentual” (Selkirk, 1984; apud. Abousalh, 1997). A mesma variante poderá ser inibida, contudo, na presença de sílabas idênticas ou semelhantes adjacentes a ela, como [sɪ], [se] e [sẽ].

(ii) A hipótese de cliticização de *cê*:

Dois grupos de fatores tentam dar conta dessa questão: a coalescência do pronome ao verbo e o tipo de oração (declarativa, interrogativa *Wh-que*, e outras interrogativas). Entende-se que, se a variante reduzida estiver se cliticizando, pode ser que a mesma tenda a vir adjacente ao verbo, uma vez que os clíticos são elementos que necessitam se “apoiar” em outro constituinte (cf. Hopper & Traugott, 1993); bem como em estruturas sintáticas com a posição de complementizador preenchida (i.e., interrogativas *wh-que*) – cf. Ramos (1997).

(iii) Atuação de questões discursivas e cognitivas:

O Paralelismo (Scherre, 1998) foi investigado em duas versões: paralelismo das variantes pronominais nas orações precedentes; e paralelismo de outras formas reduzidas adjacentes ao pronome. Além disso, verificou-se se verbos mais frequentes favoreceriam a redução do pronome.

(iv) A referência do pronome:

A referência do pronome *você/cê* tem recebido um enfoque “binário” nos diversos estudos variacionistas até então realizados (específica *versus* genérica). Vitral & Ramos (2006 [1999]), porém, destacam que, ao lado de uma referência genérica e uma específica, haveria empregos expletivos desse pronome pessoal. Corradello (1997) e Sousa (2008) também propõem referências distintas para o “você” genérico a partir de considerações pragmáticas e discursivas. Tais pesquisas, todavia, analisam estes “graus de genericidade” de um ponto de vista qualitativo ou computam o número de ocorrências do pronome nos contextos independentemente de sua pronúncia (plena ou reduzida); ou seja, não se trata de abordagens variacionistas. No presente estudo, propõe-se uma análise variacionista da correlação entre distintas referências desempenhadas pelo pronome e o emprego das variantes plena e reduzida. Com base em pressupostos teóricos da semântica (Cançado, 2004), da pragmática (Benveniste, 1991; Cervoni, 1989) e da análise do discurso (Maingueneau, 2004; Fiorin, 2008), foram estabelecidos nove empregos distintos do pronome quanto às pessoas envolvidas em sua referência.

Para isso, diferentes abordagens para a referência são testadas quantitativamente no capítulo IV, argumentando-se em favor de que uma análise “expandida” desse grupo de fatores é preferível qualitativa e quantitativamente a uma análise mais concisa.

Todas as questões anteriores visam a confirmar o caráter sistêmico desse fato variável, demonstrando que o emprego das variantes *você* e *cê* não é “aleatório” ou “livre”, mas sofre restrições tanto de caráter categórico, quanto probabilístico. Dessa forma, os interesses específicos desse estudo podem ser assim descritos:

- (i) investigar o encaixamento social e linguístico desse fenômeno variável na comunidade de fala paulistana; i.e., depreender quais fatores sociais e linguísticos correlacionam-se ao emprego do pronome de segunda pessoa nessa variedade do PB;

- (ii) ampliar as discussões sobre os fatos (sociais e linguísticos) que se correlacionam ao uso do pronome de segunda pessoa *você(s)* e da variante reduzida *cê(s)* nessa comunidade de fala, de maneira a desenvolver melhor compreensão acerca desse fenômeno;
- (iii) refinar algumas hipóteses investigadas na literatura, como é o caso da referência do pronome;
- (iv) analisar questões que até então não constituíram foco de interesse na maioria dos estudos que trataram desse caso de variação, como a atuação de questões fonético-fonológicas; e
- (v) analisar se este fenômeno constitui um caso de mudança em progresso ou de variação estável na cidade de São Paulo.

Essas discussões vêm estruturadas em quatro capítulos:

No primeiro capítulo, situa-se o objeto de estudo sincrônica e diacronicamente. O capítulo se inicia com um breve percurso histórico desse pronome, momento em que se revisa parte da literatura que trata dessa questão (Cintra, 1972; Faraco, 1996; Teixeira, 2002; Menon, 2006; Duarte & Lopes, 2003). Em seguida, faz-se uma apreciação de outras mudanças ocorridas no PB e que teriam uma relação com o fenômeno que ora se descreve. Na terceira e última parte do capítulo, são revisitadas diversas pesquisas que têm por objeto de estudo a variação no uso dos pronomes de segunda pessoa no PB.

O capítulo II, por sua vez, trata dos pressupostos teórico-metodológicos. Primeiramente, é apresentada a orientação que define esta pesquisa: a Teoria da Variação e da Mudança, ou Sociolinguística Variacionista. Discutem-se aqui o caráter sistêmico da variação e da mudança linguísticas, os tipos de mudança, os conceitos de *envelope de variação* e de *comunidade de fala*, e a metodologia adotada. Ainda neste capítulo, são apresentados os principais pressupostos teóricos que dão base às hipóteses aqui aventadas: a gramaticalização; as concepções discursivo-pragmáticas para a referência; o *Princípio do Contorno Obrigatório*; e os princípios discursivo-cognitivos do *paralelismo* e da *frequência*.

Introdução

No capítulo III, apresentam-se os *corpora* utilizados, define-se o envelope de variação desta pesquisa e desenvolve-se uma análise qualitativa dos dados. A partir dos pressupostos teóricos delineados no capítulo antecedente, são apresentadas as principais hipóteses que norteiam o estudo, traduzidas na forma de grupos de fatores (sociais e linguísticos).

No capítulo IV, apresenta-se a distribuição dos dados e interpretam-se os resultados quantitativos das análises desenvolvidas com o pacote estatístico VARBRUL (na sua versão mais recente, Golvarb X). Na primeira metade desse capítulo são apresentados os resultados da análise sincrônica com os dados da amostra mais recente (GESOL-SP-2000). Em seguida, discutem-se os resultados obtidos para amostra dos anos 1970 (NURC-SP-1970), em comparação àqueles obtidos em uma subamostra dos anos 2000 (que inclui apenas os informantes com nível superior de escolaridade).

CAPÍTULO I

Pronomes de Segunda Pessoa no Português Brasileiro

Cê vai, ocê fique, você nunca volte.

(Guimarães Rosa, “A terceira Margem do Rio” –
Primeiras Estórias)

Introdução

O presente estudo visa a investigar o uso variável do pronome de segunda pessoa (singular ou plural) *(vo)cê(s)* no português falado na cidade de São Paulo. Este capítulo, que tem por objetivo situar o objeto de estudo diacrônica e sincronicamente, encontra-se estruturado em três seções. Inicia-se com uma breve trajetória dos pronomes de segunda pessoa no português (seção 1.1). Em seguida (seção 1.2), recapitulam-se mudanças ocorridas na diacronia do português brasileiro (PB) que podem estar vinculadas direta ou indiretamente ao fenômeno aqui investigado. Em 1.3, traça-se um panorama dos principais trabalhos já realizados sobre os pronomes de segunda pessoa em diferentes comunidades de fala brasileiras.

1.1. Breve percurso histórico

Cunha & Cintra (2007), embora reconheçam o amplo uso do *você* em lugar do *tu* no Brasil, continuam a apresentar, no quadro geral dos pronomes pessoais, somente as formas *tu* e *vós* para a segunda pessoa. O mesmo se verifica em Almeida (2005), Rocha Lima (1998), Sacconi (1986) e muitos outros gramáticos da língua

portuguesa¹, independentemente da época em que as obras foram escritas. Essa classificação remonta à origem histórica desse pronome.

Em latim, a referência a uma segunda pessoa podia ser realizada por *tu*, uma forma singular utilizada para se referir a um único interlocutor; ou por *uos*, uma forma plural, também empregada para um único interlocutor, como uma forma de “respeito” (cf. Faraco, 1996; Brow & Gilman, 1972 [1960]). O latim contava ainda com formas indiretas, como *Uestra Maiestas* e *Uestra Serenitas* (cf. Faraco, 1996; Peres, 2006), consagradas ao tratamento ao imperador.

As formas latinas *tu* e *uos* deram origem aos pronomes pessoais *tu* e *vós* no português; *tu* e *vos*, em espanhol; *tu* e *vous*, em francês; *tu* e *voi*, em italiano (cf. Brown & Gilman, 1972 [1960]). Quanto ao item *você*, a concepção mais corrente entre filólogos e linguistas é a de que esse pronome teve sua gênese na expressão *vossa mercê*² (Cintra, 1972; Faraco, 1996; Teixeira, 2002; Menon, 2006; Duarte & Lopes, 2003, entre outros). Um processo semelhante ocorreu também em espanhol, no qual *vuestra merced* deu origem ao atual *usted*³.

Em sua origem, *vossa mercê* correspondia a um sintagma nominal, composto pelo pronome possessivo *vossa*, seguido do substantivo abstrato *mercê*, cujo significado era “favor”, “graça” (cf. Teixeira, 2002; Menon, 2006)⁴. Esta era a maneira com que os vassalos solicitavam favores (*mercês*) ao rei ou o agradeciam pelos benefícios (*mercês*) a eles concedidos (Menon, 2006). Nas palavras de Said Ali (1937; apud. Gonçalves, 2008: 86):

Os súbditos, dependentes sempre da mercê ou graça do príncipe, apresentavam as suas queixas e requerimentos dando-lhe o habitual *vós*. Sábiamente (sic) pediam por *mercê* e punham frequentemente *vossa mercê* por *vós*, referindo-se não à pessoa do soberano, e sim à graça e favor que dele dimanava.

¹ Nestas obras, não se encontra qualquer remissão às formas reduzidas *ocê* e *cê*, pois estas são características da língua falada, ao passo que tais gramáticas têm por base a modalidade escrita.

² Wilhelm (1979; apud. Gonçalves, 2008:99) registra o uso de *vostram mercedem* em latim. Tal forma poderia ser precursora de *vossa mercê*, no português, e de *vuestra merced*, em espanhol.

³ Biderman (1972; apud. Gonçalves, 2008:99) advoga em favor de que *vossa mercê* teria sido um empréstimo da forma de tratamento do espanhol *uestra merced*, durante o período de dominação espanhola em Portugal (final do século XVI e início do XVII).

⁴ Menon (2006) ainda destaca o uso de “minha mercê” e de “nossa mercê” em cartas emitidas pelo próprio rei, e o emprego de “vossa mercê” como uma forma de se dirigir também a Deus.

O uso frequente de *vossa mercê* levou estes dois itens a serem reanalisados como uma única unidade. *Vossa Mercê* emerge, então, como uma nova forma de tratamento ao rei, suplantando o tratamento *vós*, até então dispensado àquele soberano. É importante salientar que essa mudança não se deu de maneira abrupta nem isolada no sistema. Uma série de formas nominais de tratamento começaram a emergir a partir do século XIV, devido às transformações sociais pelas quais passava Portugal na época (Faraco, 1996).

No exemplo a seguir, o item *mercê* é empregado nas duas acepções: com o sentido de “favor/grança” (sublinhado) e como uma forma de tratamento ao rei (em negrito).

- (1) Outro sy, Senhor, os vossos Fidalgos e vossos naturaes dos vossos Regnos fazem saber a **Vossa Mercee** que elles recebem grande agravo dos Vossos Rendeiros das vossas imposições que vos poedes pela guisa que **vossa Mercee** he [...] e muitos destes, Senhor, acharedes que mais lhe levam e levaram per esta guisa do que elles ham, nem averam da conthia nem das mercees, que lhes vós fazedes, se **vossa Mercee** nom for de o temperar doutra guisa: porque senhor, vos pedem por mercee que vos lembreis delle. (Comunicação endereçada ao rei D. João I – In.: Nascentes, 1956: 115; apud. Sousa, 2008: 29; Grifos nossos)

No século XIV, o tratamento *vós* tornou-se muito generalizado para continuar a fazer referência ao rei – que passou de chefe militar a detentor do poder absoluto (cf. Faraco, 1996; Sousa, 2008). A forma *vossa mercê*, dentre outras existentes na época, foi a que se especializou para esse tratamento. O primeiro registro histórico de *Vossa Mercê* data de 1331 (Faraco, 1996), embora alguns autores defendam que seu emprego tenha se dado em um período anterior (Menon, 2006). O primeiro registro histórico de *você* de que se tem notícia viria a ocorrer apenas no século XVII, mais precisamente em 1666 (cf. Faraco, 1996).

Com o transcorrer do tempo, acontecimentos histórico-sociais propiciaram a expansão de *Vossa Mercê* para outros extratos da sociedade portuguesa (Faraco, 1996). De início, seu uso expandiu-se à nobreza – nobres, fidalgos, infantes e outros membros da corte passaram a ser tratados por esta expressão. Posteriormente, tal tratamento passou a ser reivindicado também pela burguesia nas interações com seus subalternos; e acabou por se difundir entre o povo em geral (Faraco, 1996). *Vossa mercê* surge, portanto, como um uso “honorífico” e vai se “desgastando” fonética e semanticamente até atingir um uso “comum” (Menon, 2006). O emprego frequente

dessa expressão pode ter contribuído também para o seu “desgaste fonético” (*erosão*), como se vê no percurso a seguir (Basto, 1931; apud. Gonçalves, 2008:84):

Vossa Mercê > *vossemecê* > *voss'mecê* > *vomecê*, *vom'cê* > *vôcê* (*voncê*) > *você*⁵

As mudanças pelas quais passou a expressão *vossa mercê* até a forma pronominal *você* têm sido descritas como decorrentes de um processo de gramaticalização, o qual se encontraria ainda em curso no português brasileiro atual, uma vez que esse pronome continuou a passar por alterações, como atestam suas atuais variantes reduzidas *ocê* e *cê*. A questão da gramaticalização deste pronome será retomada no capítulo II.

Alguns autores, como Duarte (1993; apud. Gonçalves), consideram que *tu* teria expandido seu uso também para a segunda pessoa do plural, substituindo o pronome *vós*; e, em um segundo momento, *você* concorreria com *tu*. Segundo esta proposta, portanto, tem-se: *vós* ~ *tu*; *tu* ~ *você(s)*⁶. Lapa (1991; apud. Gonçalves, 2008), porém, parte da hipótese de que a forma plural *vocês* é que teria entrado primeiro no sistema pronominal do português, substituindo o arcaico *vós*; e, em seguida, a forma singular *você* veio a substituir total ou parcialmente a variante *tu* a depender da região considerada. Neste caso, as alternâncias seriam: *vós* ~ *vocês*; *tu* ~ *você*.

Em geral, defende-se a entrada de *vossa mercê/você* no Brasil já como uma forma de tratamento (Sousa, 2008). Duarte (1993; apud. Lopes & Duarte, 2003) observa que a forma *tu*, que coexistia no século XIX com *você*, quase desaparece nas décadas de 1920 e 1930. O pronome *tu*, porém, voltaria a aparecer no cenário carioca no final do século XX (Paredes Silva, 2000, 2003; apud. Lopes & Duarte, 2003).

O levantamento dos usos de pronomes de segunda pessoa em peças teatrais no português europeu (PE) e no português brasileiro (PB) entre os séculos XVIII e XIX, feito por Lopes & Duarte (2003), demonstra que *tu* é a variante mais frequente

⁵ Além dessas formas, Basto (1931; apud. Gonçalves, 2008:84) destaca outras variantes regionais encontradas em Portugal: *voncê*, *vôcê* (Algarve, Alentejo e Trás-os-Montes); *voucê* (Penedono); *vaucê*, *baucê* (Barroso); *voicê* (Moncorvo); *vocei* (Alentejo); *bacê(s)* (Póvoa-de-Varzim). No território brasileiro, outras variantes regionais são: *vosmincê*, *vossuncê*, *voncê*, *vassuncê*, *vancê*, *vacê*, *mecê* (Basto, 1931; apud. Gonçalves, 2008:85); *mincê*, *oncê*, *sucê*, *suncê*, *vainicê*, *vansmicê*, *vossuncê*, *ucê* (Nascentes, 1956:119-121; apud. Peres, 2006:103).

⁶ O sinal ~ indica alternância entre as duas formas. No caso de *vós* ~ *tu*, o segundo item teria suplantado o primeiro, na proposta de Duarte.

tanto no PE quanto no PB (57% e 60%, respectivamente), embora o uso de *você* seja um pouco mais frequente no PB (13%) do que no PE (3%). A tabela a seguir esquematiza estes resultados.

Distribuição das ocorrências segundo o tratamento utilizado						
	Você	V. Mercê	Vós	Tu	Outros	Total
PE	23 (3%)	86 (12%)	107 (15%)	392 (57%)	88 (13%)	696
PB	88 (13%)	52 (8%)	46 (7%)	391 (60%)	79 (12%)	659
Total	111 (9%)	138 (10%)	135 (10%)	783 (58%)	167 (13%)	1352

Tabela 1: Formas de tratamento em peças de teatro nos séculos XVIII e XIX (Lopes & Duarte, 2003: 66)

Lopes & Duarte (2003: 69-72) evidenciam ainda que havia um favorecimento ao emprego da forma pronominal com o traço plural (*vocês*), enquanto *Vossa Mercê* era a forma preferencialmente adotada para o singular. No século XIX, o emprego de *você* mostrou-se bem mais elevado do que no século anterior, sobretudo nas relações assimétricas de superior para inferior. Além disso, o emprego de *você*⁷ nas peças teatrais se mostrava mais elevado entre personagens que pertenciam às classes populares – o que poderia refletir um emprego semelhante deste pronome entre estas classes na estrutura social da época.

Além de peças teatrais, a alternância *tu/você* no PB é atestada em outros gêneros textuais. Teixeira (1998; apud. Andrade, 2004), por exemplo, verifica esta alternância em textos literários publicados em jornais da Bahia no século XIX. Lucca (2000; apud. Andrade, 2004) também destaca a alternância *tu/vossa mercê* em correspondências em Minas Gerais no mesmo período. Em ambos os estudos, o *tu* constitui a forma de “solidariedade” entre os interlocutores (cf. terminologia de Brown & Gilman, 1972 [1960]), enquanto *você/vossa mercê* é a forma para o tratamento não-íntimo – padrão este que persiste atualmente nos dialetos em que essas duas formas são alternativas no PB (v. 1.2, neste capítulo).

⁷ Nesta época, *você* concorria com a expressão de tratamento *Vossa Mercê* e com o pronome *tu* (Lopes & Duarte, 2003).

O pronome *você*, todavia, seguiu um percurso distinto em Portugal e no Brasil. *Você* tem um emprego muito mais restrito entre os lusitanos (Ramos, 1998) e quando empregado, indica um maior distanciamento entre os interlocutores – o que o aproxima ainda a uma forma de tratamento. O *tu*, por outro lado, é a forma que tende a ser empregada para o tratamento íntimo ou “solidário” naquele país. Além disso, dependendo da região portuguesa, o uso de *você* pode mesmo ser interpretado como uma “forma ofensiva” (Hammemüllerm, 1993; apud. Gonçalves, 2008: 88), de “desdém” (Wilhelm, 1979; apud. Gonçalves, 2008: 89) ou ser considerado um uso “descortês e depreciativo” (Santos, 1985; apud. Gonçalves, 2008: 91).

No Brasil, ao contrário, *você* expandiu seu emprego, substituiu o *tu* em muitas regiões; tornou-se a forma preferida para o tratamento íntimo, e co-existe com as formas de “respeito” *o senhor / a senhora*. O uso atual de *você* e *tu* nestes dois países pode ser sintetizado no quadro a seguir – extraído de Ramos (1998):

	PE atual	PB atual
VOCÊ	Avaliação negativa em regiões rurais de Portugal (cf. Paiva Boléo 1946: 113) Uso ente iguais não-solidários e com interlocutor de status social inferior	Em expansão.
TU	Tratamento íntimo Em expansão	Uso restrito em certas regiões (3ª pes. sing.)

Quadro 1: *Você* e *Tu* no PB e no PE atualmente (Ramos, 1998: 159)

1.2. Mudanças afins

Além de *você* e *vocês*, a forma *a gente* também passou por um processo de gramaticalização, tornando-se um pronome pessoal de primeira pessoa do plural, concorrendo com o pronome “nós” (Zilles, 2002, 2005). A inserção dessas novas formas pronominais no PB (*você*, *vocês*, *a gente*) é geralmente vista como responsável pelo enfraquecimento do paradigma de flexões verbais⁸ nessa língua (Galves, 1996; Duarte, 1996). Se esta hipótese estiver correta, um paradigma de

⁸ Uma posição contrária a esta hipótese encontra-se em Ramos & Oliveira (2002), para quem a redução no paradigma de flexões verbais no PB deve-se a mudanças fonéticas ocorridas e não à entrada de novos pronomes pessoais. As autoras partem do princípio de que teria sido a queda de /d/ intervocálico que ocasionou a semelhança entre a segunda pessoa do singular e a segunda pessoa do plural: 2ª p. pl.: *podetis* > *podedes* > *podees* > **podes**; 2ª p. sg.: *podês*.

flexões verbais composto por seis morfemas (esquema A) teria se reduzido a um paradigma composto por apenas três morfemas (esquema B), ou somente dois morfemas (esquema C).

(A)	<i>eu cant-o</i>	(B)	<i>eu cant-o</i>	(C)	<i>eu cant-o</i>
	<i>tu canta-s</i>		<i>você canta-Ø</i>		<i>você canta-Ø</i>
	<i>ele(a) canta-Ø</i>		<i>ele(a) canta-Ø</i>		<i>ele(a) canta-Ø</i>
	<i>nós canta-mos</i>		<i>a gente canta-Ø</i>		<i>a gente canta-Ø</i>
	<i>vós canta-is</i>		<i>vocês canta-m</i>		<i>vocês canta-Ø</i>
	<i>eles(as) canta-m</i>		<i>eles(as) canta-m</i>		<i>eles(as) canta-Ø</i>

Quadro 2: Enfraquecimento do paradigma de flexões verbais no PB

Em decorrência dessas mudanças, uma mesma terminação verbal (morfema zero) passou a ser empregada para distintas pessoas: “*você / ele(a) / a gente canta-Ø*”⁹. A fim de se evitar possíveis ambiguidades, outra mudança teria sido desencadeada na língua: uma tendência ao preenchimento da posição de sujeito (Duarte, 1996; Galves, 1996). Tal questão é analisada por vários pesquisadores no âmbito da Teoria Gerativa, que alegam que o PB estaria passando por uma alteração do Parâmetro do Sujeito Nulo, aproximando-se de línguas como o inglês e o francês, nas quais o preenchimento do sujeito é obrigatório. De acordo com esta concepção, a tendência ao preenchimento do sujeito seria uma “resposta” do sistema, diante das mudanças ocorridas.

A possível formação de sujeitos expletivos (Duarte, 2003; Vitral & Ramos, 2006 [1999]) também é concebida por alguns autores como uma decorrência das mudanças aqui mencionadas. Duarte (2003) considera que o PB apresenta usos expletivos de pronomes pessoais (como *nós*, *a gente* e *você*) em sentenças com os verbos “ter” e “ver” com interpretações existenciais, como em (2) e (3). A autora parte do princípio de que: “*uma vez implementada uma das propriedades das línguas de sujeito não-nulo – o preenchimento dos sujeitos referenciais – o sistema*

⁹ Nos dialetos em que *tu* é uma variante, por vezes, este também é empregado com o verbo na terceira pessoa do singular, e não na segunda pessoa: “*tu canta*” (e não “*tu cantas*”).

começa a caminhar no sentido do preenchimento dos sujeitos não referenciais”, i.e., de sujeitos expletivos (Duarte, 2003: 125).

- (2) a. Você não tem mais clientela no centro da cidade.
b. ___ Não há / tem mais clientela no centro da cidade.
- (3) a. Você vê muito concreto na tua frente.
b. ___ Há / tem muito concreto na tua frente.

Vital & Ramos (2006 [1999]) também argumentam que o “*você / cê*” teria um uso expletivo em construções existenciais com o verbo “ter”. De acordo com esses autores, nestas construções, o pronome teria sofrido uma “perda de conteúdo semântico” e estaria desprovido de referência¹⁰.

Dessa maneira, teria havido um encadeamento de mudanças como mostra o quadro a seguir:

(1)	(2)	(3)	(4)	(5)
<i>Vossa Mercê > você</i>	<i>tu/vós ~ você(s)</i> <i>nós ~ a gente</i>	Enfraquecimento do paradigma de flexões verbais	Tendência ao preenchimento do sujeito.	Uso de pronomes como expletivos.

Quadro 3: Mudanças em cadeia no PB

O fenômeno em foco neste estudo também se insere em um rol de variações que vêm ocorrendo no Português Brasileiro atual. Assim como a redução *você(s) > (o)cê(s)*, encontram-se casos análogos em: *ele(s) > ei(s)* (Ramos & Correa, 2006 [1998])¹¹; *a gente > ‘enti* (Zilles, 2002); *não > num* (Cohen & Ramos, 2006); *está > tá* (Mendes, 1999; 2008). Todos esses casos de variação têm sido analisados dentro do quadro teórico da gramaticalização e podem indicar uma mudança mais ampla no PB: a formação de novos clíticos (Vital & Ramos, 2006a, Nascimento, 2009).

¹⁰ Kaiser (2006), entretanto, posiciona-se contrário a esta hipótese. No capítulo II, retoma-se essa discussão.

¹¹ De acordo com Ramos (2002; apud. Andrade, 2004) e Mattos e Silva (1989; apud. Andrade, 2004), a redução de pronomes pessoais não se trata de algo recente: no Português arcaico já se encontravam as formas reduzidas *el* e *eis* para a terceira pessoa *ele(s)* – as quais ainda não teriam um caráter clítico (cf. Ramos, 2002; apud. Andrade, 2004).

Como se viu em 1.1, o uso de *vossa mercê* como uma forma de tratamento ao rei e sua expansão a outros estratos sociais também se devem a mudanças ocorridas na estrutura social de Portugal. Estes fatos evidenciam que os processos de variação e mudança não ocorrem de maneira abrupta nem aleatória na língua, mas resultam da atuação conjunta de “forças” tanto internas quanto externas ao sistema linguístico.

1.3. Subsistemas pronominais de 2ª pessoa no PB

O pronome de segunda pessoa apresenta diversas variantes no PB atualmente. No caso da segunda pessoa do singular, as formas *tu*, *você*, *ocê* e *cê* encontram-se em variação¹². No plural, ocorrem as variantes *vocês*, *ocês* e *cês*. Tais formas, contudo, não se distribuem de maneira idêntica nas diferentes comunidades de fala. Em determinadas regiões, o pronome *tu* é uma forma bastante produtiva (com ou sem concordância com o verbo); em outras, encontram-se apenas o pronome *você* e suas variantes reduzidas.

Scherre (2010-2013)¹³, em um projeto que visa a mapear os usos de *você* e *tu* no Brasil, faz um levantamento das pesquisas já realizadas sobre o assunto no território nacional. A autora apresenta uma proposta preliminar, segundo a qual haveria seis subsistemas para o pronome de segunda pessoa no PB, resumidos a seguir (Scherre, 2010-2013: 18-19):

- (i) *Subsistema Pronominal V*: predomina o emprego de *VOCÊ* e suas variantes (de 97% e 100%).
- (ii) *Subsistema Pronominal T* com concordância muito baixa: o emprego de *tu* representa mais de 80%, dentre os quais a concordância do verbo com a segunda pessoa do singular representa menos de 10%.
- (iii) *Subsistema Pronominal T/V com concordância com o pronome TU muito baixa*: o emprego dos pronomes *tu* e *você* é equilibrado (aproximadamente 50% de *tu* e

¹² Estes também podem se alternar com as formas de tratamento *o senhor / a senhora*.

¹³ Scherre, M. M. P. *Análise e mapeamento de três Fenômenos Variáveis no Português Brasileiro*. Projeto de Pesquisa aprovado pelo CNPq para o triênio Março-2010 a Fevereiro-2013; em parceria com o Prof. Dr. Antony Julius Naro (UFES/UnB).

50% de *você*) e a concordância de *tu* com a segunda pessoa do singular é inferior a 10%.

- (iv) *Subsistema Pronominal T com concordância do pronome TU média*: o emprego de *tu* atinge um percentual bastante elevado (de 76% a 96%) e a concordância verbal com *tu* é superior a 40%.
- (v) *Subsistema Pronominal V/T ou T/V sem concordância com o pronome TU*: uso de 30% a 95% de *você* e de 5% a 70% de *tu*, sem concordância verbal.
- (vi) *Subsistema Pronominal V/T ou T/V com concordância com o pronome TU Médio-Baixa*: os percentuais de uso de *tu* e *você* ainda não seriam conhecidos, segundo a autora; a taxa de concordância verbal de *tu* com a segunda pessoa do singular é superior a 10%.

De acordo a autora, o *subsistema pronominal V* é encontrado em comunidades de fala localizadas nas regiões Centro-Oeste, Nordeste, Sul e Sudeste do Brasil. O presente estudo vai verificar que a comunidade de fala paulistana também se insere nesse subsistema, já que o *tu* não é uma forma empregada entre os paulistanos.

O *sistema pronominal T com concordância baixa* foi atestado em comunidades das regiões Sul e Norte do país. Os demais *sistemas pronominais T com concordância* foram verificados em comunidades no Sul (com concordância baixa) e no Nordeste, Norte e Sul (com concordância média). Já o *sistema pronominal T/V (ou V/T)* encontra-se em comunidades do Centro-Oeste, do Sudeste e do Sul (sem concordância) e em comunidades sulistas (com concordância médio-baixa) – cf. Scherre (2010-2013).

Dado o interesse do *subsistema V* para este trabalho, faz-se uma distinção entre este subsistema e aqueles em que o pronome *você* (e suas variantes reduzidas) concorrem com o pronome *tu*.

A variação *você* / *tu* no PB

Lucca (2005) e Dias (2007) tratam da variação entre *você* e *tu* no português falado em Brasília (região Centro-Oeste do país). A primeira autora analisa a fala de

jovens brasileiros entre 15 e 19 anos de idade, em fala espontânea em três comunidades no Distrito Federal (Brasília, Ceilândia e Taguatinga)¹⁴. Tal estudo demonstra que o emprego de *tu* é favorecido entre os homens, sobretudo em interações solidárias, com seus “pares”, e/ou quando o assunto da conversa lhes é familiar.

Dias (2007), por sua vez, analisa um *corpus* mais amplo, composto por interações espontâneas com indivíduos de três faixas etárias (13 a 19 anos, 20 a 29 anos e acima de 30 anos). Os resultados de Dias (2007) corroboram aqueles depreendidos por Lucca (2005) com relação ao gênero: são os homens quem mais empregam o pronome *tu*. Ao ampliar a faixa etária dos indivíduos, Dias (2007) verifica que há ao mesmo tempo uma mudança em progresso (quanto menor a idade dos indivíduos, maior o favorecimento da variante *tu*) e uma gradação etária (o uso de *tu* decresce com a estabilidade profissional do informante) naquela comunidade de fala.

Herênio (2006), em um estudo em Uberlândia (Minas Gerais, região Sudeste do país) e Imperatriz (Maranhão, região Norte), detecta a variação entre *tu* e *você* apenas nesta última comunidade.

Mota (2008) também desenvolve um estudo dos pronomes de segunda pessoa, com base em um *corpus* composto por entrevistas sociolinguísticas e testes de produção com informantes naturais de São João da Ponte (norte de Minas Gerais) – uma área que permaneceu de certa forma isolada diante do processo de urbanização e desenvolvimento econômico do estado mineiro. Contrariamente ao que Herênio (2006) verificou em Uberlândia, em São João da Ponte, há variação entre estas formas pronominais e o grau de intimidade entre os interlocutores foi considerado um dos fatores que mais favorece o uso de *tu*.

Na região Sudeste, o *tu* também ocorre com *você/cê* no Rio de Janeiro (Paredes Silva, 2003) e em Santos, no litoral paulista (Modesto, 2006). Paredes Silva (2003) desenvolve um estudo em tempo real de curta duração, considerando dois *corpora* do PEUL (Programa de Estudos sobre o Uso da Língua) com entrevistas sociolinguísticas – um dos anos 1980 e outro dos anos 2000; e um *corpus* dos anos

¹⁴ A amostra foi constituída por Malvar (1991; apud. Lucca, 2005).

1990 compreendendo interações espontâneas. Embora haja um reduzido número da variante *tu* nas entrevistas sociolinguísticas nos dois *corpora* do PEUL, este número é bem maior nas interações espontâneas – o que, segundo a autora, demonstra a persistência do *tu* entre os cariocas, em contextos de maior informalidade. A exemplo do que se observa em Brasília, no Rio também o emprego de pronome *tu* é favorecido na fala dos homens mais jovens.

Modesto (2006) investiga, entre outras questões, a atuação de fatores interacionais e discursivos, tais como “envolvimento”, “monitoramento” e “expressividade”, no uso das variantes *você* e *tu* em Santos. O autor conclui que *tu* é favorecido nos contextos de maior envolvimento, menor monitoramento e maior expressividade. A faixa etária e o gênero dos informantes não se mostraram correlacionados ao uso variável de *você* e *tu* entre os santistas.

É na região Sul do país, todavia, que o *tu* persiste com maior vigor. Loregian-Penkal (2004)¹⁵, por exemplo, verifica que o pronome *tu* funciona como uma “marca de identidade” aos “valores regionais” no Rio Grande do Sul e em Santa Catarina¹⁶, sendo empregado sobretudo pelas mulheres dessas localidades. Este resultado difere dos anteriores provavelmente devido às diferentes avaliações que recebe a variante *tu* no sul do país e nas comunidades de fala localizadas no Centro-Oeste e no Sudeste.

Também na região sul do país, Bolívar (2008) adota um método conhecido como “pesquisa rápida e anônima” (“R&A”) – a exemplo da pesquisa que Labov (1972) empreendeu em lojas de departamento de Nova York nos anos 1960¹⁷ – em Porto Alegre (RS). Foram analisadas interações em estabelecimentos comerciais de dois *shoppings centers* (um de classe alta e outro de classe média) e em um centro de comércio ambulante, frequentado por pessoas de classes sociais mais baixas. Bolívar (2008) avaliou as interações por parte dos vendedores, chegando à conclusão de que

¹⁵ Além da alternância entre as formas pronominais *tu* e *você*, a autora também investiga a concordância verbal com o pronome *tu*.

¹⁶ Godoy (1999; apud. Loregian-Penkal, 2004: 15) encontrou poucas ocorrências do pronome *tu* nas cidades de Irati, Londrina e Pato Branco (no interior do Paraná).

¹⁷ Labov (1972a; 2006 [1966]) investiga a pronúncia do /r/ pós-vocálico em palavras como *fourth* e *floor* na fala de empregados de três lojas de departamento de Nova Iorque (Saks, Macy’s e Klein – em ordem decrescente de *status*). O autor verifica que são as mulheres da loja Saks (ou seja, aquela de maior *status* dentre as três investigadas) que tendem a empregar a variante prestigiada (a pronúncia retroflexa do /r/). Os resultados gerais dessa pesquisa levaram à conclusão de que a variação na pronúncia do (r) é estratificada por classes sociais. Resultados similares foram obtidos em um estudo posterior (Fowler, 1986; apud Labov, 2006 [1966]).

você foi a forma preferida quanto maior o prestígio do estabelecimento comercial, especialmente na fala das mulheres¹⁸. Assim como nos estudos reportados anteriormente, as situações mais formais – no caso, interações entre desconhecidos e em ambiente de trabalho (“vendedor-cliente”) – parecem favorecer o uso de *você*, em detrimento de *tu*.

Embora muitos dos estudos sobre a variação entre *tu* e *você* admitam que esta última também co-ocorre com as formas reduzidas *ocê* e/ou *cê*, em geral tal distinção entre “*você/ocê/cê*” não é operada. As únicas exceções são as pesquisas de Queiroz Andrade (2010) e Martins (2010).

Queiroz Andrade (2010) analisa a variação entre *você*, *cê* e *tu* em Vila Planalto (Brasília)¹⁹ e conclui que *tu* tende a ser favorecido pelos homens, em relações simétricas; *você* tende a ser mais empregado pelas mulheres, em relações assimétricas; e *cê* é favorecido também nas relações assimétricas, sendo considerado um pronome “de esquiva” por Queiroz Andrade (2010) – i.e., uma variante intermediária entre *você* (“mais formal”) e *tu* (“mais informal”).

Na região norte do país, Martins (2010) trata da alternância entre *você*, *tu* e *senhor* em Tefê (Amazonas). Além dessas três formas, são analisadas também a variante *cê* e uma forma pronominal zero. Os resultados obtidos demonstram que *tu* é o pronome mais frequente naquela comunidade de fala (42,9%), seguido da forma zero (28,6%) e de *você* (22,9%). Por outro lado, as formas *senhor* e *cê* mostraram-se muito infrequentes (4,9% e 0,7%, respectivamente). Devido à baixa frequência de *cê* (8 ocorrências), o autor as analisa conjuntamente aos dados da forma plena desse pronome. De modo análogo aos estudos anteriores, também em Tefê, o *tu* é preferencialmente empregado pelas mulheres, pelos indivíduos mais jovens, e quando há maior grau de intimidade entre os interlocutores (amigos e familiares).

Nos vários estudos até aqui reportados, depreende-se que o *tu* tende a ser a forma empregada em contextos de maior intimidade entre os interlocutores, maior envolvimento com o assunto por parte do falante e menor grau de formalidade. O

¹⁸ Resultado semelhante foi obtido por Labov (2006 [1966]) para a variável (r) em NY, mas diverso do que Lorengian-Penkal (2004) verifica em relação à variação *tu / você* na região sul do Brasil.

¹⁹ Queiroz Andrade (2010) também considera o uso das formas de tratamento “o senhor/a senhora”. Estas, entretanto, ocorreram apenas na fala de um único informante.

sexo/gênero e a faixa etária do falante também são fatores importantes para o entendimento desse fenômeno variável, e sua atuação vai depender do *status* social atribuído às variantes nas diferentes comunidades de fala: em comunidades localizadas nas regiões centro-oeste e sudeste do país, os homens mais jovens tendem a empregar o *tu*; já no sul do país, são as mulheres que favorecem seu uso – exceto em interações no trabalho, com desconhecidos – como indica o estudo de Bolívar (2008).

A variação *você / ocê / cê* no PB

A grande maioria dos trabalhos realizados em comunidades cujo subsistema pronominal é do tipo *V* (cf. Scherre, 2010-2012) concentra-se em Minas Gerais. Destacam-se os trabalhos de Ramos (1997; 2000); Alves (1998); Peres (2006); e Gonçalves (2008).

Saindo do território mineiro, encontra-se também o estudo de Andrade (2004), em Brasília; de Sousa (2008), na Paraíba; e de Corradello (1997) na fala culta da cidade de São Paulo nos anos 1970. Estes dois últimos, todavia, têm um viés discursivo-pragmático, mas não variacionista. Não se tem notícia de um estudo que analise este fenômeno na cidade de São Paulo a partir de uma perspectiva variacionista.

***Você/ocê/cê* em Minas Gerais**

Em uma série de estudos desenvolvidos em Minas Gerais, Ramos (1997) é uma das primeiras a investigar a variação no uso de *você*, *ocê* e *cê* em Belo Horizonte. A autora tenta conciliar a metodologia da sociolinguística variacionista com uma concepção formalista de língua, analisando um *corpus* composto por 12 entrevistas. Foram investigados os grupos de fatores: função sintática do pronome; focalização (focalizado, não focalizado); referência do pronome (definida, indefinida); tipo e oração (interrogativa simples, interrogativa *que-que*, declarativa); posição da variante em relação ao verbo (contígua, não contígua); construções de

tópico; faixa etária (jovens, medianos e velhos)²⁰; sexo/gênero (masculino, feminino).

Os resultados obtidos por Ramos (1997) confirmam a hipótese de que *cê* estaria se comportando como um clítico na língua. A variante mais reduzida *cê* foi empregada apenas em contextos cuja função sintática era a de sujeito, exceto quando tópico ou foco. Além disso, *cê* foi a forma preferida em construções contíguas ao verbo (i.e., sem a interpolação de outros constituintes entre o pronome e o verbo) e apresentou uma maior frequência em orações declarativas e interrogativas *que-que*²¹.

Como a tabela a seguir demonstra, *cê* é mais frequente entre os indivíduos mais velhos (65%); e *você* entre os mais jovens (67%) quando a referência é definida (específica ao interlocutor). No caso da referência indefinida (genérica), tem-se um padrão inverso: os falantes mais jovens é que empregam mais frequentemente a variante reduzida (56%), enquanto os mais velhos preferem *você* (53%).

IDADE	Referência Definida						Referência Indefinida					
	CÊ		OCÊ		VOCÊ		CÊ		OCÊ		VOCÊ	
	Nº.	%	Nº.	%	Nº.	%	Nº.	%	Nº.	%	Nº.	%
Jovem	2	33	-	-	4	67	47	56	16	19	21	25
Mediano	29	40	9	13	34	47	35	63	6	11	14	26
Velho	71	65	8	7	31	28	5	33	2	13	8	53
Total	102		17		69		87		24		43	

Tabela 2: *Você, ocê e cê* conforme a referência e a Faixa etária em Belo Horizonte (Ramos, 1997: 50)

Esse resultado leva a autora a aventar a hipótese de que “*cê* estaria se especializando com referência indeterminada” – o que remete à concepção corrente em parte da literatura sobre gramaticalização, que prevê uma relação entre “abstratização semântica” e “perda de material fonético” (Hopper & Traugott, 1993).

²⁰ Não são especificadas as idades.

²¹ Clíticos são elementos átonos que precisam se “ancorar” em outro constituinte sintático, em geral o verbo (daí a contiguidade entre *cê* e o verbo corroborar a hipótese de cliticização desta variante reduzida do pronome). Ramos (1997) considera a hipótese de que as construções sintáticas com a posição de complementizador preenchida (interrogativas *que-que* – ou interrogativas “*QU-que*”) favoreceriam o uso de clíticos. Estas hipóteses serão abordadas em detalhes no capítulo III.

Embora o estudo de Ramos (1997) tenha trazido contribuições importantes a respeito da variação entre *você*, *ocê* e *cê* em Minas Gerais, a falta de medidas estatísticas, como os pesos relativos²² e de um *corpus* mais robusto compromete a interpretação dos resultados, bem como a sua generalização à comunidade de fala de Belo Horizonte (BH).

Anos depois, a autora volta a analisar a mesma comunidade de fala, desta vez comparando-a com a cidade de Ouro Preto (OP) (Ramos, 2000). Enquanto BH se caracteriza como um “grande centro cosmopolita”, OP é considerada uma região mais conservadora naquele estado. O *corpus* do estudo foi composto por entrevistas sociolinguísticas, aulas e entrevistas televisivas. Além da região geográfica do falante (BH, OP), foram analisados os grupos de fatores escolaridade, idade, sexo/gênero, contiguidade do pronome ao verbo e referência do pronome.

Seus resultados mostram maior frequência da variante reduzida *cê* em BH (55%) e uma preferência por *você* em OP (65%). A variante *ocê*, por sua vez, foi menos frequente em ambas as amostras. Segundo Ramos (2000), esse resultado deve-se à avaliação social negativa que *ocê* recebe naquela comunidade de fala: um modo de falar típico de áreas rurais, um “falar caipira”.

Com relação à função sintática, *cê* ocorreu exclusivamente na função sintática de sujeito – o que Ramos (1997) tomou como indício da especialização dessa variante com Caso²³ nominativo. A forma *ocê* foi mais frequente entre os informantes com nível de escolaridade até a 4ª série do ensino fundamental (48%). Essa mesma variante mostrou-se extremamente infrequente na fala das mulheres (6%)²⁴.

Em BH, o uso de *você* teve um percentual mais elevado nas aulas universitárias (92,8%), seguido das entrevistas televisivas (69,2%). As entrevistas sociolinguísticas foram os contextos em que esta variante ocorreu com menor frequência (33%) – o que indica, segundo a autora, que o grau de formalidade

²² Dispõe-se apenas da frequência das variantes em determinados contextos.

²³ Aqui, grafa-se “Caso” (com a inicial em letra maiúscula) para o Caso gramatical (i.e., nominativo, acusativo, dativo), distinguindo-o de outras acepções que essa palavra possui.

²⁴ As mulheres têm demonstrado, de acordo com diversos estudos sociolinguísticos (Labov, 2006 [1966]; 1972; Chambers, 1995; dentre outros), uma sensibilidade à avaliação social das variantes na comunidade de fala. Estas geralmente tendem a empregar as formas mais prestigiadas, evitando-se as variantes estigmatizadas – como é o caso de *ocê*.

poderia estar correlacionado ao emprego das variantes em estudo: quanto mais formal, maior a preferência pela variante plena *você*, considerada padrão por Ramos (1997).

Contrariamente ao que Ramos (1997) observou em BH, em OP os falantes mais jovens (24-25 anos) e os de faixa etária considerada mediana (52-63 anos) tendem a empregar *cê* com referência determinada (.76 e .63, respectivamente)²⁵. Quando a referência é indeterminada, esses mesmos informantes tendem a utilizar a variante plena *você* (.59 e .53, respectivamente). Os falantes mais velhos (acima de 77 anos) desfavorecem tanto o emprego de *cê* determinado (.44), quando o de *você* indeterminado (.42).

A autora conclui, a partir desses resultados, que a variante conservadora *você* é a preferida em OP, uma cidade historicamente “mais conservadora” do que BH. Nesta última, o processo de gramaticalização deste pronome se encontraria mais avançado, daí a variante inovadora *cê* ser mais frequente, inclusive nos contextos indeterminados.

Segundo esta hipótese, uma comunidade de fala mais conservadora, como OP, teria um ritmo de mudança mais lento do que uma cidade cosmopolita, como BH. No caso do presente estudo, a cidade de São Paulo aproxima-se histórica e economicamente mais de BH do que de OP. O capítulo IV trata dessa possível semelhança, sociolinguisticamente, mas aqui já se adianta a hipótese de que, em São Paulo, a variante inovadora *cê* poderia estar substituindo cada vez mais a variante plena *você*²⁶ com frequência e em uma multiplicidade de contextos.

A comunidade de fala de BH também foi analisada por Alves (1998). Este autor, porém, concentra-se nos usos de *você/ocê/cê* com referência indeterminada, em um *corpus* composto por entrevistas sociolinguísticas e programas de rádio e TV. Inicialmente, é realizado um levantamento dos recursos de indeterminação do sujeito empregados – cujos resultados são transcritos a seguir (Alves, 1998: 51).

²⁵ Os pesos relativos são apresentados aqui sempre por ponto seguido de duas casas decimais.

²⁶ Os resultados obtidos, todavia, não confirmam tal hipótese (v. capítulo IV).

Formas indeterminadoras		[Nº]	%
1. <i>se</i> impessoal	com infinitivo	04	1,4
	com forma finita	11	3,8
2. cv [categoria vazia]	c/ verbo na 3ª p. pl.	02	0,7
	com infinitivo	17	5,9
3. <i>você/ocê/cê</i>		131	45,6
4. a gente		73	25,4
5. as pessoas		19	6,6
6. a pessoa		11	3,9
7. cv2	[3ª p. sing. sem 'se']	07	2,5
8. o pessoal		07	2,5
9. eles		02	0,7
10. nós		02	0,7
11. ele		01	0,3
Totais		287	100

Tabela 3: Recursos de indeterminação do sujeito em Belo Horizonte (Alves, 1998: 51)

Embora em uma amostra bastante reduzida (287 ocorrências), o pronome *você/ocê/cê* e suas variantes foram os mais frequentes (46%), comparativamente a outros recursos indeterminadores. Em um segundo momento, foram investigados os fatores que controlam a variação no uso de *você*, *ocê* e *cê* indeterminados.

Das 131 ocorrências de *você/ocê/cê*, 59% são representadas pela forma plena *você*; 34% pela variante reduzida *cê*; e apenas 7% por *ocê*. Tais resultados afastam-se daqueles encontrados por Ramos (1997, 2000) na mesma comunidade. Alves (1998), porém, não se posiciona contrário à hipótese de que a variante reduzida *cê* estaria se especializando com referência indeterminada em BH, como propôs Ramos (1997): o fato de sua amostra ser constituída, em sua grande maioria, por dados provenientes de programas de rádio e TV e compor-se por informantes de faixas etárias mais velhas teria sido a causa dos resultados distintos dos apresentados pela autora.

Ainda em Belo Horizonte, Peres (2006) amplia as discussões anteriores a partir da análise de dois *corpora* separados por um intervalo de 20 anos (1982 e 2002). São analisadas sincrônica e diacronicamente as formas *você*, *ocê* e *cê*, considerando-se os mesmos grupos de fatores linguísticos investigados por Ramos (1997) – o que também tornou possível comparar os dois estudos. Peres (2006), contudo, amplia as faixas etárias de três, para cinco: crianças de 8 a 11 anos; adolescentes de 12 a 15 anos; jovens entre 16 e 30 anos; adultos medianos de 31 a 47

anos; e adultos com mais de 47 anos de idade. A tabela a seguir traz seus resultados para faixa etária, de acordo com a referência do pronome, no *corpus* de 2002.

CÊ						
FAIXA ETÁRIA	REF Definida (específica)			REF Indefinida (genérica)		
	Nº	%	p.r.	Nº	%	p.r.
V (+ de 47)	93/114	81,6	.44	93/89	49,2	.32
IV (31-47)	167/191	87,4	.51	144/205	70,3	.53
III (16-30)	133/141	94,3	.66	215/268	80,2	.63
II (12-15)	35/45	77,8	.37	31/39	79,5	.63
I (8-11)	137/162	84,6	.48	0/7	0	-
Total	565/653	86,6		483/708	68,2	

Tabela 4: Emprego de *Cê* em relação à referência e à faixa etária em BH em 2002 (Peres, 2006)

Considerando-se apenas os dados das faixas etárias III, IV e V, os resultados deste estudo diferem daqueles obtidos por Ramos (1997), que indicavam uma possível especialização de *cê* com referência indeterminada em Belo Horizonte. Por outro lado, os adolescentes (entre 12 e 15 anos) são os que mais favorecem o emprego de *cê* quando a referência desse pronome é indeterminada (.63), e ocorrem poucas ocorrências desse pronome com referência indeterminada (7 dados, todos com a variante plena) na fala de crianças entre 8 e 11 anos.²⁷

Além disso, nota-se uma possível mudança em progresso (v. Capítulo II): *cê* tende a ser empregado pelos adultos mais jovens (faixa etária III) tanto com referência determinada (.66), quanto com referência indeterminada (.63). O estudo em tempo real realizado pela autora demonstrou um padrão semelhante no que diz respeito ao uso de *cê* entre os adultos mais jovens, mas com um considerável aumento no peso relativo entre os falantes de 16 a 30 anos (.57, em 1982; e .65, em 2002).

²⁷ Embora esses resultados sejam importantes por indicarem também uma possível aquisição tardia dos recursos indeterminadores pela criança, essa questão não será discutida aqui, uma vez que o presente estudo trabalha com amostras de fala de indivíduos adultos.

FAIXA ETÁRIA	CÊ					
	BH – 1982			BH – 2002		
	Nº	%	PR	Nº	%	PR
V (+ de 47)	110/148	74,3	.38	186/319	58,3	.33
IV (31-47)	53/74	71,6	.44	314/420	74,8	.47
III (16-30)	292/365	80,0	.57	351/446	78,7	.65
Total	455/587	77,5		851/1185	71,8	

Tabela 5: Emprego de Cê em relação à faixa etária do informante em BH 1982-2002 (Peres, 2006)

Peres (2006) também verifica uma ligeira tendência, entre as mulheres, de empregar a variante reduzida do pronome em 2002 (.57), comparativamente aos homens (.43). Um cruzamento da faixa etária com o sexo/gênero do falante revela um comportamento distinto entre os falantes mais velhos: as mulheres acima de 47 anos empregam mais frequentemente a variante reduzida ($109/135 = 81\%$), do que os homens da mesma faixa etária ($77/184 = 42\%$)²⁸.

Gonçalves (2008), a partir de um *corpus* constituído de narrativas orais, analisa a variação entre *você/ocê/cê* em Arcos (norte de Minas Gerais). O autor traça um paralelo entre os seus resultados e aqueles obtidos por outros pesquisadores – como Ramos (1997), Coelho (1999)²⁹ e Peres (2006) –, chegando à conclusão de que o uso de *cê* já estaria implementado na comunidade de fala mineira, pois seu emprego supera o das demais variantes nos diversos trabalhos já realizados.

Para a referência definida (específica), *cê* é a variante mais frequente (60%) – contra 26% de *ocê* e 14% de *você*. Quando a referência do pronome é indefinida (genérica), *você* é a variante mais empregada (44%) – contra 39% de *cê* e 17% de *ocê*. Além disso, *cê* é favorecido quando há uma pausa no contexto antecedente (72%) e é majoritariamente empregado com a função sintática de sujeito da oração.

Gonçalves (2008) depreende o mesmo padrão encontrado por Peres (2006) com relação à referência da variante reduzida: a referência determinada favorece o emprego de *cê*, enquanto a indeterminada favorece a variante plena *você*.

²⁸ O capítulo IV vai mostrar resultados semelhantes para São Paulo.

²⁹ Coelho (1999; apud. Gonçalves, 2008) analisa a comunidade de São Francisco, norte de Minas Gerais. A questão mais central de seu estudo está na oposição entre zona urbana e zona rural com relação ao emprego da variante *ocê*.

Com relação às variáveis sociais, *ocê* e *cê* são mais frequentes nas áreas rurais (73,5% de *ocê* e 60% de *cê*) do que nas urbanas (26,5% de *ocê* e 40% de *cê*). Além disso, o autor verifica que estas duas variantes são mais frequentes entre os homens do que entre as mulheres, enquanto *você* apresenta um uso equilibrado (49% na fala dos homens e 51% na fala das mulheres).

Fundamentando-se em Brown & Gilman (1972 [1960]), Gonçalves (2008: 225) também considera *você* como um “pronome de poder”, empregado em interações assimétricas; *ocê* e *cê* seriam “pronomes de solidariedade”, utilizados em interações simétricas. Nos capítulos II e III a seguir, essa proposta é verificada para São Paulo.

Brasília

Saindo do âmbito mineiro, Andrade (2004) também investiga a variação entre *você*, *ocê* e *cê* em Brasília, em um *corpus* composto pela fala de crianças e adolescentes (de 10 a 14 anos) e pelos dados provenientes da fala de uma documentadora (com 25 anos na época da gravação; natural do Rio de Janeiro, mas residente em Brasília há 19 anos).

No *corpus* analisado por Andrade (2004), a variante reduzida *cê* ocorreu apenas com a função sintática de sujeito, tanto na fala da documentadora, quanto na fala dos informantes. A autora verifica que as orações interrogativas *que-que*³⁰ favoreceram o uso da variante reduzida e a presença de constituintes entre o pronome e o verbo a desfavorece. Tais resultados vão ao encontro daqueles obtidos por Ramos (1997). Além disso, o uso das variantes é correlacionado pela variável paralelismo (Scherre, 1998): uma vez pronunciada uma variante a mesma tende a continuar sendo empregada em séries de orações cujo sujeito é o mesmo. A referência genérica, por sua vez, desfavorece o emprego de *cê*, sobretudo na fala da documentadora.

De maneira semelhante ao observado por Gonçalves (2008), Andrade também nota que a variante reduzida *cê* tende a ser mais empregada pelas crianças

³⁰ Ressalva-se de que, embora a distinção entre interrogativas *que-que* (i.e., *o que que...?*) e demais estruturas *Wh-que* (i.e., *onde que*, *como que*, etc.) tenha se mostrado significativa no uso variável de *você/cê*, o número de dados do segundo tipo foi muito escasso no *corpus* de Andrade (2004) (apenas 10 ocorrências do pronome, dentre as quais somente uma foi com a variante reduzida *cê*).

das áreas rurais. A forma *ocê*, por sua vez, foi extremamente rara em ambas as comunidades: 5 ocorrências na área urbana e 2 ocorrências na área rural – o que impede extrair quaisquer conclusões³¹.

Outros estudos

Numa linha mais discursivo-pragmática, destacam-se ainda os trabalhos de Corradello (1997) e Sousa (2008). A primeira analisa 18 inquéritos do NURC-SP³², propondo que ao lado dos usos específicos do pronome *você*, há distintos empregos genéricos/indeterminados desse pronome. Tal estudo é de grande importância para esta pesquisa, uma vez que se trata da primeira análise realizada sobre esse pronome na capital paulistana. Dentre os objetivos do trabalho de Corradello (1997) estão: (i) discutir a posição das gramáticas tradicionais a respeito do pronome *você* e da indeterminação do sujeito; (ii) observar quais fatores sociais, como a faixa etária do informante, seu sexo/gênero e o tipo de inquérito (DID, D2 e EF)³³ levam a um uso mais frequente de *você* indeterminado (independentemente de sua pronúncia plena ou reduzida); e (iii) desenvolver um estudo pragmático-discursivo desse pronome no que diz respeito aos seus usos indeterminados.

Embora Corradello (1997) proponha uma análise quantitativa, esta se enquadra no que a autora chama de “sociolinguística distribucional”. Na verdade, não se trata da análise de um fenômeno variável, mas do cômputo da frequência com que ocorre “você genérico” no *corpus* (independentemente de sua pronúncia plena ou reduzida). Por outro lado, se a intenção da autora era argumentar em favor de um aumento da frequência de *você* com referência indeterminada com base no tempo aparente, isso só seria plenamente possível se fossem analisadas outras formas de indeterminação, como o fez Alves (1998).

Sousa (2008) faz uma análise semelhante à de Corradello (1997) a partir de uma amostra de fala de João Pessoa (Paraíba). Assim como no estudo anterior, o interesse não está em desenvolver uma análise variacionista, mas apenas computar a

³¹ Andrade analisa também a frequência de uso entre *você*, *ocê*, *cê* e zero.

³² Incluindo diálogos entre dois informantes, diálogos entre documentador e informante e elocuições formais/aulas.

³³ DID = diálogo entre informante e documentador; D2 = diálogo entre dois informantes; EF = elocuições formais (aula universitárias).

frequência de *você* (independentemente de sua pronúncia) no *corpus*. À parte das possíveis objeções às análises quantitativas realizadas por essas duas autoras, elas são as únicas a considerar a referência do pronome *você* além da oposição binária “determinada *versus* indeterminada”.

A variação *Você/cê* na cidade de São Paulo

Diferentemente do que se observa em Minas Gerais e Brasília, na cidade de São Paulo, a variante intermediária *ocê(s)* é improdutiva (6 ocorrências dessa variante em um total de 4639 dados). O mesmo ocorre com a forma *ucê* (3 ocorrências). Na comunidade de fala paulistana, portanto, o pronome “*você(s)*” apresenta duas variantes: uma forma foneticamente plena *você(s)*³⁴, e uma variante reduzida *cê(s)*, como nos exemplos a seguir:

- (4) a) Aí depois que **você** voto(u), **cê** fala: “Puxa vida, será que eu votei consciente?” (Marta-F2FI-2008)
b) Aí depois que **cê** voto(u), **você** fala: “Puxa vida, será que eu votei consciente?”
- (5) a) **cês** já tiveram oportunidade de estar em Belo Horizonte não?... já? (Samoel-M2SI-1970)
b) ... **Vocês** já tiveram oportunidade de estar em Belo Horizonte não?... já?

Nesta pesquisa, portanto, são investigadas apenas as variantes *você(s)* e *cê(s)*. No capítulo III, determina-se o envelope de variação e evidenciam-se os contextos em que essas duas formas funcionam como variantes de uma variável.

³⁴ A pronúncia de *você(s)/cê(s)* normalmente é realizada com a epêntese de [i] antes do [s]: *vocêis/cêis*. Neste estudo, todavia, não se fez distinção entre essas pronúncias (com ou sem um [i] epentético).

CAPÍTULO II

Pressupostos Teórico-Metodológicos

[...] o domínio de um falante nativo [...] de estruturas heterogêneas não tem a ver com multidialealismo nem com “mero” desempenho, mas é parte da competência linguística monolíngue. [...] numa língua que serve a uma comunidade complexa (i.e., real), a ausência de heterogeneidade estruturada é que seria disfuncional.

(Weinreich, Labov & Herzog, 2006 [1968])

Introdução

Este capítulo estrutura-se em cinco seções. Na primeira (2.1), são apresentados os pressupostos teóricos e metodológicos da Teoria da Variação e da Mudança, na qual esta pesquisa se fundamenta. Nas seções 2.2 a 2.5, tratam-se das bases teóricas em que se fundamentam as hipóteses aventadas sobre os usos variáveis de *você/cê* – ou seja, as hipóteses que se traduziram em grupos de fatores linguísticos (v. capítulo III).

2.1. Teoria da Variação e da Mudança Linguística

Esta pesquisa se insere no quadro teórico-metodológico da Sociolinguística Variacionista, também conhecida como Teoria da Variação e da Mudança (Labov, 2006 [1966], 1972, 1994, 2001; Chambers, 1995; Weinreich et al., 2006 [1968]; Guy, Ms. inédito). Em linhas gerais, tal perspectiva teórica assume que a variação é parte integrante do próprio sistema linguístico, e que é controlada por forças tanto internas (linguísticas), quanto externas (sociais).

A década de 1960 é considerada o marco do desenvolvimento da Sociolinguística enquanto uma disciplina (cf. Fernández, 1997a; Silva-Corvalán,

1989). Embora existam estudos sociolinguísticos em épocas precedentes – como, por exemplo, os estudos de Gauchat (1899) e Hermann (1929) em Charmey (Suíça)¹ – é na década de 1960 que a Sociolinguística começa a se firmar enquanto uma área de pesquisa com uma teoria e uma metodologia próprias. Isto se dá graças às pesquisas desenvolvidas pelo sociolinguista norte-americano William Labov. Seus trabalhos fundadores, que trataram da centralização da primeira vogal dos ditongos (ay) e (aw) em Marthas Vineyard (Labov, 1972 [1963]) e de variáveis fonológicas na cidade de Nova York² (Labov, 2006 [1966]), forneceram as bases fundamentais para um estudo sistemático da variação.

Os primeiros estudos desenvolvidos na área voltaram-se, sobretudo, para as variáveis de natureza fonológica (cf. Silva-Corvalán, 1989). A expansão dos estudos variacionistas para outros níveis de análise gerou algumas controvérsias, pois, de acordo com Silva-Corvalán (1989): a manutenção do significado ou “valor de verdade” não é tão precisa no caso de variáveis não-fonológicas (Lavandera, 1978); há uma maior dificuldade em se identificar o contexto variável (envelope de variação); e a variação morfológica ou sintática parece não apresentar uma estratificação social (Lavandera, 1978).

Esses fatos já foram observados em muitos trabalhos que tratam da variação sintática. Mendes (2005), por exemplo, investiga o uso variável das perífrases “estar + gerúndio” e “ter + particípio” na expressão dos aspectos durativo (“*Já faz tempo que ela **está** se **dedicando** ao balé*” vs. “*Já faz tempo que ela **tem** se **dedicado** ao balé*”) e iterativo (“*Eu **tô** **saindo** muito pouco*” vs. “*Eu **tenho** **saido** muito pouco*”). O autor depara-se com questões relativas à delimitação do envelope de variação e à manutenção do “mesmo” significado. As perífrases anteriores, por exemplo, não são variantes quando o aspecto é o progressivo (“*Diz que eu **estou** **tomando** banho e ligo depois*” vs. “**Diz que eu*

¹ Apud. Labov (2006 [1966]).

² Para analisar o sistema fonológico da cidade de Nova Iorque, o autor elegeu cinco fenômenos variáveis: (i) a pronúncia do /t/ em posição pré-consonantal e final; (ii) a altura da vogal /aeh/; (iii) a variação no uso da vogal /oh/; (iv) a pronúncia (fricativa, africada ou oclusiva) da consoante /th/; e (v) a pronúncia do segmento /dh/. Com exceção das duas últimas variáveis (que ocorrem no inglês falado em todos os EUA), as demais são, de acordo com o autor, características do falar nova-iorquino.

tenho tomado banho e ligo depois”) ³. Além disso, o autor verifica uma baixa correlação de variáveis sociais relativamente ao uso das perífrases anteriores (a única exceção foi a variável faixa etária).

Para o caso das variantes *você* e *cê*, quanto ao envelope de variação, embora não haja dúvidas de que as variantes sejam semanticamente “idênticas”, nem sempre elas podem ser comutadas nos mesmos contextos (v. capítulo III).

Heterogeneidade e sistematicidade

A variação linguística é um fato observável em todas as línguas. No entanto, há uma longa tradição que a situa fora da agenda de pesquisa linguística (cf. Labov, 2006 [1966]; Guy & Boberg, 1997; Guy, Ms. inédito).

Embora a Linguística Histórica do século XIX se ocupasse em examinar mudanças linguísticas, estas eram analisadas como decorrentes apenas de pressões internas, do sistema linguístico (Guy, Ms. inédito). Para os neogramáticos, por exemplo, a mudança fonética era resultante de regras que se aplicariam sem exceção (os contra-exemplos eram considerados “empréstimos” ou casos de “variação livre”). O estruturalismo saussureano (Saussure, 1922; aúd. Guy, Ms. inédito), no início do século XX, veio consolidar ainda mais essa concepção formalista, ao afirmar que a Linguística deveria se ocupar do estudo da *langue* (a língua, um sistema social compartilhado por todos os indivíduos de uma comunidade), em detrimento da *parole* (a fala, individual e idiossincrática).

Posteriormente, a Teoria Gerativa, proposta por Noam Chomsky, também opôs um “sistema abstrato da língua” (*competência linguística*) ao seu “uso real” (*performance*)⁴. Para o autor, a Linguística deveria se ocupar apenas da primeira, voltando-se para o estudo de um “falante-ouvinte ideal em uma comunidade de fala completamente homogênea” (Chomsky, 1965; apud. Guy, Ms. inédito). Dentro dessa perspectiva teórica, os casos de variação, quando não deixados de

³ Todos os exemplos citados nesse parágrafo são de Mendes (2005: 38).

⁴ Dentro dessa perspectiva, há uma sucessão de modelos teóricos: o modelo gerativo-transformacional, o modelo de princípios e parâmetros e, mais recentemente, o programa minimalista.

lado, são vistos como resultado da “mistura de dialetos”, da “coexistência de gramáticas distintas” ou como “erros de *performance*” (Labov, 2008).

Delegada à *parole* (Saussure, 1916), ou à *performance* (Chomsky, 1965; apud. Guy, Ms. inédito), a variação permaneceu, portanto, fora do escopo de investigações da chamada Teoria Linguística por um longo tempo (Labov, 2006 [1966]; Guy, Ms. inédito).

Alguns pesquisadores procuraram ampliar o conceito de “competência”, a fim de abranger não apenas questões formais, mas também sociais e culturais. Hymes (1967; apud. Silva-Corvalán, 1989), por exemplo, adota o conceito de “competência comunicativa” – anos mais tarde retomado por Fishman (1972; apud. Silva-Corvalán, 1989) sob o rótulo de “competência comunicativa sociolinguística”. Ambos estabelecem que o falante adquire não somente as regras linguísticas invariáveis de sua língua materna, mas também aquelas que regulam os usos linguísticos variáveis nos diversos contextos sociais.

Nessa mesma linha de raciocínio, Labov (2010) defende a existência de fatores cognitivos e culturais que possibilitam ao falante de uma determinada comunidade captar os padrões de comportamento linguístico (em termos probabilísticos) nos diversos contextos sociais, podendo adaptar a sua fala de acordo com a situação. Em outras palavras, o falante não apenas adquire a estrutura de sua língua, mas reconhece quais variantes são mais ou menos apropriadas a depender do contexto: um mesmo indivíduo poderá adotar um padrão de variação em um estilo de fala mais formal, por exemplo; e um outro padrão em um estilo de fala menos formal. Não se trata, todavia, de padrões categóricos, mas probabilísticos.

A Sociolinguística Variacionista assume, assim, duas premissas fundamentais. A primeira é a de que há uma “variabilidade inerente” em todas as línguas; e a segunda é a de que esta “heterogeneidade” não é “caótica” ou “livre”, mas “ordenada” (Weinreich et al., 2006 [1968]) – controlada por fatores linguísticos e extralinguísticos. Nas palavras de Weinreich et al. (2006: 36 [1968]): “numa língua que serve a uma comunidade complexa (i.e., real), a *ausência* de heterogeneidade estruturada é que seria disfuncional”.

Dado que a variação é concebida como parte integrante do próprio sistema linguístico, Guy & Boberg (1997) defendem que um modelo de gramática deve integrar tanto a variação sistemática, quanto os empregos considerados categóricos em uma língua, pois estes últimos seriam “casos extremos” da aplicação de uma regra probabilística. A gramática, portanto, teria tanto um componente *invariável*, quanto um componente *quantitativo – não-categórico e não-determinístico* (Guy & Boberg, 1997).

Relativamente aos condicionamentos sociais, há um padrão que vem se repetindo ao longo dos estudos sociolinguísticos. Destacam-se aqui três variáveis – sexo/gênero, faixa etária e escolaridade.

A variável Sexo/Gênero⁵ tem sido amplamente investigada na literatura sociolinguística desde os anos 1960. Os resultados geralmente demonstram que as mulheres tendem a ser mais “sensíveis” ao prestígio ou estigma associado às variantes (Labov, 2001; Chambers, 1995).

Fischer (1958; apud. Chambers, 1995: 109), por exemplo, analisa a variação na pronúncia do segmento final do sufixo *-ing* do gerúndio em um bairro próximo a Boston (EUA) em leituras realizadas por crianças. As variantes analisadas são a pronúncia velar *standard* [ɪŋ] e a pronúncia alveolar/dental [ɪn], estigmatizada. O autor conclui que há uma correlação entre o sexo/gênero dos informantes e o emprego dessas variantes: a pronúncia prestigiada é preferida entre as mulheres, enquanto a forma estigmatizada tende a ser evitada por elas, e favorecida entre os homens. Diversos outros trabalhos demonstram o mesmo tipo de correlação. São exemplos os estudos de Trudgill (1974) e Horvath (1985) a respeito da mesma variável em Norwich e Sydney, respectivamente; além de

⁵ Os termos “sexo” e “gênero” encerram noções distintas: enquanto o primeiro remete a uma diferenciação biológica (dividindo os indivíduos entre “homens” e “mulheres”); o termo “gênero” leva em consideração características da estrutura social e cultural em que os indivíduos se encontram inseridos (Cheshire, 2002). Além disso, não existem apenas dois gêneros, como pressupõe a divisão biológica homens *versus* mulheres. No entanto, como o termo gênero assume outras acepções na língua (gênero discursivo, por exemplo), adota-se a denominação “Sexo/Gênero”.

Wolfram (1969); Wolfram & Fasold (1974)⁶, Labov (1972), dentre muitos outros⁷.

No PB, Mollica et al. (1989; apud. Paiva, 2003), verificam a mesma atuação da variável Sexo/Gênero em um estudo da supressão da consoante vibrante /r/ - em expressões como “**problema**” e “**proprietário**” (que variam com as formas “**pobrema**” e “**propietário**”, respectivamente). As autoras observam que a variante estigmatizada é favorecida entre os homens. O mesmo se verifica em estudos que tratam da concordância nominal: a variante sem concordância (estigmatizada) tende a ser evitada pelas mulheres (Scherre, 1996; apud. Paiva, 2003).

Uma objeção, porém, pode ser feita aos estudos que trataram da correlação da variável Sexo/Gênero. Como foi visto acima, a grande maioria desses estudos adota uma classificação binária. De acordo com Eckert (1989: 246-7; apud. Chambers, 1995: 104), a variação observada entre homens e mulheres deve-se a questões de gênero, sendo a correlação com o sexo do indivíduo apenas indireta. Ainda de acordo com a autora, uma oposição binária entre homens e mulheres tem sido preferida devido à facilidade em se obter os dados para as pesquisas:

Although differences in patterns of variation between men and women are a function of gender and only indirectly a function of sex (and, indeed, such gender-based variation occurs within, as well as between, sex groups), we have been examining the interaction between gender and variation by correlating variables with sex rather than gender differences. This has been done because although an individual's gender-related place in society is a multidimensional complex that can only be characterized through careful analysis, his or her sex is generally readily observable binary variable... Because information about the individual's sex is easily accessible, data can be gathered without any inquiry into the construction of gender in that community.

(Eckert, 1989: 246-7; apud. Chambers, 1995: 104 – Grifos nossos)

No que diz respeito à alternância entre formas pronominais, a variável Sexo/Gênero também tem se mostrado atuante, e geralmente recebeu um enfoque

⁶ Todos citados por Chambers (1995: 102).

⁷ Labov (1972), por exemplo, afirma que: “*In careful speech, women use fewer stigmatized forms than men, and are more sensitive than men to the prestige pattern*” (Labov, 1972: 243); enquanto “*In stable sociolinguistic stratification, men use a higher frequency of non-standard forms than women*”. (Labov, 1990: 205).

metodológico “clássico” na literatura (“homens” *versus* “mulheres”). Laberge (1977; apud. Paiva, 2003: 35), em um estudo da variação entre *vous* e *on* no francês de Montreal (Canadá), depreende que a variante padrão *vous* tende a ser empregada pelas mulheres. Conforme foi mencionado no Capítulo I, em comunidades localizadas na região Sul do Brasil, onde a variante *tu* é prestigiada, as mulheres tendem a empregá-la (Loregian-Penkal, 2004; apud. Scherre, 2010-2013), ao passo que em outras regiões brasileiras, como o Centro-Oeste e o Sudeste, são os homens que favorecem o emprego dessa variante (Dias, 2007; Paredes Silva, 2003).

Andrade (2004) demonstra que os adolescentes brasilienses (entre 10 e 14 anos de idade) tendem a empregar a variante reduzida *cê* (.61), que é desfavorecida pelas meninas (.40). Em uma análise dessa mesma variável em Belo Horizonte, considerando informantes de distintas faixas etárias, Peres (2006: 157-158) constata uma tendência oposta: são as mulheres que favorecem o emprego de *cê*. A liderança das mulheres no emprego da variante reduzida *cê* possivelmente se dá pelo fato de não existir um estigma a ela associado – o que se coaduna à hipótese de que as mulheres tendem a liderar processos de mudança linguística, desde que a variante inovadora não seja avaliada negativamente pelos integrantes da comunidade (Labov, 2006 [1966])⁸.

Na comunidade de fala paulistana, aqui investigada, as variantes *você* e *cê* aparentemente não apresentam qualquer estigma ou prestígio a elas associado. Entretanto, embora não seja estigmatizada, a variante reduzida *cê* não se encontra registrada em gramáticas tradicionais ou livros didáticos, e os falantes a evitam na língua escrita formal⁹. Esse fato leva Ramos (1997) e Gonçalves (2008) a considerar *cê* e *ocê* como variantes “não-padrão”; e *você* como a forma “padrão”. Essas considerações são importantes para uma melhor interpretação do papel do gênero do falante no emprego das variantes pronominais em estudo, e serão retomadas no capítulo III.

⁸ A variante estigmatizada *ocê* também tende a ser evitada pelas mulheres nos diletos em que ocorre (Gonçalves, 2008; Peres, 2006; Andrade, 2004).

⁹ Nos diversos meios eletrônicos de comunicação escrita (*blogs*, *e-mails*, sites de relacionamentos, dentre tantos outros.), é bastante comum a abreviação de “você” pelas letras “vc” e, às vezes, de “cê” por “c”.

Outra variável de interesse ao estudo de correlações sociais é a Escolaridade. Como se sabe, a escola assume um papel conservador, prescrevendo os usos linguísticos prestigiados entre os informantes de maior *status* social. Esses empregos são apresentados como os únicos “corretos”, em detrimento de outras variantes (Votre, 2003). Sendo assim, pressupõe-se que indivíduos com um grau maior de escolarização tenderiam a empregar as variantes de maior prestígio (sobretudo se prescritas pelos compêndios gramaticais) e evitariam o uso de variantes estigmatizadas.

No caso do pronome em estudo, *você* é a única forma descrita pelas gramáticas tradicionais, dicionários e livros didáticos, como uma “forma de tratamento”. Por outro lado, a redução deste pronome não é enfatizada pela escola, como o são outros fenômenos, tais como a concordância nominal ou verbal (*os bancos x os banco; eles gostam x eles gosta*) e a regência verbal (*fui ao shopping x fui no shopping*). Ainda assim, o uso de *cê*, como de *tá* e de outras formas reduzidas na escrita, certamente é alvo de correção por parte do professor.

A correlação da variável escolaridade ao emprego de variantes não estigmatizadas nem prestigiadas foi verificada também em diferentes fenômenos linguísticos analisados no *corpus* PEUL (Programa de Estudos sobre o Uso da Língua). Em uma retrospectiva desses trabalhos, Paiva & Scherre (1999: 217) afirmam que:

[...] de um ponto de vista estatístico, a variável escolaridade atua igualmente sobre fenômenos menos marcados socialmente, como a monotongação de [ey], uso de artigo diante de possessivo ou uso de seu/dele, que não são alvos explícitos de um ensino regular. Este fato leva a crer que a escola, além de fazer uma pressão direta sobre o uso linguístico, atua também de forma indireta, ao iniciar e inserir o falante em uma modalidade mais rígida e mais homogênea como a escrita.

A faixa etária do informante, por sua vez, é uma questão amplamente investigada na literatura sociolinguística, uma vez que permite detectar a possibilidade de mudanças em progresso em uma comunidade de fala. Pela hipótese denominada “clássica”, a gramática do indivíduo se tornaria estável por volta dos 15 anos de idade (cf. Naro, 2003a), de modo que o estudo de informantes de distintas faixas etárias refletiria distintos “estados da língua”: um

indivíduo com 50 anos de idade em 2010, por exemplo, refletiria a fala de 35 anos atrás (i.e., 1975). Isto fundamenta o constructo do *tempo aparente* (Labov, 1994; Paiva & Duarte, 2003; Bailey, 2002), dissociando-se a oposição feita por Saussure (1969 [1916]) entre sincronia e diacronia (cf. Paiva & Duarte, 2003).

Mudança ou variação estável? – Tempo aparente e tempo real

De acordo com a perspectiva da sociolinguística variacionista laboviana, as mudanças se dão de maneira gradual em uma língua e pressupõem uma fase de variação anterior. No entanto, nem toda variação resulta necessariamente em uma mudança linguística (Weinreich et al., 1968; Bailey, 2002; Labov, 1994) – i.e., a substituição completa de uma variante pela outra, como no caso de *vós* por *vocês*. Há processos de variação que se mantêm estáveis por longos períodos de tempo.

No caso de uma mudança, vários estudos identificam um padrão, denominado curva-S (Wang & Cheng, 1970; apud. Chambers, 2002), segundo o qual a implementação da mudança é lenta nos estágios iniciais, ganha um impulso durante o processo (expandindo-se para comunidade mais geral) e volta a uma velocidade mais lenta ao término do processo (cf. Chambers, 2002).

Além disso, de acordo com o *Princípio do Uniformitarismo*, proposto por Labov (1994), os mesmos mecanismos que levam a mudanças atualmente são responsáveis por mudanças ocorridas em estágios anteriores da língua. Este princípio mostra-se útil para explicar os processos de gramaticalização, como o do pronome ora investigado (v. item 2.2, neste capítulo). Todavia, uma questão a ser colocada é: se esses mecanismos se repetem ao longo do tempo, por que tais processos afetam uma determinada variedade e não outra? Por que ocorrem em um momento determinado e não em outro? A resposta a estas questões pode tanto ser encontrada na estrutura social quanto na estrutura linguística.

O constructo do tempo aparente pode ser útil para depreender possíveis mudanças em uma língua. No entanto, a confirmação dos resultados de um estudo dessa natureza só se dá mediante a comparação destes aos de uma abordagem em tempo real (Labov, 1994; Chambers, 2002; Bailey, 2002). Considerando-se um mesmo fenômeno em dois períodos de tempo, pode-se detectar um maior ou

menor avanço de um processo de variação ou a sua estabilidade. Tal estudo pode se valer de um amplo espaço de tempo (séculos, por exemplo) ou de um intervalo menor, como de algumas décadas. Neste último caso, tem-se o chamado “estudo em tempo real de curta duração”.

Um estudo em tempo real pode ser do tipo *painel* ou *tendência*. No primeiro caso, os mesmos informantes são contatados em dois períodos de tempo – a exemplo do que fizeram os pesquisadores do PEUL (Programa de Estudo sobre o Uso da Língua) no Rio de Janeiro (cf. Paiva & Duarte, 2003). No estudo do tipo *tendência*, analisam-se dois *corpora* da mesma comunidade de fala em períodos diversos, com informantes distintos. No presente estudo, serão contrastadas amostras dos anos 1970 e dos anos 2000, cujos informantes são distintos.

A análise em tempo real pode evidenciar a implementação de uma mudança de duas maneiras: pelo aumento da taxa de *input* – i.e., da taxa de emprego de uma determinada variante relativamente às outras formas alternativas (Kroch, 1989; apud. Paiva & Duarte, 2003: 15) – ou pela ampliação dos contextos de uso dessa variante (Labov, 1972; apud. Paiva & Duarte, 2003: 15). No primeiro caso, os contextos de uso de uma dada variante permanecem inalterados, mas há um aumento constante na taxa de *input*. No segundo caso, a taxa de emprego das variantes permanece constante, mas uma delas vai ampliando seu contexto de uso ao longo do tempo. Como lembram Paiva & Duarte (2003), ambos os tipos não são mutuamente excludentes. No caso do pronome em estudo, por exemplo, a análise em tempo real desenvolvida visa a detectar: (i) se haveria um aumento na taxa de emprego de *cê* dos anos 1970 aos anos 2000 na cidade de São Paulo; e (ii) se a variante inovadora estaria se especializando em algum contexto ou ampliando seu uso a outros contextos¹⁰.

Além de evidenciar que diferentes estágios da língua podem conviver em uma mesma sincronia, a análise da variável faixa etária pode indicar uma gradação etária, i.e., quando o falante vai alterando seus padrões de uso de formas

¹⁰ De acordo com Labov (1994) e Bailey (2002), apenas a conjugação de um estudo em tempo aparente e em tempo real possibilita um estudo mais apurado dos processos de variação e de mudança linguísticas.

variantes ao longo de sua vida (Labov, 1994; Bailey, 2002), conforme os usos “mais adequados” à sua faixa etária ou por pressões como o “mercado linguístico”. Neste caso, em vez de uma mudança em progresso, verifica-se um *padrão curvilinear* quanto à faixa etária do informante (Paiva & Duarte, 2003): os grupos etários mais jovens e mais velhos aproximam-se (favorecendo o emprego de uma determinada variante), enquanto aquelas de uma faixa etária intermediária desfavorecem esta mesma variante. De acordo com Paiva & Duarte (2003), o estigma social associado a uma das formas e o mercado linguístico podem ser os responsáveis por esse padrão.

A fim de verificar se a variação entre *você* e *cê* em São Paulo constitui um caso de variação estável ou de mudança em progresso, esse pronome é aqui analisado em duas amostras (separadas por um intervalo de 30 anos), de acordo com o que será explicitado mais adiante.

Envelope de Variação

Todo estudo variacionista requer a definição de um envelope de variação. Isso significa que é necessário verificar os contextos em que as variantes podem ser comutadas uma pela outra, mantendo-se o mesmo “significado referencial” ou “valor de verdade” (Labov, 1978). Em se tratando de estudos no campo da fonologia, esse requisito não apresenta grandes desafios. No caso da pronúncia do /r/ em coda silábica, por exemplo, se uma palavra é pronunciada como po[r]ta, po[ɹ]ta ou po[h]ta, ainda que cada uma dessas três variantes possa veicular um valor social específico, a comutação de uma variante pela outra não implica mudança de significado.

À medida que se distancia do nível fonético-fonológico, a definição do envelope de variação vai se tornando cada vez mais árdua. Lavandera (1978), por exemplo, defende que a manutenção do valor de verdade ou significado referencial (Labov, 1972; 1978) só é plenamente possível no nível fonológico. Para outros níveis, a autora propõe o que chama de “comparabilidade funcional”, ou seja, os enunciados devem ser equivalentes em determinados contextos, ainda que os significados referenciais sejam distintos. Os exemplos a seguir ilustram

quão difícil é definir as variantes de uma variável e os contextos variáveis em determinados casos. A indeterminação do sujeito como variável dependente, por exemplo, requer considerar diversas variantes: *a(s) pessoa(s)*, *a gente*, *nós*, *você/cê*, *se*, como os exemplos (6a-e) evidenciam.

- (6)
- a) **A(s) pessoa(s)** vive(m) com muito medo em São Paulo.
 - b) **A gente** vive com muito medo em São Paulo.
 - c) **Nós** vivemos com muito medo em São Paulo.
 - d) **Você** vive com muito medo em São Paulo.
 - e) Vive-**se** com muito medo em São Paulo.

De um lado, em *a(s) pessoa(s)*, *se* e *você/cê*, o locutor não se inclui na referência, mas, ao pronunciar *a gente* ou *nós*, há necessariamente a sua inclusão (Ilari et. al., 1993) – o que dificultaria considerar todas estas formas como variantes de uma mesma variável.

O emprego do pronome de segunda pessoa *você(s)*, por sua vez, constitui um caso intermediário entre os dois exemplificados anteriormente. Trata-se de uma variável morfo-fonológica¹¹, uma vez que não se trata do mero apagamento dos segmentos /vo/ em uma das formas analisadas, mas sim da alternância entre dois itens distintos: o pronome forte *você(s)* e a variante reduzida¹² *cê(s)*, que encerram o mesmo significado referencial, ao funcionar, ambos, como pronome de segunda pessoa do singular/plural. No entanto, por questões outras, *você(s)* nem sempre pode ser comutado por *cê(s)* (e vice-versa), em todos os contextos. Este assunto é retomado no Capítulo III, quando se define o envelope de variação deste estudo.

¹¹ Adota-se, portanto, um ponto de vista formal para delimitar a variável dependente e as variantes em estudo: a redução ou não do pronome de segunda pessoa *você(s)*. Não são analisadas, portanto, outras possíveis formas alternativas como ‘o senhor(a)’, ou mesmos formas de indeterminação do sujeito (quando se trata da referência específica) – neste último caso, por exemplo, ter-se-ia uma variável semântico-pragmática.

¹² Alguns autores, como Petersen (2008), consideram *cê* uma forma fraca (seguindo a classificação proposta por Cardinaletti & Starke, 1999); outros como Vitral (1996, 2002, 2006a, 2006b) consideram-no um clítico ou um elemento em processo de cliticização.

A noção de comunidade de fala

Contrariamente às teorias linguísticas formalistas, que se voltam para o estudo das percepções de um único indivíduo acerca de sua língua materna, na sociolinguística variacionista, estuda-se a língua em uso e os indivíduos são importantes na medida em que integram uma comunidade de fala. De acordo com Labov (2006: 05), os indivíduos são aqui concebidos como “produto de suas histórias sociais” e a “a comunidade de fala é anterior ao indivíduo”.

A noção de “comunidade de fala”, porém, não encontra uma definição única na literatura. Com base nas várias acepções que o conceito vem adquirindo ao longo dos estudos sociolinguísticos, Fernández (1997a: 19) apresenta a seguinte definição¹³:

Una **comunidad de habla** está formada por un conjunto de hablantes que comparten efectivamente, al menos, una lengua, pero que, además, comparten un conjunto de normas y valores de naturaleza sociolingüística: comparten unas mismas actitudes lingüísticas, unas mismas reglas de uso, un mismo criterio a hora de valorar socialmente los hechos lingüísticos, unos misimos patrones sociolingüísticos.

Uma comunidade de fala, portanto, pode ser definida como “um conjunto de falantes que compartilham uma ou mais línguas”, bem como os “mesmos padrões sociolinguísticos” (que podem ser interpretados tanto em termos de produção, como em termos de avaliação linguística por parte dos falantes).

No entanto, vários problemas podem-se colocar em torno de um conceito de comunidade de fala abrangente como este. Patrick (2002), ao traçar um panorama dos vários estudos que se utilizam dessa noção, considera vários problemas, dentre os quais se destacam: (i) questões de escala (ou seja, o tamanho da comunidade considerada); (ii) problemas relativos à uniformidade (o quanto de normas/regras linguísticas e/ou sociais deve ser compartilhado pelos falantes, em termos de produção e/ou avaliação); (iii) a falta de uma tipologia para denominar os diferentes “tipos” de comunidades; e (iv) a necessidade de se investigar “comunidades dentro de comunidades” e “comunidades sobrepostas”.

¹³ Grifos nossos.

No que tange à questão de escala, observa-se que, sob o rótulo de “comunidade de fala”, têm sido classificados distintos agrupamentos humanos – desde grupos pequenos e localizados a grupos amplos e diversificados (Patrick, 2002). Com relação à uniformidade, coloca-se a questão de como definir o “quanto” de normas de produção e avaliação devem ser compartilhadas pelos indivíduos para que possam ser considerados pertencentes a uma “mesma” comunidade de fala.

Patrick (2002) enfatiza também a necessidade de se criar uma tipologia para os distintos tipos de comunidades. Uma tentativa a este respeito é apresentada por Romaine (1982; apud. Patrick, 2002), que distingue: (1) *indivíduo*; (2) *rede*; (3) *grupo social*; (4) *comunidade de fala*; e (5) *língua*. Hanks (1996; apud. Patrick, 2002), por sua vez, considera: (1) *falantes individuais*; (2) *diálogo entre duas pessoas*; (3) *interação face-a-face em grupos*; (4) *comunidades de prática*; e (5) *comunidades amplas*. Dentre estas classificações, destacam-se as “comunidades de práticas” e as “redes sociais”.

Uma “comunidade de prática” corresponde a um conjunto de indivíduos unidos em torno de um “empreendimento” comum, que se reúnem com certa frequência na busca por aquele ideal, compartilhando práticas linguísticas e sociais (Eckert & McConnell-Ginet, 1992; apud. Meyerhoff, 2002).

De acordo com Milroy (2002), o conceito de “redes sociais” emergiu entre os anos 1960 e 1970 no campo da Antropologia, e tem sido utilizado em diferentes áreas de pesquisa. Nos estudos sociolinguísticos, tal constructo foi adotado de modo mais intensivo somente a partir da década de 1980 – cf. Milroy (2002). De maneira ampla, uma rede social pode ser definida como um conjunto de indivíduos que mantêm ligações com pelo menos um membro em comum (o chamado “âncora” ou “ego” – cf. Milroy, 2002), e pode constituir uma “rede densa e múltipla” ou “redes de baixa densidade”.

Como se pode notar, o conceito de “redes sociais” aproxima-se bastante do de “comunidades de práticas”, com a diferença de que apenas estas últimas requerem a existência de um empreendimento comum entre os falantes: embora nem toda rede social seja uma comunidade de prática, uma comunidade de prática

pode ser concebida como um “tipo particular de rede social”. De maneira semelhante, as “redes sociais” e as “comunidades de práticas” podem ser consideradas subtipos de “comunidades de fala” – concebendo esta última como um hiperônimo, que abarca aquelas comunidades de nível “micro”.

Nesta pesquisa, adota-se o conceito de “comunidade de fala” em sua perspectiva “macro”, concebendo a cidade de São Paulo como uma única “comunidade de fala” – a exemplo do que Labov (2006 [1966]) propôs para a cidade de Nova York. Alguns questionamentos, contudo, podem ser feitos a este respeito, como já se discutiu em Nascimento (2008b: 03):

Definir o que seria “a comunidade de fala paulistana”, todavia, não é uma tarefa simples. Primeiramente, estamos diante de uma cidade ampla e diversificada. Em São Paulo, verifica-se, na verdade, uma ampla gama do que se poderiam considerar “pequenas comunidades” ou “sub-comunidades”, como a comunidade de imigrantes italianos, a comunidade de imigrantes japoneses, comunidades da periferia paulistana, etc. Diante de uma grande miscelânea de indivíduos, cabe a pergunta se, de fato, estaria correta a afirmação de que este se trata de um estudo da “comunidade de fala paulistana”. Poder-se-ia questionar, por exemplo, sobre quais seriam as características (linguísticas e sociais) que identificam um indivíduo como membro desta “comunidade”.

Neste estudo, parte-se do princípio de que, mesmo se tratando de um amplo e diversificado centro urbano cosmopolita, os paulistanos diferenciam intuitivamente o seu modo de falar “paulistano” de outros falares, como o “paulista”, o “gaúcho”, o “carioca”, o “mineiro”, etc. Em outras palavras, utiliza-se aqui o conceito de comunidade de fala para o que se convencionou chamar “dialetos regionais”. Foram considerados paulistanos aqueles indivíduos que nasceram e foram criados em São Paulo, ou que tenham nascido em outras localidades, mas que aqui se fixaram antes de completarem um ano de idade, independentemente da origem regional de seus pais.

Ao assumir essa perspectiva, não se está excluindo a possibilidade de estudos de nível “micro” dos vários segmentos sociais e urbanos dessa cidade, ou seja, um enfoque de “sub-comunidades” dentro dessa comunidade maior e que

considere questões de natureza etnográfica/antropológica, mas apenas delimitando o campo desse estudo, de acordo com os objetivos desse estudo¹⁴.

Padrões quantitativos e metodologia

Para comprovar a sistematicidade da variação linguística, a Sociolinguística adota uma metodologia quantitativa de análise, que permite averiguar se as hipóteses aventadas pelo pesquisador correlacionam-se ou não ao emprego das variantes em estudo e, a partir dos resultados obtidos, depreender como se dá o encaixamento social e linguístico de um fato variável.

O programa estatístico VARBRUL¹⁵ tem tido um amplo emprego na literatura variacionista (Guy & Zilles, 2007; Scherre & Naro, 2003). Tal programa opera com um modelo matemático de regressão logística (cf. Naro, 2003b), que permite desenvolver uma análise multivariada – na qual, computa-se o efeito de uma determinada variável (grupo de fatores), levando-se em consideração o conjunto de todas as outras variáveis investigadas naquela pesquisa (cf. Guy, 2007 [1993]).

Pode-se dizer que tal análise opera tanto com uma *estatística descritiva* (Fernández, 1997b: 310) – que computa os dados, fornecendo, por exemplo, os números absolutos e as percentagens de uso das variantes –, quanto com uma “*estatística de inferências*” (Fernández, 1997b: 310) – que permite generalizar os resultados quantitativos para a comunidade de fala analisada como um todo. Este segundo tipo é expresso por medidas estatísticas como é o caso dos pesos relativos¹⁶, que refletem a tendência com que uma determinada variante é empregada nos diversos contextos investigados. Tal método estatístico possibilita ampliar o alcance da interpretação dos resultados obtidos: a partir de uma amostra reduzida, composta pela fala de uma parcela dos indivíduos de uma comunidade,

¹⁴ De acordo com Patrick (2002), o conceito de “comunidade de fala” não é uma “*uma entidade pré-definida*”, mas algo “*construído a cada nova pesquisa*” e que depende fundamentalmente dos interesses do pesquisador.

¹⁵ No presente estudo, adota-se o Varbrul em sua versão para Windows, Goldvarb X (Sankoff, Tagliamonte & Smith, 2005). <http://individual.utoronto.ca/tagliamonte/goldvarb.htm>

¹⁶ Os pesos relativos são medidas que variam de 0 a 1.0 e indicam a tendência com que uma determinada variante tende a ocorrer em um dado contexto. Valores acima de .5 indicam um favorecimento ao emprego de uma determinada variante, enquanto valores abaixo de .5 desfavorecem-na.

pode-se obter um padrão de variação replicável a toda a comunidade (Fernández, 1997b; Guy & Zilles, 2007).

O Varbrul indica quais grupos de fatores analisados apresentam ou não uma significância de um ponto de vista estatístico. Convencionalmente, um grupo de fatores é significativo, se há um valor de significância menor ou igual a 0.05, ou seja, se há menos do que 5% de chances da hipótese nula¹⁷ ser verdadeira (i.e., o grupo de fatores não estar correlacionado ao emprego das variantes); e, conseqüentemente, 95% de chances de que um determinado grupo de fatores correlacione-se ao fenômeno em estudo (cf. Guy & Zilles, 2007).

Mantendo-se os mesmos grupos de fatores em duas análises quantitativas com o Varbrul, é possível também calcular a significância estatística de versões diferentes desse grupo de fatores, através de um teste de *qui-quadrado* (cf. Guy & Zilles, 2007). A subtração dos valores do *log likelihood* (*logaritmo de verossimilhança*; outra medida estatística dada pelo Varbrul) de ambas as análises, multiplicada por 2, fornece o valor do *qui-quadrado*. A depender do grau de liberdade (i.e., a diferença entre o maior e o menor número de fatores analisados nas duas análises¹⁸), pode-se calcular a significância estatística daquele grupo de fatores¹⁹. No capítulo IV, são apresentados os resultados quantitativos em que se adota essa metodologia.

2.2. Gramaticalização

Além das questões teóricas em torno da análise variacionista a que se dedicou o item 2.1 acima, este capítulo também trata das bases teóricas em que se aventaram as hipóteses sobre a variação entre *você(s)* e *cê(s)*. Duas das hipóteses de natureza linguística (a cliticização da variante reduzida e a referência do *você/cê*) encerram-se no paradigma da gramaticalização, dadas as próprias origens e a evolução desse pronome, conforme relata brevemente o capítulo I.

¹⁷ A hipótese nula indica uma não-correlação entre a hipótese aventada e a variável (Guy & Zilles, 2007).

¹⁸ Mantendo-se inalteradas as demais variáveis investigadas.

¹⁹ No Anexo 2, há uma tabela com os valores de *qui-quadrado* em diferentes graus de liberdade (cf. Guy & Zilles, 2007).

A gramaticalização é entendida aqui como um tipo de mudança linguística na qual itens lexicais passam a desempenhar funções gramaticais, ou itens “menos gramaticais” tornam-se “mais gramaticais” ainda (Kurylowicz, 1965; apud. Moura Neves, 2001: 52)²⁰. A escala a seguir, de Hopper & Traugott (1993: 07), ilustra o *continuum* desse processo – concebido como unidirecional.

ITEM LEXICAL > PALAVRA GRAMATICAL > CLÍTICO > AFIXO FLEXIONAL

O sintagma nominal *vossa mercê* deu origem ao pronome pessoal *você*, desenvolvendo-se até a forma reduzida *cê* (considerada um clítico por alguns autores – como Vitral, 1996):

VOSSA MERCÊ > VOSSEMECÊ > [...] > VOCÊ > OCÊ > CÊ

As principais mudanças por que passam os itens em processo de gramaticalização situam-se em diferentes níveis linguísticos (cf. Gonçalves et. al, 2007). No plano fonético-fonológico, o item tende a sofrer *erosão*, ou seja, uma “perda de substância fonética” e uma “perda de proeminência acentual” (elementos tônicos podem tornar-se átonos). No plano semântico, destaca-se um processo denominado *bleaching* (“desbotamento semântico” ou “abstratização semântica”), via metáforas, metonímias e inferências pragmáticas. No nível morfossintático, formas plenas e com relativa independência sintática tornam-se “mais dependentes” contextualmente – é o que se ocorre, por exemplo, quando itens plenos tornam-se clíticos ou afixos (formas presas).

De um ponto de vista fonológico, *vossa mercê* foi reduzindo-se até as atuais formas (v)ocê (duas sílabas), e “cê” (uma sílaba). Morfossintaticamente, diversos autores defendem que *cê* estaria se cliticizando (Vital, 1996) e, com a continuidade do processo, poderia vir a se tornar um afixo, adjungindo-se ao verbo (Nascimento, 2009).

²⁰ Castilho (2006, 2009) propõe uma abordagem distinta para a gramaticalização, denominada “multissistêmica”. Essa proposta estabelece que os processos de Gramaticalização (na qual estão incluídos os subprocessos de Fonologização, Morfologização e Sintaticização), Semantização, Lexicalização e Discursivação seriam “*processos de criação linguística*” que ocorrem simultaneamente, “*não se devendo estabelecer entre eles relação de derivação nem de determinação*” (Castilho, 1997).

Sintaticamente, *vossa mercê* foi reanalisado como uma única expressão: recategorizou-se como uma forma de tratamento e, posteriormente, como um pronome pessoal. Com relação à sua semântica, *vossa mercê* (“vosso favor”) foi reinterpretado, através de processos metafóricos e metonímicos, como uma forma de tratamento ao rei (de “os favores prestados” para “aquele que prestava os favores”).

Outra característica de um item em processo de gramaticalização é o aumento de sua frequência de uso (frequência *token*) e o aumento da variedade de contextos que o item passa a integrar (frequência *type*) – Bybee (2002). Conforme mencionado no capítulo I, a tendência ao preenchimento do sujeito no PB pode ter contribuído para o uso mais frequente desse pronome (aumentando assim sua frequência *token*). Quanto à frequência *type*, a elevada frequência do pronome “você/cê” como recurso indeterminador do sujeito (Alves, 1998) pode ser indício de que o uso deste pronome estaria se expandindo aos contextos com referência indeterminada.

Mais ou menos gramaticalizado?

Lehmann (1995 [1982]; apud. Gonçalves & Carvalho, 2007: 70-78) estabelece um conjunto de critérios a fim de “medir” o grau de autonomia dos itens em estágios mais avançados do processo: quanto menos autônomo, mais gramaticalizado o item²¹.

Com base em três aspectos (*peso*, *coesão* e *variabilidade*), aplicados aos eixos paradigmático e sintagmático, o autor estabelece seis parâmetros:

- (i) integridade (*peso paradigmático*): o “peso” fonético-fonológico e semântico de um item; ao longo do processo de gramaticalização, o elemento passa por um desgaste tanto no plano da expressão (*erosão*) quanto no plano do conteúdo (*bleaching*); a expressão *vossa mercê* amalgamou-se e passou por reduções

²¹ Hopper (1991; apud. Gonçalves & Carvalho, 2007: 79-85) também formula princípios para “medir” o grau de gramaticalização de itens que se encontram em estágios mais recentes desse processo. Por tratar de um processo de gramaticalização mais antigo na língua, este trabalho discute apenas o modelo proposto por Lehman (1995 [1982]).

fonéticas até o atual pronome *você* e suas variantes *ocê* e *cê*; quanto à “perda de conteúdo semântico” (ou “abstratização semântica”), o valor inicial *vossa graça* / *vosso favor* deu lugar a um significado gramatical (2ª pessoa do singular);

- (ii) paradigmaticidade (*coesão paradigmática*): o “tamanho” do paradigma que o item integra; *você/cê* integra o paradigma dos pronomes pessoais, um paradigma “fechado”, contrariamente ao paradigma dos substantivos, como era o caso de *mercê*;
- (iii) variabilidade paradigmática: maior “obrigatoriedade” de uso do item relativamente a outras possibilidades de “escolha” existentes em um dado paradigma; em relação aos usos genéricos, Alves (1998), por exemplo, verifica uma maior frequência de uso de *você/cê* como uma estratégia de indeterminação do sujeito, comparativamente a outros recursos indeterminadores.
- (iv) escopo (*peso sintagmático*): itens mais gramaticalizados teriam um escopo menor do que itens menos gramaticalizados; tal fato é mais visível na gramaticalização de verbos plenos a auxiliares do que na gramaticalização de nomes.
- (v) conexidade (*coesão sintagmática*): a relação sintagmática que o item estabelece com outros elementos no sintagma; no caso do pronome em estudo, se a forma *cê* estiver se tornando um clítico, ela tenderia a vir preferencialmente em posição adjacente (*coalescente*) ao verbo, podendo a ele se adjungir futuramente, em um estágio mais avançado de gramaticalização.
- (vi) variabilidade sintagmática: grau de mobilidade do constituinte dentro do sintagma. Supõe-se que, quanto mais gramaticalizado o item, mais fixa a posição que ocupa.

O capítulo III traz análises em que esses parâmetros são levados em conta.

Cê: clítico ou forma fraca?

A forma pronominal reduzida *cê* tem sido analisada por alguns autores como decorrente de uma nova etapa no processo de gramaticalização. Considerando o percurso *item lexical* > *item gramatical* > *clítico* > *afixo* (Hopper & Traugott, 1993), *cê* pode ser concebido como um clítico, e algumas restrições ao seu emprego corroboram essa hipótese. No entanto, nem todos os autores concordam com essa classificação. Encontram-se três posicionamentos a este respeito: (i) *cê* é considerado um clítico (Vital, 1996; Ramos, 1997); (ii) *cê* é considerado uma forma fraca (Petersen, 2008); (iii) *cê* estaria passando por um processo de cliticização (Vital, 2006b).

A hipótese (i) apóia-se em restrições sintáticas, (supostamente) categóricas, ao uso da variante *cê*. Essa hipótese também tem sido investigada de um ponto de vista variável por pesquisadores como Ramos (1997), Gonçalves (2008) e Peres (2006), que observam que a presença de constituintes entre o pronome e o verbo (interpolação) desfavorece o emprego da variante reduzida.

O posicionamento (ii) fundamenta-se na tripartição pronominal, proposta por Cardinaletti & Starke (1999; apud Petersen, 2008: 289-290). Seguindo critérios fonológicos, morfológicos, sintáticos e semânticos, os autores dividem os pronomes em três subtipos: pronomes fortes, formas fracas e clíticos. Dentre as características que diferenciam os clíticos das formas fracas está o fato de que aqueles devem se apoiar sobre outro constituinte (normalmente o verbo), não ocorrem como complemento de preposição, nem podem receber foco ou ênfase, dado seu caráter átono. O fato de *cê* “violar” alguns destes critérios, leva Petersen (2008) a classificá-lo como uma *forma fraca*.

De acordo com a autora, *cê* é a uma forma fraca porque “*apresenta um comportamento muito distinto dos demais clíticos do PB*” (Petersen, 2008: 289): pode ocorrer como enclítico ao verbo (*Vou fazer cê feliz²²*) e permite a inserção de outros constituintes interpolados entre o pronome e o verbo (*cê realmente gosta desse restaurante.*).

²² Exemplo de Vital (1996).

Os argumentos quantitativos apresentados por Ramos (1997) em favor da cliticização de *cê* também são questionados por Petersen (2008), para quem os resultados “*não comprovam o estatuto de clítico da proforma cê*”, uma vez que “*não excluem a possibilidade de se tratar de outro tipo de pronome*” (Petersen, 2008: 289) – i.e., uma forma fraca.

Vital (2002, 2006a), com base em evidências fonéticas, assume uma posição intermediária: *cê* estaria “em processo de cliticização”²³ – cf. hipótese (iii) acima. O autor considera problemática uma divisão estrita entre “clíticos” e “formas fracas”, pois:

os termos “clítico” e “forma fraca” não têm estatuto explicativo real e [...] a cliticização é mais bem compreendida como um processo gradual e heterogêneo que prevê mais etapas do que as que normalmente têm sido admitidas na literatura.

(Vital, 2002: 389)

Adicionalmente, a interpolação (presença de constituintes entre o pronome e o verbo) era um fenômeno bastante comum no português arcaico com elementos como *se*, *me*, *lhe*, etc. (Martins, 1996; apud. Vital, 2006b). Os exemplos (7-10), de Martins (1996; apud. Vital, 2006b: 96) ilustram como distintos constituintes sintáticos podiam ocupar uma posição intermediária em relação ao complexo “clítico + verbo”.

- (7) Que [se adeãte segue]. (1355)
- (8) que [sse nõ deue] a estender esse maravedi. (1296)
- (9) o quall pardieiro [lhe logo emprazarô]. (1494)
- (10) que [as Ao dicto Mosteiro deuiã] Alguas pessoas. (1357)

Existem, assim, evidências históricas de que o processo de cliticização ocorreria em um *continuum* – o que apóia a hipótese de que *cê* estaria passando por um processo de cliticização semelhante no PB atual²⁴ (Vital, 2002, 2006a,

²³ Andrade (2004) também apresenta uma discussão a este respeito e chega à conclusão de que *cê* aproxima-se de uma forma fraca atualmente; porém, essa autora também partilha da concepção de Vital (2002), para quem *cê* estaria passando por um processo de cliticização.

²⁴ No português europeu (PE) atual, a interpolação restringe-se a registros mais formais da escrita (Martins, 1996; apud. Vital, 2006b).

b). Na presente pesquisa, assume-se esta hipótese, que será testada quantitativamente, com um grupo de fatores que é descrito no capítulo III.

Como já se mencionou anteriormente, outros processos de cliticização são apontados na literatura: a redução do pronome de terceira pessoa (*eles* > *eis/es*) – Ramos & Corrêa (2006); a variação entre *a gente*, *a'ente* e *'ente* (Zilles, 2002); a redução do advérbio de negação “não” em posição pré-verbal, (*não* > *num*) – Ramos (2006); e a redução do auxiliar “*estar*” em perífrases com o gerúndio “*está fazendo* > *tá fazendo*” (Mendes, 1999; 2008).

Nascimento (2009), ao revisitar os trabalhos anteriores, observa que, com exceção do “*não/num*” pré-verbal, os itens reduzidos *cê*, *eis*, *'ente* e *tá* (reduções de *você*, *eles*, *a gente* e *estar*, respectivamente) permitem a interpolação de constituintes entre eles e o verbo. Além disso, a comparação dos resultados demonstra que as formas *cê*, *num* e *tá* tendem a ser evitadas em estruturas com interpolação (especialmente com elementos não-clíticos) – o que reforça a hipótese de que estariam também em um processo de cliticização em curso no PB atual. Tais fenômenos apontam para a atuação do *Princípio do Uniformitarismo* (Labov, 1994), segundo o qual os mesmos princípios de mudança podem ser atuantes em diferentes estágios da língua – v. capítulo I.

As discussões anteriores servem de fundamento para o grupo de fatores Coalescência do pronome ao verbo e também se ligam de maneira menos direta a hipóteses de condicionamentos semânticos e fonético-fonológicos, que se apóiam na hipótese de que a variante reduzida do pronome estaria se cliticizando no PB. Essas questões serão tratadas qualitativamente no capítulo III, e testadas quantitativamente no capítulo IV.

***Bleaching* e a noção de Referência**

Alguns autores que tratam da gramaticalização defendem uma correlação entre *erosão* (perda de material fonético) e *bleaching* (perda / abstratização de conteúdo semântico). Desta assunção, resultaria que, quanto mais gramaticalizado

um item, tanto menor seria sua substância fonética e menos concreto o seu conteúdo semântico.

Se a erosão fonética é um processo indiscutivelmente visível, a noção de *bleaching*, entretanto, não é tão fácil de ser definida. De um lado, autores como Heine & Reh (1984; apud. Hopper & Traugott, 1993:87) caracterizam a mudança semântica como uma “dessemantização”, i.e., uma “perda” ou “enfraquecimento” dos traços semânticos do item. De outro lado, Castilho (1997) posiciona-se contrário à noção de *dessemantização* ou “perda” de traços semânticos. Para o autor, trata-se apenas de “*modificações de sentidos*”.

Hopper & Traugott (1993: 93) reconhecem a dificuldade em se precisar o fenômeno:

‘Bleaching’ must therefore be taken to be a very relative notion, and one that pertains almost exclusively to late stages of grammaticalization. An important question for the future research is what exactly constitutes bleaching, when it can be said to set in, and how it correlates with morphosyntactic generalization.

A noção de *bleaching* geralmente encontra-se associada a um *continuum* que vai de um significado “mais concreto” a outro “mais abstrato”. Croft (1990; apud. Hopper & Traugott, 1993:157), a partir dos traços [$\pm$ definido] e [$\pm$ referencial], estabelece a seguinte escala²⁵:

$$+ \text{def} / + \text{ref} > - \text{def} / + \text{ref} > - \text{def} / - \text{ref}$$

Por esse percurso, o autor sugere haver primeiro uma abstratização semântica, devido à perda do traço [+ definido], que conduziria, posteriormente, à formação de expletivos (desprovidos também de referencialidade).

Vital & Ramos (2006: 86 [1999]) refinam a escala anterior, adotando para isso as noções de “referencialidade virtual” e “referencialidade real” (Milner, 2003 [1982]). A “referência virtual” independe do ato de enunciação, pois nos remete a uma determinada ‘classe de indivíduos/seres’. Por exemplo, o substantivo *gato* refere-se virtualmente a um mamífero, felino, de quatro patas,

²⁵ def = definido; ref. = referencial.

etc. Ao ser atualizado no discurso, esse nome adquire também uma “referência real”: *O gato da Ana sumiu.*

Os pronomes dêiticos de primeira e segunda pessoa do singular, entretanto, não possuem uma referência virtual. Benveniste (1991: 281) os define como aqueles que “[...] não podem existir como signos virtuais, não existem a não ser na medida em que são atualizados na instância do discurso [...]”²⁶. Tais elementos, portanto, apresentariam apenas uma referência real (recuperada pelo discurso), mas não uma virtual²⁷.

A partir dessas considerações, Vitral & Ramos (2006 [1999]) propõem a seguinte escala para a gramaticalização de *vossa mercê*:

$$\text{Referência virtual} > \text{Referência real} > \text{sem referência} \\ [+ \text{ def} > - \text{ def}]$$

Conforme os autores, o sintagma nominal *vossa mercê* era dotado de ambas as referências (real e virtual). Quando este se gramaticaliza como o pronome pessoal *você*, perde sua referência virtual e permanece apenas com uma referência real – o que corresponde à primeira etapa da escala acima: *Referência virtual > Referência real*). Na etapa final do processo, haveria a perda das duas referências (virtual e real) – o que levaria à formação de expletivos.

Cabe enfatizar, todavia, que esse processo se dá em um *continuum*. Sendo assim, a transferência de uma significação para outra costuma ser acompanhada por fases intermediárias, que se caracterizam por ambiguidades semânticas e estruturais – ou, como se prefere nesta pesquisa, significações híbridas:

“(...) A transferência de um sentido ‘literal’ para outro ‘figurado’ e o de um domínio de conceptualização para outro promovem o deslizamento de um sentido mais concreto para um mais abstrato. Essa movimentação normalmente é intermediada por uma ambiguidade semântica.

(Heine et. al., 1991. Trad. : Gonçalves et al., 2007: 33 – Grifos nossos)

²⁶ Grifos nossos.

²⁷ A referência virtual é entendida como o significado lexical de um item (Milner, 2003: 86). Como os dêiticos apresentam uma significação gramatical (*você* / *cê* = ‘segunda pessoa do singular’), mas não uma lexical, eles não possuiriam tal referência.

Além de se referir ao interlocutor, como em (11), o pronome *você/cê* (2ª p. sg.) pode apresentar uma interpretação abrangente, referindo-se não só à pessoa com quem se fala, mas às pessoas em geral, como em (12)²⁸.

- (11) Que mais, **Van**, que **cê** que(r) sabe(r)? (Jéssica-F1FI-2007)
- (12) Hoje em dia ou **você** economiza e paga um convênio médico, ou **cê** corre pro PS de bairro em bairro. (Jéssica-F1FI-2007)

Moura Neves (2000: 463), explica assim a noção de “fazer referenciação genérica” com o pronome “você”:

O pronome **VOCÊ**, embora seja forma de pessoa envolvida no discurso (**segunda pessoa**), pode indicar referência genérica. A indeterminação nesse caso, é muito forte (**VOCÊ** = uma pessoa, seja qual for) [...]

Para os propósitos da análise variacionista, vale a pena propor que a referência pode ser analisada em um *continuum* e não apenas através de uma oposição binária (‘determinada’ *versus* ‘indeterminada’).

O emprego do pronome *você* em contextos indeterminados foi objeto de investigação de Corradello (1997), Alves (1998), e Sousa (2008) e tem recebido os mais diferentes rótulos na literatura linguística: “genérico” (Corradello, 1997), “indeterminado” (Corradello, 1997), “indefinido” (Ramos, 1997), “referência arbitrária” (Duarte, 2003)²⁹. Tal emprego não encerra apenas um sentido e uma referência, mas também apresenta funções discursivas, tais como estabelecer uma maior proximidade com o interlocutor, inserindo-o hipoteticamente no discurso e, em alguns casos, “preservar a face” do próprio locutor (v. capítulo III).

Frege (1978), ao tratar da referência³⁰ de nomes próprios³¹ e de sentenças declarativas, considera que os primeiros referem-se a um ser ou entidade

²⁸ A referenciação genérica ocorre apenas com a forma singular do pronome de segunda pessoa (*você/cê*). Por essa razão, na análise do grupo de fatores “Referência” (v. Cap. IV), os empregos genéricos não foram analisados.

²⁹ No presente estudo, adotam-se os termos “específico” e “determinado” como sinônimos; o mesmo se aplica aos termos “genérico” e “indeterminado”; evita-se os termos “definido/indefinido” (exceto quando se estiver reportando aos trabalhos dos autores que os adotaram), uma vez que, como considera Milanez (1982; apud. Corradello, 1997), o termo “indefinido” é reservado na semântica a expressões indefinidas de terceira pessoa.

³⁰ Frege (1978) opera uma distinção entre sentido e referência. O sentido não corresponde nem ao objeto situado no mundo (referente), nem à representação mental particular de cada interlocutor, mas a um conceito “único”. O clássico exemplo fornecido pelo autor é o da imagem da lua projetada em um telescópio: a lua

particular, situado no mundo real; ao passo que a referência de uma sentença declarativa corresponde ao seu valor de verdade (verdadeira ou falsa). Em uma sentença como “*O Lula venceu as eleições desse ano*”, por exemplo, o sintagma nominal “*O Lula*” refere-se a um indivíduo particular do mundo; e a sentença pode ser verdadeira ou falsa, a depender do momento de sua enunciação.

Apoiada em Lyons (1977), Cançado (2005: 79) destaca cinco tipos de referência que têm como expressão referencial um sintagma nominal (SN): *referência singular definida; sintagmas nominais definidos não-referenciados; referência geral distributiva e coletiva; referência indefinida específica e não-específica; e referência genérica*. Dentre estes, interessa particularmente a última. No exemplo a seguir, a referência do pronome não aponta unicamente para o interlocutor, mas se estende a uma classe de indivíduos: “os brasileiros”.

- (13) Mas como **o brasileiro** tem memória curta política... eu acho que isso acaba não acontecendo. Em quatro anos **você** esquece em quem **cê** votou... (Flávio-M2SI-2003)

No presente estudo, adota-se o termo “referência” tanto para a “referência determinada”, quanto para os usos “genéricos” do pronome *você/cê(s)*.

2.3. Concepções discursivo-pragmáticas para a referência

Os pronomes pessoais têm um caráter *fórico* (Moura Neves, 2000: 449), ou seja, sua referência se estabelece por meio da enunciação (*exofórico*) ou do enunciado (*endofórico*). No primeiro caso, têm-se os pronomes de primeira e segunda pessoas do singular (*eu; tu/você*), também chamados de *dêiticos*. Cervoni (1989)³² caracteriza estes últimos como “*símbolos-índices*”, pois dispõem tanto de um significado gramatical convencional (‘primeira pessoa do singular’ e

corresponde ao referente, propriamente dito; a imagem projetada na lente do telescópio é o seu sentido; enquanto a imagem projetada na retina de um observador é uma projeção mental particular e variável a cada novo observador.

³¹ A classificação “nomes próprios” de Frege se difere daquela apresentada pela gramática tradicional, na medida em que abarca não apenas nomes próprios propriamente ditos, mas também descrições definidas como “O atual presidente do Brasil”, “O meu vizinho”, uma vez que as mesmas apontam para um ser singular no mundo.

³² O autor considera a distinção feita por Jakobson (apud. Cervoni, 1989).

‘segunda pessoa do singular’, respectivamente) – daí serem classificados como *símbolos*³³ –, como também requerem a presença de um referente na situação de comunicação – daí o rótulo *índices*³⁴. No entanto, embora possuam um significado gramatical, sua interpretação é dependente do discurso, da enunciação.

Stalnaker (1982: 78; apud. Corradello, 1997: 69-70), por exemplo, opõe as questões semânticas intrínsecas aos termos às questões pragmáticas envolvidas no processo de referenciação. De acordo com o autor:

O referir-se é algo que as pessoas fazem com os termos e não algo feito pelos próprios termos. É por isso que a referência é um problema da pragmática, e é por isso que o papel de um termo singular depende menos da categoria sintática ou semântica do próprio termo (nome próprio, descrição definida, pronome) do que do locutor [,] do contexto, e das pressuposições do locutor naquele contexto.

Maingueneau (2004: 111), por sua vez, propõe três tipos de referência: (i) baseadas na *enunciação*; (ii) baseadas no *cotexto* [= texto, enunciado]; e (iii) referências que não se baseiam nem na *enunciação*, nem no *cotexto*.

Quando o pronome *você/cê(s)* se refere ao próprio locutor, o referente encontra-se na situação enunciativa – correspondendo ao primeiro tipo de referência proposto por Maingueneau (2004). O segundo caso necessita de uma correferência textual, a exemplo das anáforas e catáforas prototípicas. Ao passo que o terceiro (embora não especificado pelo autor) pode ser concebido como aqueles que requerem informações fora do âmbito contextual, como o “conhecimento de mundo” dos interlocutores, etc.

Com os usos genéricos do pronome *você/cê*, esse processo de referência não ocorre de maneira direta: o pronome não faz referência (unicamente) ao indivíduo com quem se fala. O ouvinte entende que a referência não se aplica a ele exclusivamente, mas a um conjunto mais abrangente de indivíduos, do qual ele é apenas um exemplo (Corradello, 1997).

Um processo semelhante ocorre em francês (Maingueneau, 2004: 126-127), em que o pronome *vous* (cuja tradução em português pode ser *você(s)*) pode

³³ Cf. a terminologia de Peirce, adotada também por Jakobson (apud. Cervoni, 1989).

³⁴ Cf. a terminologia de Peirce, adotada também por Jakobson (apud. Cervoni, 1989).

se referir a uma classe profissional, em uma situação “virtual”, estando “associado a um presente não dêitico em um fragmento não embreado (Maingueneau, 2004: 127).

Fiorin (2008: 27) assim distingue os processos de debreagem e embreagem³⁵:

Ao contrário da **debreagem**, que é a colocação fora da instância de enunciação da pessoa, do espaço e do tempo do enunciado, a **embreagem** é “o efeito de retorno à enunciação”, produzido pela neutralização das categorias de pessoa e/ou espaço e/ou tempo, assim como pela denegação da instância do enunciado. [...]

A **embreagem actancial** diz respeito à neutralização das categorias de pessoa. Toda embreagem pressupõe uma debreagem anterior. Quando o presidente diz “O Presidente da república julga que o Congresso Nacional deve estar afinado com o plano de estabilização econômica”, formalmente temos uma debreagem enunciativa (um *e/e*). No entanto esse *e/e* significa *eu*. Assim, uma debreagem enunciativa (instalação de um *eu*) precede a embreagem, a saber, a neutralização da oposição categórica eu/ele em benefício do segundo membro do par [...].

Conforme o autor, a “*embreagem actancial diz respeito à neutralização da categoria de pessoa*”, ou seja, ao emprego de uma pessoa em lugar da outra. Assim, é plausível pensar em um processo de embreagem actancial também para os empregos genéricos do pronome *você*³⁶.

Esta questão é retomada mais adiante, quando se trata das várias referenciações que o *você* genérico pode assumir no discurso. Antes, porém, convém esclarecer a acepção que “referência” assume em análises que se voltam para o discurso e os vários tipos de correferência.

Referenciação e correferência

Alguns trabalhos que adotam uma perspectiva discursivo-pragmática partem do princípio de que a construção do sentido e da referência (não apenas dos dêiticos, mas de qualquer expressão linguística) só se estabelece plenamente no transcorrer do discurso. A este respeito, Mondada & Dubois (2003: 17 [1995]) afirmam que:

³⁵ Grifos nossos.

³⁶ Isso é contrário à proposta de Maingueneau (2004), para quem os usos genéricos de *vous* não seriam casos de embreagem, por situarem-se fora do âmbito da enunciação.

[...] as categorias e os objetos de discurso pelos quais os sujeitos compreendem o mundo não são nem preexistentes, nem dados, mas se elaboram no curso de suas atividades, transformando-se a partir dos contextos.

Os mesmos autores preferem adotar o termo *referenciação*, em vez de *referência*:

[...] falaremos de *referenciação*, tratando-a, assim como a categorização, como advindo de práticas simbólicas mais que de uma ontologia dada. Como diz Rastier [1994: 19], a referenciação não diz respeito a “uma relação de representação das coisas ou dos estados de coisas, mas de uma relação entre o texto e a parte não linguística da prática em que ele é produzido e interpretado”.

(Mondada & Dubois, 2003: 20 [1995])

Há, portanto, uma distinção entre *referenciação* e *referência*. Enquanto esta última concepção se volta para a relação entre expressões linguísticas e os objetos situados ou encontráveis no mundo real (cf. a semântica formal clássica), a *referenciação* é um processo mais complexo que se produz no e pelo discurso, podendo ser alterada ou modificada a cada nova enunciação ou mesmo dentro de uma mesma enunciação (Koch & Marcuschi, 1998).

Ilari et al. (1993: 82) também partem do princípio de que a interpretação dos pronomes pessoais requer o estudo do texto (discurso) em que o mesmo é proferido:

O que parece certo é que a interpretação dos pronomes (não só dos pessoais, mas de todos os fóricos) é do âmbito da **semântica textual**, não da semântica da palavra, embora seja identificável em cada um dos fóricos (os pessoais e alguns outros pronomes) uma instrução muito firme para a interpretação. É que os fóricos incorporam uma indicação precisa de relações, seja com o ato discursivo, seja com outras porções da peça do discurso, situando-se a variedade na possibilidade de mudanças do elemento referido ou recuperado, não na relação em si.³⁷

Para estabelecer esse processo de referenciação via discurso, tais autores consideram a existência de *funções conexas à função anafórica* (e catafórica). Além da correferência prototípica (anáfora ou catáfora), o pronome de terceira pessoa do plural, por exemplo, pode se referir também a (Ilari et al., 1993): (i) um

³⁷ As sublinhas são nossas.

conjunto de indivíduos evocados por meio de um ‘*abstrato institucional*’ (*Prefeitura de São Paulo* = “*eles*”); (ii) referência a um gênero ou espécie (*baiano* = “*eles*”); (iii) remissão a um elemento do mesmo *frame*: *Japão* = “*eles*” = ‘*os japoneses*’).

Nos três casos, estabelece-se uma correferência indireta graças à presença de “pistas” referenciais no contexto. Essa concepção de referência mais ampla, apoiada no discurso, tem em Brown & Yule (1983: 215; apud. Micheletti, 2000) seus grandes expoentes. Estes autores consideram três tipos de constituintes antecedentes para se estabelecer um processo de referência no discurso: (i) *antecedentes nominais* (anáfora-textual prototípica); (ii) expressão *predicativa antecedente* (a referência se estabelece com termos que não necessariamente compartilham os mesmos traços de gênero e número); e *novos predicados* (quando não houver nenhum antecedente nominal ao pronome).

Koch & Marcuschi (1998) também diferenciam quatro processos de referenciação. Com base nos critérios de *correferência* e *co-significação* (que se pode entender como a manutenção dos traços de gênero, número e pessoa, etc.), os autores propõem que a referenciação pode ser explícita ou implícita. Esta última subdivide-se de acordo com a vinculação da referência: textual, contextual ou situacional.

Com base nos pressupostos teóricos de Ilari et al. (1993), Brown & Yule (1983; apud. Micheletti, 2000) e Koch & Marcuschi (1998), este trabalho adota uma concepção discursivo-pragmática para a referência do pronome *você/cê*. A análise qualitativa dos dados com relação à referência é feita no capítulo III.

Análise da referência na literatura sociolinguística

A discussão precedente sustenta a posição, tomada nesta pesquisa, de que uma oposição puramente binária da referência (“determinada” *versus* “indeterminada”) não dá conta de expressar a complexidade envolvida no processo de referenciação. Além disso, a anáfora e a catáfora prototípicas não são os únicos processos de correferência; esta pode se dar de maneira indireta,

recuperando-se antecedentes que não compartilham todos os traços e/ou por inferências pragmáticas.

Tomar a indeterminação da referência como um *continuum* foi a opção em trabalhos que tratam da variação no preenchimento do sujeito pelo pronome de terceira pessoa do plural “eles”, mas ainda tem sido muito pouco explorada em estudos variacionistas sobre o pronome de segunda pessoa.

Micheletti (2000), ao investigar o preenchimento ou não do sujeito pronominal de terceira pessoa do plural “eles”, distingue três graus de indeterminação desse pronome, conforme a presença ou ausência de “pistas” referenciais: Grau 1 = indeterminado com pistas referenciais no texto (elo de referência explícito); Grau 2 = indeterminado com pistas referenciais imersas no contexto (elo de referência implícito); Grau 3 = ausência de pistas (plenamente indeterminado).

Micheletti & Franchetti (1996) também adotam uma divisão ternária para a referência indeterminada³⁸. Souza (2007), por sua vez, retoma essa proposta e sugere um detalhamento maior desses graus de indeterminação, levando em conta a natureza das “pistas referenciais”. São estabelecidas, assim, três tipos de “pistas” (Souza, 2007: 64-67): *Genérica* (remete a uma classe ou espécie de indivíduos); *Coletiva* (refere-se a um grupo de indivíduos); e *Locativa* (apresenta-se um cenário definido). As “pistas” referenciais denominadas “genéricas” diferem das “coletivas”, segundo Souza (2007: 64), porque apenas as primeiras permitem uma pluralização sem qualquer alteração semântica.

Além desses três tipos de “pistas referenciais”, a autora analisa também os casos em que há uma referência indefinida, ou seja, quando não existe nenhuma “pista” referencial presente no contexto³⁹.

³⁸ Micheletti & Franchetti (1996) também seguem os pressupostos de Brown & Yule (1983), para tratar da indeterminação do sujeito, distinguindo três graus: (i) “pistas” referenciais presentes no texto; (ii) “pistas” imersas no contexto; (iii) ausência de “pistas”.

³⁹ A análise quantitativa empreendida por Souza (2007) revela que os tipos de constituintes antecedentes no discurso correlacionam-se ao emprego do pronome de terceira pessoa do plural: quando as “pistas” referenciais são de locativos e genéricos, há um favorecimento ao preenchimento do sujeito pelo pronome (pesos relativos de .84 e .81, respectivamente). Os SNs coletivos, por sua vez, favorecem bem menos o preenchimento do sujeito (.54), ao passo que a ausência de pistas referenciais desfavorece essa variante (.22).

Já entre os estudos sociolinguísticos sobre a variação no uso do pronome *você/cê*, é comum a investigação da referência como um grupo de fatores. Tais pesquisas, todavia, geralmente adotam uma oposição binária (referência determinada vs. referência indeterminada)⁴⁰. São exemplos os trabalhos de Ramos (1997, 2000), Andrade (2004), Peres (2006) e Gonçalves (2008)⁴¹.

Vital & Ramos (2006 [1999]) ampliam um pouco essa distinção para três empregos do pronome *você*: específico (14), genérico (15) e expletivo (16)⁴².

(14) De que lado **cê** tá?

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

(16) Em Kioto **você** tem aquela confusão nas ruas.

A ocorrência do pronome em (16) é considerada como desprovida de referência e de sentido pelos autores. Neste caso, o *você/cê* apenas ocuparia a posição sintática de sujeito – a exemplo do que ocorre em outras línguas, como o inglês, o francês e o italiano – em que os constituintes *there / it, il e ci* funcionam como expletivos (cf. Vital & Ramos, 2006: 87 [1999] e Avelar, 2006).

Duarte (2003) também defende que o PB apresenta usos expletivos do pronome pessoal *você* (assim como de outros pronomes, como *nós* e *a gente*) quando sujeitos em construções com os verbos *ter* e *ver* em sentido existencial – é o que se observa nos exemplos (2) e (3), retomados em (17) e (18) a seguir⁴³.

(17) a) **Você** não tem mais clientela no centro da cidade.

b) __ Não há/ __ não tem mais clientela no centro da cidade.

(18) a) **Você** vê muito concreto na tua frente.

b) __ Há / __ Tem muito concreto na tua frente.

⁴⁰ Também denominadas de referências “definida” versus “indefinida” ou “específica” versus “genérica” em determinados estudos.

⁴¹ O mesmo se verifica em estudos da variação entre os pronomes *você/cê* e *tu*: a referência é tratada como uma variável independente com apenas duas oposições (determinada versus indeterminada) – Lucca (2005), Dias (2007), Modesto (2006), Queiroz Andrade (2010).

⁴² Os exemplos (14-15) são de Ramos (1997). O exemplo (16) é de Vital & Ramos (2006:87 [1999]).

⁴³ Os exemplos são de Duarte (2003: 03).

Uma análise das ocorrências (supostamente) expletivas do *você/cê* é relevante na pesquisa que aqui se desenvolve. Se esse pronome estiver, de fato, funcionando como expletivo, esse uso pode corresponder a um novo estágio em seu processo de gramaticalização – Vitral & Ramos (2006 [1999]).

Os estudos que defendem esta hipótese, entretanto, tratam apenas do preenchimento da posição do sujeito. Duarte (2003), por exemplo, investiga o preenchimento ou não da posição à esquerda dos verbos “ter”, “ver” e “haver” (não apenas por pronominais, mas também por sintagmas locativos e temporais)⁴⁴. Callou & Avelar (2002; apud. Callou & Avelar, 2007) também pesquisam o preenchimento do sujeito por *você* em construções com o verbo “ter” em um *corpus* dos anos 1990⁴⁵. Em tais estudos, entretanto, não se investiga qual a pronúncia do pronome *você* (se plena ou reduzida) tende a ser empregada.

Vitral & Ramos (2006 [1999]), por sua vez, apresentam apenas uma análise qualitativa dos dados, embora destaquem a importância de se desenvolver um estudo que compute a frequência das variantes do pronome nos contextos existenciais:

Para se verificar então se o uso de *você* como expletivo constitui um problema para a suposição corrente na literatura [sobre processos de gramaticalização] de que haveria uma correspondência biunívoca entre perda de conteúdo semântico e perda de substância fonética, será preciso investigar, de forma comparativa, a frequência dos usos de *cê* e *você* nestes ambientes.

(Vitral & Ramos, 2006: 88 [1999])

No presente estudo, os empregos genéricos serão diferenciados dos supostos expletivos, a fim de verificar se a variante reduzida tenderia ou não a ser empregada nos contextos em que o pronome supostamente estaria desprovido de uma referência.

Cabe destacar, contudo, que a formação de sujeitos expletivos no PB está longe de ser consensual. Kaiser (2006) posiciona-se contrário a essa hipótese.

⁴⁴ Com relação ao verbo “haver”, Duarte (2003) verifica apenas o posicionamento de expressões locativas e temporais à sua esquerda. No caso do verbo “ver”, a posição do sujeito foi categoricamente preenchida nos dois corpora analisados (datados de 1980 e 2000). Com relação ao verbo “ter”, 25% das ocorrências do *corpus* mais recente apresentavam um constituinte à sua esquerda (Duarte, 2003: 06).

⁴⁵ Os autores verificam uma maior frequência de preenchimento do sujeito entre os informantes mais jovens.

Para o autor, o PB seria uma língua intermediária no que diz respeito ao parâmetro do sujeito nulo – distinguindo-se tanto do espanhol (língua cujo preenchimento do sujeito é mais restrito), quanto do francês (língua cujo preenchimento da posição de sujeito é obrigatório); e, conseqüentemente, não haveria a formação de expletivos no PB como decorrência dessa alegada mudança paramétrica. Duarte (2003: 131) também faz a ressalva de que nem sempre a distinção entre usos expletivos e genéricos é facilmente observável, “*podendo um sujeito expletivo [...] ser facilmente confundido com um referencial [...]*”.

Avelar (2006), que havia considerado o pronome *você* um expletivo (Callou & Avelar, 2001), retoma o assunto anos depois, argumentando que tal emprego corresponderia, na verdade, a um uso indeterminado/genérico: “*The version of ‘você’ appearing in existentials is an instance of indeterminate pronoun, with a generic interpretation.*” (Avelar, 2006: 02).

Para esse autor, que adota a perspectiva teórica do Programa Minimalista (Chomsky, 1995, 2000; apud. Avelar, 2006: 08), esse pronome não corresponde a um “expletivo verdadeiro”. De um ponto de vista sintático, *você/cê* teria sido movido de dentro de um sintagma preposicionado locativo para a posição de sujeito da sentença na estrutura de superfície⁴⁶:

There are evidences that *você* is not a true expletive in Brazilian Portuguese existential clauses. In the analysis I have proposed, the pronoun is initially merged into a locative PP and then moves to the specifier of TP/InflP.

(Avelar, 2006: 08)⁴⁷

No capítulo III, discutem-se alguns dados a este respeito, procurando analisar a referência ou a ausência de referência desse pronome nestes contextos.

Ao lado das oposições binárias ou ternárias da referência do *você/cê*, outras investigações tratam os empregos genéricos desse pronome em subtipos a depender de questões discursivas e pragmáticas – é o que se encontra nas análises de Corradello (1997) e Sousa (2008).

⁴⁶ Viotti (2000) trata da atribuição de Caso aos PPs locativos em sentenças existenciais do PB e de outras línguas, contrapondo alguns modelos teóricos dentro da Teoria Gerativa e do Programa Minimalista.

⁴⁷ TP = *temporal phrase* (sintagma temporal); InflP = *inflectional phrase* (sintagma flexional); PP = *prepositional phrase* (sintagma preposicionado).

Sousa (2008: 123-128) investiga as alterações na forma, na função e na referência desse pronome a partir de uma perspectiva histórica e funcionalista. Considerando um *corpus* dos anos 1990⁴⁸ do português falado em João Pessoa (Paraíba), a autora propõe três empregos para este item quanto à sua referência: com valor de segunda pessoa (P2)⁴⁹; com valor de primeira pessoa (P1); e genérico⁵⁰.

De acordo com Sousa (2008), os empregos em P1 e P2 se aproximam na medida em que *EU-TU* (i.e. a primeira e a segunda pessoas) são consideradas “pessoas” da enunciação (cf. Benveniste, 1991), ao passo que os usos genéricos estariam fora dessa instância – situando-se no campo de uma terceira pessoa, ou seja, uma “não-pessoa”, nos termos de Benveniste (1991). Para a autora, os traços [+ humano], [+ouvinte], [+presente no discurso] oporiam a 1ª e a 2ª pessoas, de um lado, à 3ª pessoa, de outro.

No entanto, embora se reconheçam distintos empregos “genéricos”, não faz parte dos objetivos da pesquisadora considerá-los como etapas sucessivas no processo de gramaticalização desse item nem desenvolver um estudo variacionista.

Os resultados indicam uma maior frequência dos empregos genéricos (independentemente da pronúncia do pronome), seguidos pelo “uso com valor de primeira pessoa-EU” e pelo uso específico. Ainda que a frequência *type* (Bybee, 2002, 2003) seja importante no processo de gramaticalização, a confiabilidade desses resultados fica prejudicada, pois pode ter ocorrido uma maior incidência de construções genéricas do que específicas apenas devido ao tipo de *corpus* adotado (entrevistas sociolinguísticas) e/ou ao assunto da entrevista.

⁴⁸ O *corpus* adotado por Sousa (2008) foi organizado pelo Projeto de Variação Linguística da Paraíba (VALPB).

⁴⁹ A denominação P1, P2, P3, P4, P5 e P6 é de Câmara Jr. (1980) e corresponde, respectivamente a: 1ª p.sg., 2ª p.sg., 3ª p.sg., 1ª p.pl., 2ª p.pl. e 3ª p.pl. Os exemplos são de Sousa (2008: 123, 125 e 128).

⁵⁰ Na verdade, o próprio emprego denominado de P1 por Sousa (2008) corresponde um subtipo de empregos genéricos, havendo entre eles apenas uma distinção em relação às pessoas envolvidas na referência do pronome.

De maneira similar, Corradello (1997: 07) analisa uma amostra de 18 inquéritos do NURC/SP⁵¹, propondo uma análise pragmático-discursiva⁵² para o pronome *você*. De acordo com a autora, quando indeterminado, o *você* pode fazer referência: (i) *ao locutor + interlocutor + “todo mundo”* ($1^a + 2^a + 3^a$ pessoas); (ii) *ao locutor + grupo específico de pessoas, menos o interlocutor* ($1^a - 2^a + 3^a$ pessoas); (iii) *ao próprio locutor* ($1^a - 2^a - 3^a$ pessoas); ou (d) *a uma 3ª pessoa* ($-1^a - 2^a + 3^a$ pessoas).

Corradello (1997) desenvolve um estudo que denomina “sociolinguístico distribucional”, computando apenas a frequência de uso do “você” genérico (independentemente de sua pronúncia plena ou reduzida) em relação às variáveis sociais faixa etária, sexo-gênero, situação de fala e grau de formalidade. Nesse estudo quantitativo, porém, não se distinguem os vários tipos de indeterminação descritos em sua análise qualitativa.

As mesmas críticas traçadas à análise de Sousa (2008) também se aplicam ao estudo de Corradello (1997), uma vez que a confiabilidade das análises quantitativas em ambas as pesquisas fica prejudicada, pois são analisados apenas os contextos em que o pronome indeterminado ocorre, sem computar os contextos em que ele poderia ter ocorrido (i.e., os outros contextos indeterminados). Adicionalmente, faz-se necessário considerar também medidas estatísticas mais precisas do que as frequências.

À parte as críticas acima, de um ponto de vista qualitativo, a análise de Corradello (1997) destaca-se por abranger um maior número de oposições da referência do pronome de segunda pessoa do singular, quando comparada aos outros estudos da referência desse pronome.

Em Nascimento (2008a), parte-se em busca de “expressões contextuais” que pudessem delimitar o campo referencial do *você* genérico. Propõem-se como critérios os traços [$\pm$ referencial], [$\pm$ definido], [$\pm$ exofórico] e [$\pm$ n° contável de

⁵¹ Os 18 inquéritos do NURC/SP analisados por Corradello (1997) compreendem 6 elocuições formais (EF), 6 diálogos entre dois informantes (D2) e 6 diálogos entre informante e documentador (DID). Na pesquisa diacrônica desenvolvida no presente estudo, 6 dentre os 18 inquéritos no formato DID da década de 1970 são os mesmos analisados por Corradello (1997).

⁵² A autora se fundamenta em pressupostos de (Possenti, 1988; apud. Corradello, 1997) e Maingueneau (1996; apud. Corradello, 1997).

indivíduos], a fim de estabelecer uma escala de (in)determinação para a referência do pronome *você/cê* (Nascimento, 2008a). Foram identificados cinco empregos desse pronome no que diz respeito à sua referência: específico; híbrido (ou ambíguo); genérico1 (com “pistas referenciais” explícitas no contexto discursivo); genérico2 (sem “pistas referenciais”); e expletivos.

Os cinco tipos de referência do pronome *você/cê* foram situados em uma escala crescente de genericidade (Nascimento, 2008a):

A hipótese que norteia essa classificação é a de que, se estiver correta a relação entre “perda de material fônico” e “perda de conteúdo semântico” (Heine & Reh, 1984; apud. Hopper & Traugott, 1993:87) ou “abstratização semântica”, pode-se esperar que a variante reduzida *cê* tenderia a ser empregada quanto mais indeterminada for a sua referência, ou quando a mesma estiver desprovida de uma referência (no caso dos usos expletivos).

Em Nascimento (2008a) analisa-se um *corpus* composto por 12 inquéritos (6 entrevistas sociolinguísticas e 6 aulas universitárias) gravados em 2003. Os resultados indicam que a variante reduzida tende a ser empregada com referência específica ao interlocutor e é desfavorecida nos empregos expletivos. Tais resultados, que vão de encontro à hipótese anterior, indicam uma não simultaneidade entre os processos de “abstratização semântica/perda de conteúdo semântico” e “redução fonética”. Os graus intermediários de indeterminação, por outro lado, mostraram-se neutros ao emprego das variantes em estudo. Desse modo, a presença ou não de “pistas” referenciais no contexto não se apresentou relevante ao emprego das variantes plena e reduzida do pronome.

O quadro a seguir traz uma síntese das classificações propostas para o estudo da referência do pronome *você* apresentadas até aqui:

Ramos (1997) e outros	Vitral & Ramos (2006 [1999])	Sousa (2008)	Corradello (1997)	Nascimento (2008a)
Definida Indefinida	Definida Indefinida Expletivo	P2 P1 Genérica	[2 ^a p.sg.] 1 ^a + 2 ^a + 3 ^a pess. 1 ^a + 3 ^a (– 2 ^a) pess. 1 ^a (– 2 ^a – 3 ^a) pess. – 1 ^a – 2 ^a + 3 ^a pess.	Específica Híbrida Genérico1 (com ‘pistas’) Genérico2 (sem ‘pistas’) Expletivo

Quadro 4: A referência do *você/cê* e seus subtipos

No capítulo III, propõe-se uma nova classificação para a referência do pronome *você/cê*, levando em consideração as pessoas envolvidas no processo de referenciação – a exemplo de Corradello (1997), mas com algumas modificações e inclusões.

2.4. O Princípio do Contorno Obrigatório

De acordo com o Princípio do Contorno Obrigatório (*Obligatory Contour Principle* – OCP), elementos idênticos não podem ocorrer adjacentes (McCarthy, 1986; apud. Suzuki, 1998). Esse princípio, inicialmente postulado como uma restrição categórica, teve versões mais brandas dentro da Teoria da Otimalidade (Prince & Smolensky, 1993; apud. Guy & Boberg, 1997)⁵³.

O Princípio do Contorno Obrigatório mostrou-se bastante fecundo à explicações de processos de dissimilação nos níveis segmental e suprasegmental. São exemplos canônicos a não adjacência de tons idênticos em línguas tonais (Leben, 1973; apud. Guy & Boberg, 1997) e a alteração de traços em segmentos adjacentes ou próximos – como no exemplo latino: /*sol-alis*/ = *sol-aris* (‘solar’) (Kent, 1936: 250; apud. Suzuki, 1998: 12).

Guy & Boberg (1997) tratam esse princípio como uma regra variável, que atua não apenas em termos categóricos, mas também probabilísticos. De acordo

⁵³ Suzuki (1998), por exemplo, chega a propor uma nova denominação para o princípio: *GOCP* (*Generalized OCP*), segundo o qual as restrições teriam um caráter universal, mas poderiam ser violadas, pois obedecem a um ranqueamento.

com os autores, uma vez que a variação é parte integrante do próprio sistema linguístico – como prevê a Teoria da Variação e da Mudança linguística – nada mais natural do que os mesmos princípios que governam as restrições categóricas em uma língua também governem questões variáveis:

(...) in a theory that postulates inherent variability within the grammar [the theory of Linguistic Variation and Change], it is perfectly plausible that some principles could govern both variable and invariant properties.

(Guy & Boberg, 1997: 151)

Os autores defendem que o OCP é um desses princípios, tomando como exemplo a sua aplicação à variação no uso dos segmentos /t/ e /d/ em final de palavra em inglês. O apagamento ou não desses segmentos é condicionado pelo contexto fônico antecedente, tanto de um ponto de vista categórico, quanto variável. Primeiramente, inexistente a sequência de duas oclusivas coronais idênticas /t, d/ na língua inglesa – o que demonstra a atuação categórica do OCP. Este princípio, por seu turno, também atua de maneira a favorecer o apagamentos de /t/ e /d/ quando o segmento antecedente compartilha um maior número de traços com esses fonemas – evidenciando um efeito probabilístico do *Princípio do Contorno Obrigatório* (Guy & Boberg, 1997).

As oclusivas coronais /-t, -d/ apresentam os traços fonológicos [- son], [- cont] e [+ cor]⁵⁴. Quando, no contexto precedente, há uma consoante oclusiva não-coronal [-son, - cont, - cor], uma sibilante [- son, + cont, + cor] ou a nasal coronal /n/ [+ son, - cont, + cor], o apagamento de /t, d/ é favorecido, uma vez que os segmentos adjacentes compartilham dois traços fonológicos. No caso das nasais não-coronais [+ son, - cont, - cor], das fricativas não sibilantes [- son, + cont, -cor] e das laterais [+ son, + cont, + cor], há uma menor tendência de /-t, -d/ serem apagados, uma vez que o segmento fônico precedente compartilha com o seguinte apenas um traço fonológico. O estudo de Guy & Boberg (1997), portanto, oferece um bom exemplo de como o *Princípio do Contorno Obrigatório* pode funcionar como uma regra variável.

⁵⁴ Onde: [son] = sonorante; [cont] = contínuo; e [cor] = coronal.

O *Princípio de Alternância Rítmica*, por sua vez, defendido por Selkirk (1984: 52; apud. Abousalh, 1997: 68) estabelece que, nas línguas naturais tende-se a evitar tanto a sequência de duas sílabas acentuadas (“choques de acento”), quanto sequências longas de sílabas átonas (“lapsos acentuais”):

- (i) Toda sílaba forte num nível métrico *n* deve ser seguida por, pelo menos, uma posição fraca naquele nível;
- (ii) Qualquer posição fraca num nível métrico *n* pode ser precedida por no máximo uma posição fraca naquele nível.

Uma das estratégias possíveis para se evitar o encontro de duas sílabas acentuadas é a “retração do acento”, como em⁵⁵:

<i>thirteen men</i>	→	<i>thirteen men</i>
(. *) (*)		(* .) (*)

Ambos os casos (choques acentuais e lapsos acentuais) podem ser considerados desdobramentos de *Princípio de Contorno Obrigatório* (Abousalh, 1997), pois se evitam sequências idênticas.

A exemplo do que propõem Guy & Boberg (1997), o *Princípio de Contorno obrigatório* é concebido aqui em termos probabilístico e serve para justificar três hipóteses, que são traduzidas em grupos de fatores no capítulo III.

2.5. Princípios discursivos e cognitivos

Como se abordou no item precedente, a referência do pronome, apesar de sua natureza semântica, tem uma forte relação com o discurso e com a pragmática. Além dessa, outras duas questões discursivas foram investigadas como hipóteses que podem estar correlacionadas ao emprego de *você* e *cê*: o paralelismo de formas e a frequência do verbo seguinte ao pronome. Além de estarem associadas ao discurso, ambas têm um caráter cognitivo.

⁵⁵ Cf. Collischonn (1996) e Abousalh (1997).

O “*Paralelismo*” ou “*princípio do processamento paralelo*” (Scherre, 1998), por exemplo, tem se correlacionado a diferentes fenômenos linguísticos (cf. Paiva & Scherre, 1999). Esse princípio pode ser definido como a manutenção de uma mesma forma em uma série; ou seja, trata-se da influência que formas antecedentes exercem sobre as subsequentes.

Destacam-se, a este respeito, os estudos realizados sobre a concordância de número no Português Brasileiro (Naro & Scherre, 1996; Scherre & Naro, 1991; Scherre, 1988, 1991, 1997 – apud. Scherre, 1998), cujos resultados indicam que a marca de concordância leva à sua manutenção nos contextos subsequentes, assim como a ausência de concordância leva também à sua ausência.

Conforme Scherre (1998), o fato de marcas já realizadas (portanto, previsíveis) tenderem a ser mantidas no contexto seguinte é considerado “*contra-funcional*” por alguns linguistas. Scherre (1998), por seu lado, defende o caráter funcional do Paralelismo. Anos antes, a mesma autora já havia considerado que este princípio não se trata de uma mera “*repetição mecânica de formas*”, mas de um processo mais complexo:

(...) a forma de atuar da variável Paralelismo formal mostra que os falantes são compelidos a usar formas semelhantes por algum princípio mental associativo, que pode estar ligado a uma das formas da mente humana operar, refletido no comportamento humano em geral.

(Scherre, 1988; apud. Scherre, 1998: 42)

Paiva & Scherre (1999) também defendem que a sistematicidade dessa variável, cuja atuação foi identificada em diversos níveis linguísticos e em várias comunidades de fala não só no PB, faz com que ela adquira o *status* de um “*universal de processamento linguístico*”.

Nos estudos sobre o pronome de segunda pessoa no PB, esse princípio também tem se mostrado atuante. Paredes Silva (2003), por exemplo, verifica a atuação do paralelismo na variação entre *você*, *cê* e zero no dialeto carioca. Andrade (2004), em um estudo desse pronome em Brasília, também observa que, uma vez pronunciado *você*, ou *cê*, o falante tende a continuar empregando as mesmas formas nas orações seguintes.

Outra questão importante apontada por Paiva & Scherre (1999) é a de que o Paralelismo pode afetar a atuação de outras variáveis. As autoras citam exemplos de estudos em que uma análise dos dados sem ocorrências no contexto anterior (isolados ou primeiros de uma série) evidenciaram a correlação de determinados grupos de fatores sociais. Tal fato demonstra que, além da questão discursivo-cognitiva, a análise desse princípio se faz importante também para um melhor controle das outras hipóteses analisadas na pesquisa.

Outro princípio de natureza discursiva e cognitiva diz respeito à frequência com que itens linguísticos são empregados pelos falantes. Como se discutiu em 2.2, a frequência tem sido uma questão de grande interesse em estudos que tratam de processos de gramaticalização (Bybee, 2002): o emprego elevado de um item, por exemplo, pode deflagrar ou acelerar seu processo de gramaticalização. Além disso, essa questão faz parte dos interesses das abordagens sobre a língua em uso (Pierrehumbert, 2001; Poplack, 2001 – apud. Hopper & Bybee, 2001). Nas palavras de Hopper & Bybee (2001: 03):

The frequency with which certain items and strings of items are used has a profound influence on the way language is broken up into chunks in memory storage, the way such chunks are related to other stored material and the ease with which they are accessed.

Segundo essa assunção, itens cujo emprego é mais frequente são mais facilmente “estocados” pelos falantes e seu acesso também é facilitado. Além disso, a alta frequência de emprego de um elemento também é considerada responsável pelo processo de “habituação” (Haiman, 1994; apud. Hopper & Bybee, 2001): o uso constante pode o levar a uma “perda/mudança” de certos traços – o que conduz ao *bleaching* envolvido nos processos de gramaticalização, como se discutiu em 2.2.

A Hipótese de Redução Probabilística (*Probabilistic Reduction Hypothesis*) prevê ainda que, quanto maior frequência de um constituinte, maior sua probabilidade de ser reduzido (cf. Jurafsky et al., 2001: 230). Jurafsky et al. (2001), por exemplo, verificam que, em palavras funcionais do inglês (*a, the, in, of, to, and, I, it* e *you*), processos de redução vocálica são afetados pela frequência

de palavras vizinhas; em itens lexicais, porém, o autor verifica que o apagamento de /-t, -d/ finais é influenciado apenas pela frequência da palavra que contém aqueles segmentos.

Ainda que os propósitos e a metodologia do estudo de Jurafsky et al. (2001: 230) e do que aqui se apresenta sejam distintos, nosso interesse também recai sobre o efeito da frequência das palavras do contexto seguinte ao pronome (no caso, o verbo) como um possível favorecedor à sua redução.

Corradello (1997) também destaca a frequência de determinadas expressões, como *você vê*, *você veja* e *você pode ver* em um *corpus* dos anos 1970. Conforme a autora, o verbo “ver” possuiria uma significação hipotética, e esse emprego do pronome apresenta referência específica ao interlocutor, mas desempenha ao mesmo tempo a função discursiva de um “marcador conversacional”:

[...] as expressões **você vê**, **você sabe** determinam claramente a referência ao interlocutor. Sua função no discurso, entretanto, parece ser melhor explicada pelos estudos da Análise da Conversação, por sugerir tratar-se de um “marcador conversacional” que, nas palavras de Marcuschi: “*são sinais característicos da linguagem falada, muitas vezes desprovidos de conteúdo semântico, que servem para organizar a conversação, interligando os vários segmentos do discurso e engajando os interlocutores no diálogo.*”⁵⁶

No presente estudo, considera-se também que o emprego de “(vo)cê + V”, em construções como as exemplificadas acima, poderia estar passando por um processo de automatização (Bybee, 2002), segundo o qual a repetição de itens, em princípio independentes, levam-nos a serem processados e “estocados” na mente dos indivíduos como uma única unidade. No próximo capítulo, retoma-se essa discussão, ao se definir o envelope de variação desta pesquisa.

⁵⁶ Grifos nossos.

CAPÍTULO III

Análise Qualitativa

Afinal, se uma língua tem de ser estruturada, a fim de funcionar eficientemente, como é que as pessoas continuam a falar enquanto a língua muda, isto é, enquanto passa por períodos de menor sistematicidade?

(Weinreich, Lavov & Herzog, 2006 [1968])¹

Introdução

Neste capítulo é definido o envelope de variação e são estabelecidos os grupos de fatores acompanhados de suas justificativas linguísticas e sociais.

3.1. As amostras de dados

O *corpus* composto para esta pesquisa contém duas amostras do português paulistano falado, separadas por um intervalo de aproximadamente trinta anos.

A primeira das amostras é constituída de 36 entrevistas sociolinguísticas gravadas entre 2003 e 2009. Estas entrevistas integram um *corpus* maior, que está sendo organizado pelo Grupo de Estudos e Pesquisa em Sociolinguística (GESOL) do Departamento de Linguística da Universidade de São Paulo. Esta amostra foi denominada aqui *GESOL-SP-2000*.

A segunda amostra compõe-se de 18 inquéritos no formato de diálogo entre informante e documentador (DID). Os informantes são paulistanos com nível superior de escolaridade e as entrevistas foram gravadas nos anos 1970 por pesquisadores do Projeto Norma Urbana Culta (NURC) do estado de São Paulo. Neste estudo, esta amostra é denominada *NURC-SP-1970*.

¹ Trad.: Marcos Bagno.

O *corpus* de pesquisa do GESOL é composto de entrevistas sociolinguísticas, aulas universitárias e redações escolares – aqui, foi utilizada parte das primeiras. Esse material, em sua maior parte, foi coletado e transcrito por alunos de Sociolinguística, coordenados pelo Prof. Dr. Ronald Beline Mendes.

As gravações das entrevistas se deram entre 2003 e 2009 e foram coletadas por documentadores (homens e mulheres), geralmente na faixa etária dos 20 aos 30 anos de idade, com grau superior de escolaridade. Foi fornecido aos documentadores um roteiro prévio para a entrevista, o que proporcionou uma uniformidade do *corpus*. Os temas abordados foram “bairro”², “família”, “transporte”, “saúde”, “política” e “educação”³. Este último tema deveria ser abordado necessariamente na seção final da entrevista. As gravações foram realizadas com o consentimento prévio dos informantes. Foi solicitado aos documentadores, contudo, que não dissessem aos informantes que se tratava de um estudo linguístico, de língua portuguesa, ou algo relacionado à linguagem, a fim de se chegar o mais próximo possível do vernáculo do informante, contornando, dentro do possível, o chamado *paradoxo do observador* (Labov, 2006 [1966])⁴.

Em virtude do reduzido número de dados com referência específica ao interlocutor e do fato de que as entrevistas foram realizadas com distintos documentadores, desenvolveu-se um estudo à parte com os dados de alguns documentadores. Metodologicamente, não é possível analisar a fala de todos os documentadores, uma vez que: (i) alguns deles não são paulistanos; (ii) não foram recolhidos os dados pessoais de todos os documentadores, nem puderam ser todos recontactados; e (iii) sobretudo em entrevistas cujo informante é da terceira faixa

² O “bairro” é um tema mais recente que integra as entrevistas de 2009. Os demais temas permeiam todos os anos.

³ Em alguns casos, incluíram-se também testes de percepção, em que o documentador questionava o informante a respeito de algumas estruturas linguísticas (pronúncia do /r/ em coda silábica; emprego de *nós* e *a gente*; e uso do gerundismo – *vou fazer* versus *vou estar fazendo*). Além disso, algumas entrevistas trazem uma seção final com leitura de palavras e um trecho de um texto, seguindo a proposta de entrevista sociolinguística de Labov (2006 [1966]). A leitura de textos, a leitura de palavras e os testes de percepção tinham como alvo outros casos de variação. Além disso, nem todas as entrevistas adotadas na presente amostra apresentam essa configuração. Por esses motivos, quando fosse o caso, tais trechos das entrevistas foram desconsiderados na análise.

⁴ Labov (2006 [1966]) denomina de “paradoxo do observador” o fato de que o pesquisador precisa observar dados de fala enunciados como se o falante não estivesse sendo observado.

etária, o documentar às vezes refere-se a ele/ela exclusivamente pela forma de tratamento *o senhor/ a senhora*, que não são objeto da presente pesquisa. Dentre as 36 entrevistas sociolinguísticas dos anos 2000, apenas 12 documentadores são paulistanos. Dentre estes, dois empregaram exclusivamente a forma de tratamento “*o/a senhor(a)*” e uma entrevista apresentava problemas de áudio na fala da documentadora.

A análise dos dados dos documentadores foi, portanto, reduzida a um conjunto de 9 indivíduos e atendeu a dois objetivos: analisar as variáveis independentes “referência do pronome” e “tipo de oração” com um número mais robusto de dados de referência específica e em orações interrogativas.

Para preservar a identidade dos informantes e documentadores, os nomes apresentados são fictícios. Além disso, os exemplos não fornecem informações pessoais acerca deles.

Os quadros a seguir trazem os perfis dos 9 documentadores considerados na análise dos dados e dos 36 informantes que compõem a amostra GESOL-SP-2000.

DOCUMENTADORES PAULISTANOS AMOSTRA GESOL-SP-2000						
Informante	Sexo	Sexo D-I	Idade	Escolar.	Ano	Código ⁵
01-Rosane	F	F-F	21	Superior	2008	Rosane-F1SD-2008
02-Júlia	F	F-F	21	Superior	2008	Júlia-F1SD-2008
03-Alice	F	F-M	23	Superior	2006	Alice-F1SD-2006
04-Karina	F	F-M	27	Superior	2007	Karina-F1SD-2007
05-Ellen	F	F-F	27	Superior	2007	Ellen-F1SD-2007
06-Ivana	F	F-F	28	Superior	2009	Ivana-F1SD-2009
07- Camila	F	F-M	32	Superior	2007	Camila-F1SD-2007
08-João	M	M-F	37	Superior	2007	João-M2SD-2007
09-Dino	M	M-F	53	Superior	2007	Dino-M3SD-2007

Quadro 5: Amostra GESOL-SP-2000 (documentadores paulistanos)

⁵ As siglas adotadas após a pseudônimo do informante indicam o sexo/gênero, a faixa etária, o nível de escolaridade e seu papel na entrevista sociolinguística: F = feminino; M = masculino; 1 = primeira faixa etária; 2 = segunda faixa etária; 3 = terceira faixa etária; S = nível superior de escolaridade; F = ensino Fundamental I ou II; M = ensino Médio; I = informante; D = documentador.

INFORMANTES DA AMOSTRA GESOL-SP-2000						
Informante	Sexo	Sexo D-I	Idade	Escolar.	Ano	Código
01-Fabíola	F	M-F	22	Superior	2007	Fabíola-F1SI-2007
02-Fabiana	F	F-F	32	Superior	2003	Fabiana-F1SI- 2003
03- Camila	F	F-F	22	Superior	2006	Camila-F1SI-2006
04-Ingrid	F	F-F	35	Superior	2007	Ingrid-F2SI-2007
05-Vanessa	F	M-F	45	Superior	2003	Vanessa-F2SI-2003
06- Cecília	F	M-F	38	Superior	2007	Cecília-F2SI-2007
07- Silvia	F	M-F	52	Superior	2006	Silvia-F3SI-2006
08- Denise	F	F-F	72	Superior	2009	Denise-F3SI-2009
09-Zoraide	F	F-F	53	Superior	2003	Zoraide-F3SI-2003
10- Pedro	M	M-M	20	Superior	2007	Pedro-M1SI-2007
11-Mauro	M	F-M	31	Superior	2003	Mauro-M1SI-2003
12- Fabrício	M	F-M	24	Superior	2007	Fabrício-M1SI-2007
13- Rogério	M	F-M	36	Superior	2007	Rogério-M2SI-2007
14-Flávio	M	M-M	35-45	Superior	2003	Flávio-M2SI-2003
15- Edilson	M	M-M	37	Superior	2007	Edilson-M2SI-2007
16-José	M	M-M	63	Superior	2006	José-M3SI-2006
17-Frederico	M	F-M	50+	Superior	2003	Frederico-M3SI-2003
18-Sebastião	M	M-M	50+	Superior	2009	Sebastião-M3SI-2009
19-Jéssica	F	F-F	21	Fundam.	2007	Jéssica-F1FI-2007
20-Cláudia	F	F-F	20-25	Fundam.	2007	Cláudia-F1FI-2007
21-Aline	F	M-F	20-32	Fundam.	2004	Aline-F1FI-2004
22-Edith	F	F-F	36	Fundam.	2006	Edith-F2FI-2006
23-Marta	F	F-F	36	Fundam.	2008	Marta-F2FI-2008
24-Gisele	F	F-F	45	Fundam.	2008	Gisele-F2FI-2008
25-Clara	F	F-F	58	Fundam.	2007	Clara-F3FI-2007
26-Maria	F	F-F	52	Fundam.	2008	Maria-F3FI-2008
27-Rosana	F	F-F	65	Fundam.	2009	Rosana-F3FI-2009
28-Fábio	M	F-M	25	Fundam.	2007	Fábio-M1FI-2007
29-Bruno	M	M-M	30	Fundam.	2008	Bruno-M1FI-2008
30-Gustavo	M	F-M	27	Fundam.	2006	Gustavo-M1FI-2006
31- Jefferson	M	F-M	35	Fundam.	2008	Jefferson-M2FI-2008
32-Luís	M	M-M	43	Fundam.	2007	Luís-M2FI-2007
33- Daniel	M	F-M	44	Fundam.	2008	Daniel-M2FI-2008
34- Marco	M	M-M	57	Fundam.	2007	Marco-M3FI-2007
35- Ulisses	M	M-M	66	Fundam.	2006	Ulisses-M3FI-2006
36- Valter	M	M-M	51	Fundam.	2007	Valter-M3FI-2007

Quadro 6: Amostra GESOL-SP-2000 (Informantes)

O *corpus* do NURC-SP é composto por inquéritos em três formatos: Diálogo entre informante e documentador (DID); diálogo entre dois informantes (D2) e elocuções formais (EF)⁶. Uma seleção desse *corpus* encontra-se publicada em três volumes (Castilho & Preti, 1986-1990). Dos 18 inquéritos utilizados neste estudo, 9 foram publicados. No entanto, apenas suas transcrições (ortográficas) encontram-se publicadas e estas não visam a distinguir a pronúncia do pronome em análise (plena ou reduzida), assim como não há a distinção de outras questões relevantes para esta pesquisa (como a pronúncia do elemento seguinte, por exemplo: *está/tá*, *não/num*, etc.). Fez-se necessário, portanto, ouvir as 18 gravações. Uma primeira versão das transcrições dos 9 inquéritos restantes foi realizada pelos alunos do curso de Sociolinguística, que utilizaram esse material nas pesquisas realizadas durante o curso. Fez-se uma revisão de todas as transcrições (publicadas ou não), a fim de distinguir se a pronúncia do pronome era foneticamente plena ou reduzida e checar o contexto em que este ocorre.

As gravações do *NURC-SP* foram realizadas durante a década de 1970 por linguistas pesquisadores filiados a esse projeto, cujo objetivo era compor um *corpus* do português culto falado em cinco capitais brasileiras com mais de um milhão de habitantes na época: São Paulo, Rio de Janeiro, Recife, Salvador e Porto Alegre. O Projeto NURC em São Paulo é coordenado pelo Prof. Dr. Dino Preti (USP) e pelo Prof. Dr. Ataliba Teixeira de Castilho (UNICAMP; USP) e conta atualmente com duas sedes, uma na USP (sob responsabilidade da Profa. Dra. Marli Quadros Leite) e outra no Instituto de Estudos da Linguagem (IEL), na UNICAMP⁷. A pesquisa do material necessário deu-se na sede localizada na USP⁸.

Os assuntos das entrevistas variavam entre: “casamento, família”; “casa”; “profissões”; “vestuário”; “tempo”; “finanças”; “futebol, esportes” e “educação”. Na maioria das entrevistas, depreende-se um senso de formalidade por parte tanto dos documentadores (duplas formadas por duas mulheres ou por um homem e

⁶ Aulas universitárias.

⁷ Para maiores informações: <http://www.flech.usp.br/dlc/nurc/historico.htm>

⁸ Agradeço à Profa. Dra. Marli Quadros Leite, que gentilmente permitiu o acesso às gravações das entrevistas do NURC-SP.

uma mulher) quanto por parte dos informantes. Infelizmente, não fez parte dos objetivos dos empreendedores do projeto a gravação de informantes com níveis de escolaridade mais baixa. Como não há entrevistas com esse perfil em outros *corpora* da cidade de São Paulo coletados na mesma época, o estudo diacrônico foi feito com uma subamostra de *GESOL-SP-2000*, de modo que sua comparação com o *NURC-SP-1970* pudesse ser quantitativamente adequada.

INFORMANTES AMOSTRA NURC-SP-1970						
Informante	Nº Inq.	Sexo	Idade	Escolar.	Ano	Código
01-Valéria	DID-116	F	28	Superior	1972	Valéria-F1S-NURC-1972
02-Luzia	DID-17	F	25	Superior	1971	Luzia-F1S-NURC-1971
03-Márcia	DID-76	F	25	Superior	1972	Márcia-F1S-NURC-1972
04-Felícia	DID-234	F	44	Superior	1974	Felícia-F2S-NURC-1974
05-Ana	DID-228	F	38	Superior	1974	Ana-F2S-NURC-1974
06-Fátima	DID-251	F	34	Superior	1974	Fátima-F2S-NURC-1974
07-Francine	DID-242	F	60	Superior	1974	Francine-F3S-NURC-1974
08-Margarida	DID-212	F	56	Superior	1974	Margarida-F3S-NURC-1974
09-Carminda	DID-280	F	72	Superior	1976	Carminda-F3S-NURC-1976
10-Mathias	DID-161	M	25	Superior	1974	Mathias-M1S-NURC-1974
11-Leonel	DID-46	M	25	Superior	1971	Leonel-M1S-NURC-1971
12-Marcos	DID-28	M	29	Superior	1972	Marcos-M1S-NURC-1972
13-Marcelo	DID-137	M	41	Superior	1972	Marcelo-M2S-NURC-1972
14-Irineu	DID-79	M	38	Superior	1972	Irineu-M2S-NURC-1972
15-Assis	DID-172	M	37	Superior	1974	Assis-M2S-NURC-1974
16-Airton	DID-250	M	69	Superior	1974	Airton-M3S-NURC-1974
17- Leonardo	DID-71	M	67	Superior	1972	Leonardo-M3S-NURC-1972
18- Gustavo	DID-134	M	60	Superior	1972	Gustavo-M3S-NURC-1972

Quadro 7: Amostra NURC-SP-1970 (DID = diálogo entre informante e documentador)

3.2. Variantes e o envelope de variação

Além das variantes *você* e *cê*, foram encontradas as formas *ocê* e *ucê* nos *corpora* analisados. No entanto, o número de ocorrências destas foi tão reduzido (apenas 06 ocorrências de *ocê* e 03 de *ucê* em mais de 4000 dados) que se optou por excluí-las da análise quantitativa.

Foram codificadas separadamente as ocorrências cuja pronúncia (ou a gravação) não permitiu uma audição plena da variante empregada. Em alguns casos, houve dúvida se o falante pronunciou [vo.'se] ou ['vse], com a elisão do [o]. Em outros casos, a dúvida era sobre se o falante teria pronunciado rapidamente a forma reduzida *cê*, empregado o indeterminador *se*, pronunciado [sɪ], ou um “zero”. Estes casos totalizam 117 ocorrências (54 de *vcê* e 63 de *cê* “duvidosos”) em um total geral de 4639 ocorrências, e também foram excluídos. A tabela a seguir expõe a distribuição geral das variantes nos dois corpora e na fala dos documentadores da amostra GESOL-SP-2000.

VARIANTES	Informantes (36) GESOL-SP-2000		Documentadores (9) GESOL-SP-2000		Informantes (18) NURC-SP-1970		TOTAL	
	Nº	%	Nº	%	Nº	%	Nº	%
VOCÊ	1616	52,7	639	54,9	263	64,1	2518	54,30
CÊ	1367	44,6	488	41,9	140	34,1	1995	43,00
[vse] (?)	<u>37</u>	1,2	<u>15</u>	1,3	<u>02</u>	0,5	<u>54</u>	1,16
CÊ (?)	<u>38</u>	1,2	<u>21</u>	1,8	<u>04</u>	1,0	<u>63</u>	1,35
OCÊ	05	0,2	-	-	01	0,2	06	0,13
UCÊ	02	0,1	01	0,1	-	-	03	0,06
Total	3065	100	1164	100	410	100	4639	100

Tabela 6: Variantes empregadas nas amostras do NURC-SP-1970 e GESOL-SP-2000

Uma variante deve ser contabilizada em relação ao total dos contextos em que ela ocorre ou poderia ter ocorrido. Pode-se questionar, então, se os contextos em que há um sujeito nulo deveriam ser incluídos na análise:

- (19) eu acho que assim são dois... são dois duas coisas importantes né? ... **Você** dá o acesso e **Ø mante(r)** a qualidade ou **Ø briga(r)** pela qualidade né? (Fabiana-F1SI-2003)

Neste caso, as duas ocorrências de sujeito nulo apresentam a mesma referência do pronome “você”, que as antecede. Disso resulta que sujeitos nulos poderiam constituir uma terceira variante a ser estudada. Empregos como esses, no entanto, não foram analisados neste estudo, uma vez que eles requerem a

análise de outra variável dependente: o preenchimento ou não do sujeito de segunda pessoa do singular – que se encontra fora do escopo desta pesquisa.

Analisa-se, assim, somente as ocorrências do pronome efetivamente pronunciadas (exceto as pronúncias duvidosas e as formas pouco frequentes *ocê* e *ucê*).

Variável dependente: <i>pronomes de segunda pessoa (singular ou plural)</i>		
Variantes no PB	Variantes na cidade de SP	Variantes investigadas
<i>Você(s)</i>	<i>Você(s)</i>	<i>Você(s)</i> <i>Cê(s)</i>
<i>Cê(s)</i>	<i>Cê(s)</i>	
<i>Ocê(s)</i>	<i>Ocê(s)</i>	
<i>Ucê(s)</i>	<i>Ucê(s)</i>	
<i>Vcê(s)?</i>	<i>Vcê(s) ?</i>	
<i>Tu (sg.)</i>	-	
<i>Zero</i>	<i>Zero</i>	

Quadro 8: Variável e variantes do pronome de segunda pessoa

Foram computadas as ocorrências desse pronome tanto nos contextos cuja referência é específica ao interlocutor, quanto naqueles em que o pronome apresenta uma referência genérica/indeterminada. Nesse último caso, é possível questionar se outros recursos indeterminadores poderiam ser empregados – tais como *nós*, *as pessoas*, *se*, etc. Uma análise dos recursos indeterminadores, todavia, também não constituiu foco de interesse da presente pesquisa.

Como foi discutido nos pressupostos teóricos (v. capítulo I), um requisito fundamental para que duas formas possam ser consideradas variantes de uma mesma variável é a possibilidade de uma forma substituir a outra no mesmo contexto, preservando-se o mesmo significado. Embora *cê* possa ser sempre comutada por *você*, não são todos os contextos em que a variante reduzida pode ser empregada em lugar da forma pronominal foneticamente plena.

Com base em dados intuitivos, Vitral (1996) apresenta uma série de contextos nos quais o uso de *cê* seria agramatical, i.e. em que *você* teria um emprego categórico. De acordo o autor, a variante reduzida *cê* restringe-se aos

contextos em que o pronome desempenha a função sintática de sujeito (20), exceto se posposto ao verbo (24) ou modificado por advérbios (26). Quando o pronome é complemento de preposição (22), o emprego de *cê* seria vetado, pois este ambiente sintático não permite a ocorrência de clíticos. A forma *cê* também seria agramatical com a função sintática de objeto direto (21) e em de construção de tópico (25). Estas duas últimas restrições justificam-se, conforme o autor, dado a ênclise ter se tornado um fenômeno raro no PB falado atualmente.

SUJEITO	(20) a) (V)ocê pediu pra sair. b) Cê pediu pra sair.
OBJETO DIRETO	(21) a) Eu amo (v)ocê . b) Eu amo * cê .
COMPLEMENTO DE PREPOSIÇÃO	(22) a) E falei pra você . b) Eu falei pr'ocê . c) Eu falei * pra cê .
PÓS-VERBAL	(23) a) (V)ocê foi o culpado. b) Cê foi o culpado. (24) a) Foi (v)ocê o culpado. b) Foi * cê o culpado.
PREPOSTO [TÓPICO]	(25) a) (V)ocê , ele não viu. b) * Cê , ele não viu.
MODIFICADO POR ADVÉRBIO	(26) a) Só (v)ocê tava mentindo. b) Só * cê tava mentindo.
ISOLADO	(27) A: – Quem vai sair? B: – (V)ocê . / – * Cê .
COORDENAÇÃO	(28) a) Ele e (v)ocê podem votar contra. b) * Ele e cê podem votar contra.

Quadro 9: Restrições sintáticas ao emprego de *CÊ* (exemplos de Vitral, 1996)

Tais restrições sintáticas ao emprego da variante reduzida *cê* levam Vitral (1996) a considerá-la um clítico. Uma vez que clíticos são itens que não possuem uma autonomia, eles precisam se “apoiar” em outro constituinte a eles adjacente (cf. Hopper & Traugott, 1993), normalmente o verbo – disso resulta que clíticos não são empregados em enunciados isolados, como em resposta a uma pergunta (27). Seu caráter átono, também os impede de serem topicalizados, integrarem uma estrutura coordenada (28), ou se encontrarem sob o escopo de um advérbio (26) – contextos nos quais receberiam um acento de palavra.

Algumas objeções podem ser feitas à análise de Vitral (1996). Em primeiro lugar, o contexto sintático objeto do verbo pode inibir o emprego de *cê* não devido à posposição do pronome ao verbo, como propõe autor, mas devido à sua função sintática. Embora a ênclise seja um fenômeno raro no PB atual, isso é válido para os clíticos acusativos – que parece não ser o caso de *cê*. Para Ramos (1997), *cê* vem se especializando com Caso nominativo no PB, dado a função sintática de sujeito ser a única que não oferece restrições ao seu uso. Os exemplos a seguir, entretanto, demonstram que a variante reduzida do pronome também pode ocorrer com Caso acusativo, quando seguida de orações reduzidas de particípio ou de gerúndio:

(29) Eu vi **cê caído** na rua.

(30) Vi **cê saindo** da festa ontem.

Em (29) e (30), *cê* está integrando uma *Small Clause*, na qual funciona como sujeito (Vital, 1996). A questão fundamental, portanto, parece ser a função sintática que o pronome exerce (sujeito), e não o Caso nominativo, propriamente dito. Nas sentenças anteriores, por exemplo, *cê* recebe Caso acusativo, atribuído pelo verbo da oração principal através de um processo de *Marcação Excepcional de Caso* (Miotto, 2004: 192).

Outro fato a ser considerado é a dupla função desempenhada pelo pronome em determinados contextos. Nos exemplos a seguir, o pronome é simultaneamente complemento da preposição (ou objeto do verbo da oração principal) e sujeito da oração encaixada:

(31) “Não, *cê* vai compra[r], *cê* vai com a bicicleta eu **deixo *cê i(r)***...” (Marco-M3FI-2007)

(32) Não dá pra conversar tanto, né? (nesses lugares)... *cê* conhece pessoas...é uma forma **de *cê mante(r)*** né? seu/... a vida social... (Fabiola-F1SI-2007)

(33) As pessoas fazem campanha na televisão **p’ *cê vota(r)***, mas vota(r) pra quê? (Jefferson-M2FI-2008)

Em (31), *cê* é sujeito de uma oração encaixada, que funciona como “objeto direto” da oração principal. Nos exemplos (32-33), embora o pronome venha

precedido por uma preposição, dentro da oração encaixada em que se encontra, *cê* desempenha a função sintática de sujeito dessa oração – daí serem denominados de “dupla função” sintática (cf. denominação de Andrade, 2004). Tais contextos mostraram-se variáveis no presente estudo.

Quando complemento de preposição sem exercer a função sintática de sujeito, como em (34) a seguir, *cê* foi empregado com a contração da preposição “para/pra”, reduzida apenas ao segmento fônico [p] – de maneira à preposição e o pronome formarem uma única sílaba “p’cê” [‘pse]. Casos como estes vão contra a hipótese de Vitral (1996), para quem esses contextos impediriam o emprego da variante reduzida; e funcionam como argumento para Petersen (2008) advogar em favor da classificação de *cê* como uma forma fraca.

- (34) Ah, difícil eu fala **p’cê** ... que nem, depois eu tô aqui hoje na casa da minha mãe. (Bruno-M1F-2008)

Seguindo a hipótese de Kato (1999; apud. Andrade, 2004: 52) – para quem posições marcadas e proeminentes na sentença são ocupadas apenas por pronomes fortes (e não por formas fracas ou clíticos⁹) – nos dados apresentados por Andrade (2004), a variante reduzida *cê* não ocorre em estruturas topicalizadas, clivadas ou em uma coordenação¹⁰. Nestes casos, a forma plena *você* foi categórica. Como Andrade (2004) lembra, todavia, o fato de um item não figurar em alguns *corpora* pode ser simplesmente uma questão de amostragem e não uma restrição categórica. Com relação às orações clivadas, Braga (1996; apud. Andrade, 2004: 52-53) distingue seis subtipos, apresentados a seguir¹¹:

⁹ Andrade (2004), seguindo a classificação entre Pronomes Fortes, Formas Fracas e Clíticos (Cardinaletti & Starke, 1994; apud. Andrade, 2004) considera *cê* uma forma fraca, que estaria rumo a um processo de cliticização. Neste estudo, investiga-se quão avançado estaria o processo de cliticização desse pronome.

¹⁰ A respeito das construções coordenadas, por exemplo, a ordem do pronome na estrutura parece ser relevante. É plausível uma sentença com a variante reduzida *cê* desde que a mesma seja o primeiro elemento da coordenação, embora soe “estranha” a mesma estrutura quando o pronome reduzido é o segundo elemento coordenado.

¹¹ Os exemplos foram extraídos de Andrade (2004) ou de nosso próprio *corpus* de pesquisa.

1. Clivadas propriamente ditas	(35)	Sua mãe me ligou na casa, ou foi <u>você</u> que quis ouvir? (In.: Andrade, 2004: 53)
	(36)	É <u>você</u> que escolhe isso? (Pedro-M1S-2007)
2. Construções É-Que	(37)	<u>Você</u> é que tinham que cuidar. (In.: Andrade, 2004: 53)
	(38)	“Mas parece que <u>você</u> é que é da Bahia pelo jeito (de) você fala(r):: algumas coisa(s)...” (Cláudia-F1F-2007) ¹²
3. Pseudo-clivada	(39)	Quem vai mostra(r) ao aluno como faze(r) as conexões é <u>você</u> . (Fabiana-F1S-2003)
4. Foco SER	(40)	Já vi que <u>você</u>s são é danados. (In.: Andrade, 2004: 53)
5. Que-Foco	(41)	<u>Você</u>s que levaram. (In.: Andrade, 2004: 53)
	(42)	“Nossa! Olha, <u>você</u> que escreveu? melhorou bem! ” (Silvia-F3S-2006)
6. Duplo foco	(43)	É <u>você</u> é que sofre. ¹³

Quadro 10: Restrições ao emprego de *CÊ* em orações clivadas

Dos seis contextos acima, apenas as clivadas propriamente ditas e as construções “É que”, “Que Foco” e “Foco Ser” ocorrem no *corpus* analisado por Andrade (2004), sempre com a variante plena do pronome. No estudo que aqui se apresenta, apenas as clivadas “Que-Foco” demonstraram variação:

(44) Ela falou “**cê que** é aquela mãezinha que brigou com a médica?” (Jéssica-F1SI-2007)

Embora haja forte restrição ao uso de *cê* em determinadas estruturas sintáticas¹⁴, faz-se necessário um levantamento empírico no *corpus* de estudo, a fim de verificar se os supostos contextos categóricos o são de fato na comunidade de fala analisada e, assim, definir o envelope de variação deste estudo. A tabela 7,

¹² As aspas indicam que se trata de um discurso reportando.

¹³ Adaptado de “É a gente é que sofre” (Braga, 1991; apud. Andrade, 2004: 53). Não foram obtidos dados de *você/ocê/cê* com essa estrutura sintática nem em Andrade (2004) nem nesta pesquisa.

¹⁴ Andrade (2004), ao analisar a função sintática do pronome (sujeito, objeto de verbo, complemento de preposição e dupla função), verifica que a forma *cê* ocorre somente com a função sintática de sujeito no *corpus* por ela analisado. Nos demais contextos (*você*) foi categórico. O mesmo padrão é depreendido por Ramos (1997) e Gonçalves (2008). Peres (2006) e Coelho (1999; apud. Gonçalves, 2008), por outro lado, computam raras ocorrências de *cê* com a função sintática de objeto de verbo e complemento de preposição. Esses resultados demonstram que, embora sejam ambientes bastante desfavorecedores ao emprego da forma reduzida do pronome, não se trata de contextos categóricos *a priori*.

na página seguinte, traz os percentuais de emprego das variantes com relação à sua função sintática e à estrutura sintática em que o pronome se encontra¹⁵.

Quando o pronome exerce uma dupla função sintática (objeto e sujeito simultaneamente), o uso de *cê* foi reduzido (26 ocorrências em 111 dados). O emprego dessa variante foi ainda mais infrequente (02/182) quando é apenas complemento de uma preposição (e.g., *eu disse p'cê*), o que indica uma forte restrição ao uso dessa variante nesse contexto. Quando objeto direto, em coordenação ou sujeito posposto ao verbo, observou-se ocorrência categórica da variante plena.

Nos dados dos documentadores, quando o pronome-sujeito integrava construções estabelecidas previamente à entrevista sociolinguística (um conjunto de perguntas comuns aos documentadores), ele foi computado separadamente das demais ocorrências do pronome com essa função sintática. Nas construções “pré-estabelecidas”, o uso de *cê* foi muito pouco frequente (10%). Além da distribuição quantitativa das variantes ser distinta nesses casos, trata-se de construções que se repetem nas entrevistas – razão pela qual foram excluídas do envelope de variação.

¹⁵ Todas as ocorrências do pronome foram computadas, inclusive truncamentos e repetições.

CÊ								
	FUNÇÃO / ESTRUTURA SINTÁTICA	INFs GESOL Anos 2000		DOCs GESOL Anos 2000		INFs NURC Anos 1970		CÊ Total Nº
		Nº	%	Nº	%	Nº	%	
FUNÇÃO SINTÁTICA	Sujeito simples	1185/2407	49,2	476/968	49,2	126/331	38,1	1787
	Sujeito em construções pré-estabelecidas	---	---	06/60	10,0	---	---	6
	Sujeito / objeto em coordenação	0/1	0	0/6	0	-	-	0
	Sujeito posposto ao V	0/2	0	-	-	0/1	0	0
	Objeto direto	0/9	0	0/4	0	0/4	0	0
	Complemento de preposição	2/110	1,8	0/44	0	0/28	0	2
	Dupla função (SUJ-OBJ)	24/94	25,5	2/8	25,0	0/9	0	26
ESTRUTURA SINTÁTICA	Clivadas propriamente ditas	0/3	0	0/1	0	-	-	0
	Pseudo-clivadas	0/4	0	-	-	-	-	0
	Clivadas “é-que”	0/3	0	-	-	-	-	0
	Clivadas Foco-SER	-	-	-	-	-	-	-
	Clivadas duplo Foco	-	-	-	-	-	-	-
	Clivadas “Que-foco”	4/9	44,4	-	-	-	-	4
	Tópico	0/13	0	0/1	0	0/3	0	0
	Modificada por advérbio	0/12	0	0/1	0	0/3	0	0
OUTROS	<i>“marcadores conversacionais”</i>	77/90	85,6	2/2	100	8/10	80,0	87
	<i>Expressões (quase) cristalizadas</i>	55/78	70,5	1/1	100	7/12	58,3	63
	Repetições / hesitações	24/62	38,7	8/18	44,4	1/3	33,3	33
	Repetições de SVs	9/24	37,5	1/6	16,7	-	-	10
	Truncamentos	9/94	9,6	8/27	29,6	2/4	50,0	19
	Truncamentos c/ verbo subentendido	16/30	53,3	5/9	55,6	0/1	0	21
	Isolados	0/9	0	0/3	0	-	-	0
	Usos enfáticos	0/4	0	0/2	0	-	-	0
TOTAL		1405/3058	45,9	509/1161	48,1	144/409	35,2	2058/ 4628

Tabela 7: Distribuição geral de CÊ conforme a Função/Estrutura Sintática e o contexto

Nas estruturas clivadas, topicalizadas ou quando o pronome está sob o escopo de um advérbio, o emprego da variante reduzida restringiu-se às orações

clivadas do tipo “Que Foco” (e.g., *cê que sabe*). Em face dos poucos dados obtidos, seria prematuro afirmar que, naqueles contextos, a forma plena seria categoricamente empregada. No entanto, como não houve nenhuma ocorrência da variante reduzida nas demais construções clivadas, apenas as orações clivadas *Que-Foco* foram analisadas quantitativamente.

Para não inflacionar os dados, certos tipos de ocorrências também foram excluídas: repetições, truncamentos, construções (quase) cristalizadas na língua.

As repetições (ou “paralelismos plenos”) foram divididas em dois tipos: aquelas em que o falante repete somente a variante pronominal (45) e aquelas em que o falante pronuncia o mesmo sintagma verbal (46). Em ambos os casos, considerou-se somente a última ocorrência da série repetida.

(45) mas é... **cê** / **cê** *nota* pela pela aplicação dos professor e... e a cobrança, né? (Ulisses-M3F-2006)

(46) (...) o serviço de saúde da saúde de do do estado não tem... **cê num consegue!**...**cê num consegue**... EU levei uma amiga minha. (Clara-F3F-2007)

Além disso, muitas vezes o pronome ocorre frequentemente em uma mesma estrutura sintática e encerra uma função discursiva, funcionando como uma espécie de “marcador conversacional” (Corradello, 1997) – como nos exemplos a seguir:

(47) Então **cê vê**... o próprio comitê entre São Paulo e Rio optou por Rio né? (Fabiana-F1S-2003)

(48) São obrigados né, são/ estão sujeitos aí a [hes.] aceita(r) aquilo que... **cê veja**, quantas e quantas aí... (Frederico-M3S-2003)

(49) Quando o Lucas e o Denis nasceram, a gente tava... então **cê imagina**, depois aí então, logo a gente começo a cáí, né? (Frederico-M3S-2003)

(50) Então foi meio corrido, mas assim foi gostoso pelo lado de:... **cê sabe?** chega(r) de noite e sabe(r) que tá tudo no lugar. (Edith-F2F-2008)

(51) Não, ele escrevia bem, sabia... olha, **cê acredita?**... (José-M3S-2006)

(52) Quando eu tô sem pacote sem coisa... eu passo na/ na catraca... **cê entendeu?** (Denise-F3S-2009)

(53) ... Tem... cantores, **cê acha?** a pessoa vai, se elege, consegue o voto... (Marta-F2F-2008)

Todas as ocorrências anteriores são pronunciadas com uma determinada entonação: uma espécie de pergunta retórica (*Cê sabe?, cê entendeu?*) ou expressando ênfase (*cê vê!, cê veja!, cê imagina!*). Tais exemplos podem ser parafraseados por *vê!, veja!, imagina!, sabe?, acredita?, entendeu?* Foi notório o emprego constante de *cê entendeu?* no decorrer de duas entrevistas, sempre com a variante reduzida *cê* ou com um zero (*entendeu?*). No exemplo abaixo, tal expressão é empregada com e sem o pronome em um mesmo enunciado:

- (54) o professor tá dando uma matéria toca o telefone o cara atende o tele/... **cê entendeu?**... [tossiu]... então que(r) dize(r)... isso daí prejudica quem?... não prejudica o professor... ele tá lá ganhando o dinheiro dele... prejudica você que é o aluno... **entendeu?** (Marco-M3FI-2007)

Outros casos considerados “expressões (quase) cristalizadas” na língua, embora não apresentem uma entonação prototípica às anteriores são aqueles em que o *você/cê* é empregado com os verbos *ver, saber, achar, acreditar, imaginar e entender* (independentemente das terminações verbais) seguidos de uma oração iniciada pelo complementizador *que* (55-56); e as expressões *cê me pego(u)* e *p’cê ter (uma) ideia* (57). Estas duas últimas, por exemplo, ocorrem sempre com variante reduzida do pronome e também aparentam funcionar como uma única unidade de processamento (v. capítulo II).

- (55) ... **você vê que**... que é tão drástico aqui o voto que se você não vota(r), você, por exemplo, vai faze(r)... tenta(r) faze(r) um concurso público, se você ... num tive[r]... é... votado... você perde a vaga. (José-M3SI-2006)

- (56) Eles adora(m) estuda(r)... **cê pensa que** eles falta(m) na escola?... eles não falta(m) não... (Claudia-F1FI-2007)

- (57) Tem gente que vai de carro busca(r) o leite **p’ cê te(r) ideia**. (Jefferson-M2FI-2008)

Na tabela 7 mais acima, há uma elevada frequência de *cê* tanto nos denominados “marcadores conversacionais” quanto nas “estruturas (quase) cristalizadas”. Enquanto a média geral de emprego da variante reduzida é em torno de 45%, seu percentual de emprego nessas construções é em torno de 70 a 80% (considerando-se os dados dos anos 2000). Tais ocorrências foram também descartadas do envelope de variação não somente devido à questão quantitativa,

mas porque se tratam de estruturas idênticas que, muitas vezes, concentram-se apenas em algumas entrevistas.

Excluíram-se também as ocorrências em que o pronome encontra-se isolado (58) ou é pronunciado com uma entoação enfática (59), pois, em ambos, a variante plena foi categórica (v. tabela 7).

- (58) Doc.: *(sabe que) eu comecei a faze(r) nutrição lá... em dois mil e um.*
Inf.: *aonde?*
Doc.: *lá na CaRlos de Campos*
Inf.: ***você**?....* (Denise-F3S-2009)

- (59) Obrigado **VOCÊ** por te(r) paciência de me ouvi(r)...¹⁶ (Fábio-M1F-2007)

Por outro lado, considerou-se as ocorrências em que o pronome não era acompanhado por um verbo, mas este era subentendido:

- (60) Inf.: *e catava esses matinhos assim as folhas maiores era o dinheiro.*
Doc.: *[risos] eu {brincava assim também}*
Inf.: ***cê também**? [risos]* (Denise-F3S-2009)

- (61) Ela casou com um cearense... **cê num cons/** *cê num vai comer macarrão e colocar farinha de mandioca no macarrão...* (Marco-M3F-2007)

Em (60), subentende-se o verbo “brincar”: “*cê também [brincava disso]?*”; e em (61), tem-se tanto o advérbio de negação “num” quanto o truncamento possivelmente do verbo “conseguir”.

Em suma, o envelope de variação deste estudo define-se: pelas variantes *você* e *cê* nos contextos em que estas desempenham a função sintática de sujeito na ordem canônica do PB (SV, i.e. sujeito-verbo, ainda que o verbo esteja subentendido) ou uma dupla função (objeto-sujeito) – exceto em estruturas topicalizadas ou clivadas (a única exceção são as clivadas do tipo *Que-Foco*). Foram excluídas também as ocorrências do pronome em uma coordenação, sob o escopo de um advérbio, enfáticas, os truncamentos, as primeiras ocorrências em

¹⁶ A transcrição em CAIXA ALTA aqui indica que a pronúncia do pronome foi enfática.

repetições e as construções “(quase) cristalizadas” ou que funcionam como “marcadores conversacionais.

ENVELOPE DE VARIAÇÃO		
FUNÇÃO SINTÁTICA	Sujeito simples	✓
	Sujeito composto	-
	Sujeito posposto ao V	-
	Objeto direto	-
	Complemento de preposição	-
	Dupla função (SUJ-OBJ)	✓
ESTRUTURA SINTÁTICA	Clivadas propriamente ditas	-
	Pseudo-clivadas	-
	Clivadas “é-que”	-
	Clivadas Foco-SER	-
	Clivadas duplo Foco	-
	Clivadas “Que-foco”	✓
	Tópico	-
	Modificada por advérbio	-
OUTROS	<i>“marcadores conversacionais”</i>	-
	<i>Expressões (quase) cristalizadas</i>	-
	Repetições / hesitações	-
	Repetições de SVs	-
	Truncamentos s/ verbo	-
	Trunc. c/ verbo subentendido	✓
	Isolados	-
	Usos enfáticos	-

Tabela 8: Envelope de variação

Cabe lembrar ainda que, embora os contextos (quase) categóricos ou que funcionam como “marcadores conversacionais” tenham sido descartados da análise quantitativa dos dados, eles funcionam como evidências de que a variante reduzida do pronome estaria se cliticizando na língua e especializando-se com a função sintática de sujeito.

3.3. Grupos de Fatores

Variáveis Extralinguísticas

Foram investigados seis grupos de fatores extralinguísticos: três sociais e três interacionais. Estes últimos foram analisados em 2/3 da amostra dos anos 2000 (24 entrevistas sociolinguísticas)¹⁷. Como não se dispõe das informações necessárias para o estudo das variáveis interacionais na amostra dos anos 1970, sua análise restringe-se à amostra mais recente.

Sexo/Gênero

No capítulo II, discutiu-se que, em diversos estudos sociolinguísticos, as mulheres têm demonstrado maior “sensibilidade ao prestígio social” atribuído às variantes; ao mesmo tempo em que tendem a liderar processos de mudança (Labov, 2001; Chambers, 1995). Com base nesses pressupostos, duas hipóteses podem ser aventadas:

- As mulheres tenderiam a liderar esse processo de mudança em direção à forma inovadora *cê*, uma vez que, na língua falada, essa variante mostra-se neutra em relação a uma avaliação negativa ou positiva; ou
- As mulheres seriam mais “conservadoras”, preferindo a variante “padrão” *você* (a mais antiga na língua, a única registrada pelas gramáticas e menos “informal” do que a forma reduzida *cê*).

Faixa Etária

Se a faixa etária do informante pode refletir estágios anteriores da língua – cf. “hipótese clássica” (v. Capítulo II), pergunta-se:

- a variante *cê* tende a ser empregada pelos mais jovens e evitada quanto mais elevada é a idade do informante? Isto evidenciaria um típico padrão de mudança em progresso, detectada através do constructo *tempo aparente*.

¹⁷ Uma vez que a constituição das amostras não tinha por objetivo coletar tais informações, a sua investigação só se tornou possível graças a um recontato com os documentadores da amostra mais recente (GESOL-SP-2000) – dentre os quais apenas 24 (em um total de 36) deram um retorno. Agradeço a Livia Oushiro por ceder informações de recontato com alguns dos documentadores, cuja fala foi analisada em sua pesquisa de mestrado *Uma análise variacionista para as interrogativas QU-* (inédito).

- comparando os resultados obtidos para a variável faixa etária nas duas amostras analisadas (NURC-SP-1970 e GESOL-SP-2000), a variação *você/cê* constitui um caso de mudança em progresso ou permanece estável ao longo desses 30 anos?

Nível de Escolaridade

Dado que a variante plena é a única veiculada pelas escolas e pelos compêndios gramaticais, uma vez que é evitada na língua escrita, a análise do grupo de fatores escolaridade visa a responder à seguinte questão:

- falantes com maior nível de escolaridade tenderiam a empregar a variante plena do pronome?

Para responder a essa questão, foram considerados dois níveis de escolaridade: superior (concluído ou em curso); e Fundamental (I ou II). Dentre os informantes que não cursaram nível superior, dezesseis interromperam seus estudos entre a 4ª série do ensino Fundamental I e a 8ª série do ensino Fundamental II; e dois informantes apresentam o ensino médio (um havia parado os estudos no segundo ano e o outro o havia concluído). Como estes dois últimos informantes não diferiram muito no emprego das variantes em relação aos informantes com nível fundamental, foram agrupados com aqueles.

Conforme foi discutido no capítulo I, Brown & Gilman (1972 [1960]) opõem os pronomes de “poder” aos pronomes de “solidariedade”. A partir desta oposição, a exemplo do que Gonçalves (2008) propôs para o estudo das formas *você/ocê/cê* em Minas Gerais, pergunta-se acerca da variação *você/cê* se ela se organiza com base na oposição “poder”/“solidariedade”.

Tal hipótese foi investigada mediante a análise de três grupos de fatores interacionais: o grau de intimidade entre os interlocutores; a (as)simetria em relação ao sexo/gênero dos interlocutores e o papel dos interlocutores na entrevista sociolinguística (documentador ou informante).

Grau de intimidade entre os interlocutores

Foram analisados seis tipos de relação entre os interlocutores: parente, amigo, amigo de amigo, colega, conhecido e desconhecido (i.e., não se conheciam previamente à entrevista). Quanto maior o grau de intimidade entre os interlocutores, haveria uma maior informalidade na conversação, o que poderia propiciar o uso da variante reduzida *cê*.

Sexo/Gênero dos interlocutores

Com vistas a analisar a simetria ou assimetria na relação entre os interlocutores, foram estabelecidos quatro fatores: relações simétricas (ambos os interlocutores são do mesmo sexo: M-M ou F-F)¹⁸; relações assimétricas (locutor-interlocutor: M-F ou F-M). Aventa-se a hipótese de que a variante reduzida do pronome (uma forma que indicaria “solidariedade” para com o interlocutor) tenderia a ser empregada nas relações simétricas. Alguns estudos que tratam da variação entre *tu* e *você* (Lucca, 2005; Dias, 2007), por exemplo, constataam que os homens tendem a empregar o “tu” em interações com indivíduos do mesmo sexo/gênero (M-M).

O papel do falante na entrevista sociolinguística

Esse grupo de fatores objetiva averiguar se a variante reduzida do pronome tende a ser evitada na fala dos documentadores, uma vez que eles estão conscientes dos objetivos da entrevista e encontram-se, muitas vezes, em posse de anotações ou de um roteiro a ser seguido.

O quadro a seguir sintetiza as hipóteses aqui aventadas, sob a forma de grupos de fatores (variáveis independentes).

¹⁸ Onde: F = feminino; M = masculino.

VARIÁVEIS EXTRALINGUÍSTICAS		FATORES
SOCIAIS	Sexo/Gênero	masculino feminino
	Faixa etária	1ª faixa etária (20-32) 2ª faixa etária (35-45) 3ª faixa etária (50 ou +)
	Escolaridade	superior fundamental / Médio
INTERACIONAIS	Sexo/gênero dos interlocutores (Doc.-Inf.)	masculino-masculino feminino-feminino masculino-feminino feminino-masculino
	Grau de intimidade entre os interlocutores (Relação entre os interlocutores)	parente amigo amigo de amigo colega conhecido desconhecido
	“Papel” do falante na entrevista	informante documentador

Quadro 11: Grupos de fatores extralinguísticos

Variáveis Linguísticas

Variáveis fonético-fonológicas

Com vistas a verificar a cliticização da variante reduzida *cê* de um ponto de vista fonético-fonológico, Vitral (2002; 2006a) desenvolve análises acústicas das formas *cê* (redução de *você*), *se* (clítico prototípico, monossílabo átono) e *Zé* (nome próprio, monossílabo tônico).

O autor analisa a leitura de cinco informantes (três mulheres e dois homens), aos quais se solicitou que lessem duas vezes pequenos textos em que apareciam as formas *cê*, *se* e *Zé* (Vitral, 2006a: 55)¹⁹. Foi desenvolvida uma análise acústica em que se mediram os parâmetros de intensidade (I) e duração (D), a fim de verificar se, de um ponto de vista fonético, a variante reduzida *cê* aproximar-se-ia de um clítico (*se*) ou de um monossílabo tônico (*Zé*). A média geral dos valores obtidos é apresentada na tabela a seguir, de Vitral (2006a: 57).

¹⁹ A proposta do autor era a de analisar os três itens entre sílabas átonas; em início de oração; e entre sílabas tônicas. As sentenças foram pronunciadas duas vezes por cada um dos falantes. Os dados foram editados pelo programa Soud Designer e analisados acusticamente pelo programa WinPitchi.92.

<i>Se</i>	<i>Cê</i>	<i>Zé</i>
I = 26,6 dB	I = 26,3 dB	I = 31,3 dB
D = 0,129 ms	D = 0,138 ms	D = 0,194 ms

Tabela 9: Duração (D) e Intensidade (I) das Formas *se*, *cê* e *Zé* (Vital , 2006a: 57)²⁰

Para Vital (2006a), apesar da hierarquia obtida entre as três formas analisadas, as distinções encontradas entre *cê*, *se* e *Zé* no que diz respeito à intensidade e duração permitem aproximar *cê* de um clítico²¹.

Ciriaco, Vital & Reis (2004) também desenvolvem análises acústicas das formas reduzidas *cê* (redução de *você*), *ez* (cujo correspondente pleno é o pronome *eles*) e *num* (redução de *não*). Os dados analisados foram obtidos através de um experimento realizado em uma cabine acústica apropriada com quatro informantes de 18 a 21 anos (2 homens e 2 mulheres) – todos estudantes de Letras da Universidade Federal de Minas Gerais (UFMG). Nesse experimento, o pronome *cê* foi contrastado com o monossílabo tônico *Cê* (apelido de Cecília). Os resultados para o parâmetro duração foram²²:

Realização + Ambiente	Duração (ms)
Ce(nome) + sílaba átona	222,6
Ce(pronome) + sílaba átona	152,5
Ce(nome) + sílaba tônica	271,7
Ce(pronome)+ sílaba tônica	224,9

Tabela 10: Duração de *cê* (pronome) e *Cê* (nome) – adaptada de Ciriaco et al., 2004:147-148)

Estes resultados demonstram que, relativamente à duração, *cê* (pronome) apresenta valores menores do que a forma tônica correspondente *Cê* (nome). Uma

²⁰ Onde: dB = decibéis; MS = milésimos de segundos.

²¹ A distinção entre os segmentos [s] e [z] não ocasionaria problemas a essa interpretação, conforme Vital (2006a), pois estudos de Santos (1987; apud. Vital, 2006a: 58) e Stevens & Klatt (1968; apud. Vital, 2006a: 58) demonstram que as fricativas sonoras teriam uma duração menor (aproximadamente 50 ms.) comparativamente às fricativas surdas. A distinção entre a abertura das vogais [e], [i] e [ɛ] também não causaria problemas, pois estudos no inglês americano verificaram uma intensidade de 0,07 dB para as vogais /e/ e /ɛ/, ao passo que a diferença entre os segmentos /e/ e /i/ encontrada é de 3,5 dB (Lehiste, 1970: 120; apud. Vital, 2006a: 58).

²² Utilizou-se do programa Sound Designer II para editar os dados. A análise acústica foi realizada com o programa Macquiner.

comparação das médias das durações destas duas formas com aquelas do reflexivo *se* (um monossílabo átono) é apresentada na tabela abaixo:

Elemento	Duração (ms)
CÊ(nome)	233,4
Cê(pronome)	204,9
Se(reflexivo)	132,9

Tabela 11: Duração de *cê* (pronome), *CÊ* (nome) e *se* (CIRÍACO et al., 2004:149)

Tanto as análises de Vitral (2006a), quanto as análises de Ciríaco et al. (2004) apontam para uma posição intermediária do pronome reduzido *cê* em relação a monossílabos tônicos e a monossílabos átonos. No entanto, diferentemente da análise de Vitral (2006a), Ciríaco et al. (2004) verificam que *cê* (pronome) encontra-se mais próximo a um tônico do que a um elemento átono no quesito duração.

Peres (2006: 85-96) também defende que a escolha entre *você*, *ocê* e *cê* relaciona-se a questões fonético-fonológicas, como estrutura da sílaba, acentuação, ritmo e velocidade da fala. A autora observa que, em Belo Horizonte, a utilização da variante *ocê* com a função sintática de objeto de preposição geralmente ocorre em uma estrutura em que também ocorre a redução das preposições: *c'ocê*, *d'ocê*, *pr'ocê*. Tal fato estaria relacionado, conforme Peres (2006), ao padrão silábico canônico do PB, CV (consoante-vogal), ou ao padrão CCV (consoante-consoante-vogal).

Apoiada em Collischonn (1996), a autora defende que uma construção como “*Vi *cê*” seria evitada não por questões sintáticas, mas por questões fonéticas. Segundo a autora, evita-se tal estrutura para que as duas sílabas acentuadas não ocorram adjacentes. O mesmo não ocorre em uma sentença como “Vi *cê* ontem” (Peres, 2006: 90): nesse caso, o acento da forma *cê* teria sido deslocado para a primeira sílaba do elemento “ontem”.

De acordo com Peres (2006), casos como estes possibilitam a ocorrência de *cê*, com a função sintática de objeto direto devido à presença de outro elemento à sua direita, fazendo com que o acento da variante reduzida se desloque. Essa

hipótese, entretanto, vai em direção oposta à formulada nos itens a seguir. Peres (2006) concebe *cê* portador de acento, que perderia sua proeminência acentual, deslocando-se esse acento para outro elemento à sua direita. Defende-se aqui que, em uma construção como “*Vi cê ontem*”, *cê* pode ser empregado por se tratar de uma forma menos proeminente do ponto de vista acentual, localizada entre duas sílabas tônicas. Ao contrário, uma sentença como “*Vi Zé ontem*” tenderia a ser evitada pela adjacência de três sílabas tônicas²³.

Os três grupos de fatores de natureza fonético-fonológica apresentados a seguir fundamentam-se na “perda de proeminência acentual da variante reduzida *cê*” (Ciríaco et al., 2004; Vitral, 2006a) e nos pressupostos teóricos do *Princípio do Contorno Obrigatório*²⁴ (Guy & Boberg, 1997) – do qual se tratou no capítulo II.

Tonicidade da sílaba seguinte

Considerando-se a variante reduzida *cê* como uma forma intermediária entre monossílabos tônicos e monossílabos átonos – cf. Vitral (2002, 2006a) e Ciríaco et al. (2004) e que as línguas naturais tendem a evitar “choques acentuais” (Selkirk, 1984: 52; apud. Abousalh, 1997), nos casos em que ao pronome se segue uma sílaba tônica, duas estratégias podem ser adotadas a fim de se evitar o encontro de duas sílabas acentuadas:

Retração do acento: o acento passaria da última para a penúltima sílaba do pronome pleno *você*²⁵:

(62)	<u>você</u> <u>sabe</u>	→	<u>vôce</u> <u>sabe</u>
	(. *) (* .)		(* .) (* .)

²³ Peres (2006) não desenvolve uma análise quantitativa em que questões rítmicas sejam consideradas como grupos de fatores.

²⁴ O *Princípio do Contorno Obrigatório* é aqui adotado em uma perspectiva bem ampla, referindo-se a questões que se encontram no nível da sílaba, embora convencionalmente este seja empregado na literatura que trata de traços fonológicos.

²⁵ Ponto (.) indica que a sílaba é átona; asterisco (*) indica que a sílaba é acentuada.

Emprego de uma forma menos proeminente: *cê* tenderia a ser empregada nos contextos em que ao pronome se segue uma sílaba tônica, devido à menor proeminência acentual desta variante:

- (63) você sabe → ce sabe
 (. *) (* .) (.) (* .)

A segunda hipótese constitui objeto de investigação nesta pesquisa. Foram considerados três contextos em relação à tonicidade da sílaba seguinte ao pronome:

a) tônica:

- (64) Independente de se **você** cumpre doze horas de trabalho. (Fabiana-F1S-2003)

b) átona:

- (65) É como se **você** recebesse uma/ funções extras né? (Fabiana-F1S-2003)

c) intermediários? (*tá* / *num*)²⁶:

- (66) O pessoal falava ‘não, **cê** tá falando errado’. (José-F1S-2006)

- (67) ...**cê** num precisa i(r) na na montanha russa {e tal aquelas coisas. (Denise-F3S-2009)²⁷

Objetiva-se com a análise desse grupo de fatores investigar se a variante reduzida *cê* tenderia a ocorrer preferencialmente nos contextos cuja sílaba do elemento seguinte é tônica, evitando-se, assim, um “choque acentual”. Caso se verifique esta tendência, ter-se-ia mais um argumento em favor de que *cê* estaria se cliticizando.

Tonicidade da sílaba antecedente

Controlou-se também a tonicidade da sílaba antecedente ao pronome. Diferentemente da hipótese anterior: uma vez antecedido por sílabas átonas, a variante *cê* poderia ser a preferida, visto seu caráter intermediário entre elementos

²⁶ Decidiu-se classificar separadamente os itens reduzidos *tá* e *num*, pois estes também apresentam uma proeminência acentual intermediária entre elementos tônicos e átonos²⁶ – cf. Ciriaco et al. (2004) e Mendes (1999).

²⁷ Além destes, foi considerado um quarto fator: os casos em que há hesitações ou pausas no contexto seguinte.

tônicos e elementos átonos (Vital, 2006a). Essa hipótese apóia-se também no *Princípio de Alternância Rítmica* de Selkirk (1984; apud. Abousalh, 1997), segundo o qual se tende a evitar não só a adjacência de sílabas tônicas, mas também longas sequências de sílabas átonas (“lapsos de acento”). Embora a forma *cê* possivelmente tenha uma menor proeminência acentual relativamente à última sílaba do pronome *você*, ela teria uma maior proeminência comparativamente à primeira sílaba desse pronome: *você* < *cê* < *você*.

Dissimilação: constituinte /SE/ adjacente

Outra questão formulada é se a presença de elementos semelhantes à variante reduzida *cê* no enunciado poderiam inibir o seu emprego.

Uma sequência como “*se cê se sentir bem*” tenderia a ser evitada, preferindo-se o emprego da forma plena do pronome: “*se você se sentir bem*”? Tal fenômeno também se relaciona ao chamado *Princípio do Contorno Obrigatório*.

Uma das estratégias para se evitar esses encontros é o apagamento de um dos elementos idênticos. Assim, “*Faculdade de Letras*” pode ser pronunciada como “*Faculda(de) de Letras*”²⁸, um caso de haplologia, fenômeno que pode ocorrer não só com elementos idênticos, mas também com sequências de segmentos semelhantes (*lei(te) de coco* e *cal(do) de cana*) (Leal, 2004). Neste estudo, consideram-se semelhantes à variante *cê* os seguintes elementos:

- (i) *se* = conjunção, pronunciada [sɪ]: *se você fosse assim...*
- (ii) *se* = pronome reflexivo, pronunciada [sɪ]: *cê se manteve sentado*.
- (iii) a primeira sílaba de uma palavra começando por [sɛ] ou [sẽ]: *pra você se(r) assim; cê sempre vem aqui, né?*

Os segmentos [sɛ], [sẽ] e [sɪ] não foram diferenciados, devido à pequena quantidade de dados obtidas nos *corpora* analisados. Para facilitar a remissão a esses elementos, será utilizada a notação /SE/.

²⁸ Os parênteses indicam que o segmento foi apagado.

Também foi considerada a posição em que /SE/ ocorre em relação ao pronome (anterior, posterior ou ambas):

a) anterior: /SE/ + pronome:

- (68) **se** cê num tivesse esses curso cê num podia faze(r) vestibular... só que eu ao invés de faze(r) um desses três... eu fiz um curso técnico... [sɪ.se] (Denise-F3S-2009)

b) posterior: pronome + /SE/:

- (69) Eu acordava cedo, passava o dia todo... na atividade, tomando suco, só me alimentando bem... cê **sente** a diferença....pôxa, o ar né. [se.'sê] (Fabíola-F1S-2007)
- (70) Então eu sinto que ele vai ter muita facilidade para escreve(r) (o Dênis)... sabe quando cê **se** realiza (no filho)? [se.sɪ] (Silvia-F3S- 2006)

c) anterior e posterior: /SE/ + pronome + /SE/:

- (71) Sim... aqui na Brasilândia... não sei **se** cê **se** lembra... seu pai deve se lembra(r)... [sɪ.se.sɪ] (Maria-F3F- 2006; Nota: pronunciado com a forma pronominal plena)

d) não há elementos /SE/ adjacentes:

- (72) ... mais acontecimento assim... como assim que cê que(r)...? (Cláudia-F1F- 2008)

Variáveis Morfossintáticas

Coalescência do pronome ao verbo

Ao mesmo tempo em que a variante reduzida *cê* aproxima-se de clíticos mais antigos na língua (*se*, *me*, *lhe*, etc.), dadas as restrições (quase) categóricas ao seu uso em determinados contextos (isolado, complemento de preposição, etc.), *cê* distingue-se desses mesmos clíticos atualmente, por apresentar uma maior mobilidade sintática. Diferentemente de clíticos canônicos, *cê* permite que diferentes constituintes sintáticos ocorram em uma posição intermediária entre ele e o verbo, desde clíticos até sintagmas maiores, como pode ser observado nos exemplos a seguir:

a) Advérbio de negação (pleno/tônico: *não*; ou foneticamente reduzido: *num*)²⁹:

²⁹ *Num* assemelha-se mais a clíticos do que ao seu correspondente pleno '*não*', que é um monossílabo tônico.

- (73) a) Mesmo assim, (em) caso de urgência, cê num tem um atendimento legal. (Jéssica-F1F-2007)
b) Mesmo assim, (em) caso de urgência, cê não tem um atendimento legal. (exemplo teórico)

b) Clíticos:

- (74) É que agora cê me conheceu agora... né... faz alguns meses... totalmente mudada... eu era bem mais gorda... (Ingrid-F2S-2007)

c) “que” em construções “Que Foco”:

- (75) [...] cê que sabe ((riu)) o que eu acho é o seguinte... (Ulisses-M3F-2006)

d) Monossílabos tônicos (*já, só*):

- (76) Ah, é! Só aqueles carrão (assim), cê já brilha o olho! (Jéssica-F1F-2007)
(77) “Bom, ah não, eu sou professora da USP” – “é? Tá bom [[risos]] que mais? cê trabalha ou não?... Ah! cê só dá aula?” (Fabiana-F1S-2003)

e) Elementos discursivos/marcadores conversacionais (*eh, né?*):

- (78) cê né? ter que sair cedo de casa...

f) Advérbios ou locuções adverbiais:

- (79) mas... sei lá cê também tem que te(r) um pouco de humanidade, não sei. (Sílvia-F3S-2006)

g) Sintagmas verbais ou nominais³⁰:

- (80) a) “conversa com o seu irmão né?... tenta passa(r)/ você que entrô na USP tenta passa um pouquinho”. (Sílvia-F3S-2006)
b) “conversa com o seu irmão né?... tenta passa(r)/ cê que entrô na USP tenta passa um pouquinho”.

h) Pausas e hesitações³¹:

- (81) se a gente for tomar uma vacina no:: com o plano de saúde você tem que paga(r) à parte... no convênio né?... então cê:: ... acaba usando o serviço público. (Fábio-M1F-2007)

Uma vez que a interpolação de constituintes entre o clítico e o verbo é um fenômeno que ocorria com clíticos prototípicos no português arcaico (Martins, 1996; apud. Vitral & Ramos, 2006 – v. capítulo II), pode-se aventar a seguinte

³⁰ Apenas o exemplo (80a) foi efetivamente pronunciado pelo informante.

³¹ Em todas as transcrições apresentadas nos exemplos, pausas são representadas por reticências. Quando se tratar de hesitações por parte do falante, utiliza-se a notação [hes.].

hipótese: haveria uma maior tendência de emprego da variante reduzida quando esta se encontrar adjacente ao verbo?

No caso de haver constituintes interpolados entre o pronome e o verbo, pode-se nos questionar também se a extensão fônica dos mesmos correlaciona-se ao emprego de *cê*. Outra distinção a ser estabelecida é sobre a posição que os constituintes interpolados apresentam no sintagma. Dentre os exemplos anteriores, alguns elementos interpolados têm uma posição fixa no enunciado – é o caso do advérbio de negação “não/num”, pré-verbal na comunidade de fala paulistana; e dos clíticos (uma vez que a ênclise é um fenômeno raro atualmente no PB falado). Outros elementos, todavia, apresentam uma posição móvel no sintagma – como os advérbios e os sintagmas nominais e verbais. Pode-se supor, assim, que estes últimos desfavoreceriam o emprego da variante reduzida do pronome.

A fim de investigar essas questões, postulou-se um grupo de fatores que, além de investigar se há interpolação de constituintes, também distingue os tipos de constituintes intervenientes, a exemplo de Andrade (2004).

Tipo de oração interrogativa

Apoiada nos pressupostos da Teoria Gerativa, Ramos (2006 [1997]) analisa se haveria alguma correlação entre o tipo de oração e a pronúncia plena ou reduzida do pronome *você*. São investigados três contextos – os exemplos (82-84) são da autora (p. 48):

a) declarativas:

(82) Maria, *cê* tá muito calma.

b) interrogativas simples:

(83) O *que você* gosta então? / Como *você* acha os ursinhos?

c) interrogativas *wh-que* (interrogativas *que-que*):

(84) O *que que cê* gosta então? / Como *que cê* acha os ursinhos?

Ramos (2006 [1997]) parte da hipótese de que se a variante reduzida *cê* estiver se tornando um clítico fonológico, haveria uma tendência desta ocorrer

quando a posição de núcleo da projeção CP³² estiver preenchida. Conforme a autora, “a sequência [especificador + núcleo lexicalizado + clítico fonológico] não é surpreendente, visto que *cê* teria que ser adjungido ao núcleo do complementizador na forma fonológica” (Ramos, 2006: 49). Logo, se clíticos fonológicos adjungem-se ao núcleo de CP, se *cê* for um clítico ou estiver se cliticizando, construções *wh-que* poderiam favorecer o seu emprego.

Ramos (1997) e Peres (2006) verificam que as orações interrogativas *que-que* favorecem o emprego da variante reduzida – confirmando a hipótese anterior. Neste estudo, trabalha-se com oito contextos sintáticos³³:

- a) declarativas: *Você/cê sabe disso.*
- b) interrogativas sem constituinte interrogativo: *Você/cê conhece?*
- c) interrogativas *Q-in-situ*: *Você/cê vai fazer o quê?*
- d) interrogativas ‘que’: *O que você/cê acha?/ Por que você/cê ...?*
- e) interrogativa ‘que-que’: *O que que você/cê acha?*
- f) interrogativas *é-que*: *O que é que você/cê acha?/ Como é que você/cê veio?*
- g) outras interrogativas *QU-* (*onde, como, quando....*): *Onde você/cê vai?*
- h) outras interrogativas *QU-que*: *Como que você/cê faz pra chegar ao banco?*

Como Andrade (2004) demonstra que uma distinção entre as estruturas interrogativas em que há dois *que* em sequência (*O que que*) e aquelas formadas por constituintes interrogativos distintos de “*que*” seguidos do complementizador (*onde que, quando que, como que, etc*) é importante para explicar a variação no uso do pronome *você/cê* em Brasília, manteve-se essa distinção no presente estudo, e ampliou-se essa oposição para as interrogativas *qu*. Essa distinção tem por objetivo analisar se a variante reduzida tende a ser empregada quando a posição do complementizador é preenchida; ou simplesmente quando há um “*que*” no contexto antecedente (seja ele complementizador ou não) em estruturas interrogativas.

³² Conforme a terminologia adotada na Teoria Gerativa, CP corresponde ao “Complementizer Phrase”.

³³ Oushiro (2010) propõe a seguinte classificação para as estruturas interrogativas de constituinte (*interrogativas-Q*): interrogativas *qu* (*Onde você estuda?*); interrogativas *qu-que* (*Onde que você estuda?*); interrogativas *é-que* (*Onde é que você estuda?*); e interrogativas *qu-in-situ* (*Você estuda onde?*). Aqui utiliza-se parcialmente essa classificação. A oposição entre “declarativas”, “interrogativas simples” e interrogativas com a posição de complementizador preenchida, por sua vez, é de Ramos (1997).

Flexão do pronome

Alguns autores defendem que *vocês* teria entrado primeiro no sistema, em substituição ao *vó* (v. capítulo I)s. Nesse sentido, pode-se pensar que a gramaticalização poderia estar em um estágio mais avançado nas ocorrências do plural do pronome de segunda pessoa, do que nas do singular – o que pode ser expresso em um favorecimento da redução do pronome quando se trata do plural.

No capítulo anterior, discutiu-se que a frequência de uso (Bybee, 2002) também é um fator importante para o avanço dos processos de gramaticalização. Provavelmente, no dia-a-dia, referir-se a um único interlocutor é algo bem mais frequente do que a dois ou mais indivíduos. Sendo assim, pode-se pensar em uma hipótese contrária à anterior: a redução do pronome seria favorecida no singular, dado seu uso possivelmente mais frequente no dia-a-dia dos falantes.

Embora existam dois tipos de segunda pessoa do plural: (i) referente a dois ou mais interlocutores presentes na situação de comunicação; ou (ii) referindo-se ao interlocutor (2ª pessoa) mais uma terceira pessoa, esta última ausente na enunciação, não se fez tal distinção devido à pequena quantidade de dados³⁴ de segunda pessoa do plural obtida.

Variáveis semânticas: Referência do pronome

A noção de “referência” está sendo empregada aqui tanto para a referência determinada (específica ao interlocutor), quanto para os usos indeterminados (genéricos) – a exemplo do que propõe Cançado (2005). Embora não seja objetivo desenvolver um estudo discursivo-pragmático do pronome *você/cê*, a análise se coaduna às propostas de Brow & Yule (1985; apud. Micheletti, 2000), Koch & Marcuschi (1998), Micheletti (2000), Corradello (1997) e Nascimento (2008a) – apresentadas no capítulo II –, pois se considera a presença de “pistas” contextuais e de inferências pragmáticas para estabelecer as distintas referenciações que o pronome *você/cê* adquire no discurso. Aqui, não se faz uma

³⁴ Como as entrevistas do NURC-SP-1970 foram realizadas em duplas de documentadores, há uma maior incidência do primeiro tipo de *(vo)cês* na amostra de fala de 1970. Nos anos 2000, todavia, salvo alguma situação em que terceiros intervieram no decorrer da entrevista, as poucas ocorrências de *vocês/cês* aparecem em discursos reportados.

distinção entre os termos “referência” e “referenciação”, embora o último se mostre mais adequado aos propósitos deste estudo.

Diferentemente dos demais grupos de fatores, para a análise da referência, são consideradas apenas os dados do pronome de segunda pessoa do singular, pois a segunda pessoa do plural possui sempre um referente específico.

Para facilitar a remissão aos distintos empregos do pronome em relação à sua referência, adotam-se as denominações de Câmara Jr. (2006 [1980]; apud. Sousa, 2008):

P1	-	1ª pessoa do singular
P2	-	2ª pessoa do singular
P3	-	3ª pessoa do singular
P4	-	1ª pessoa do plural
P5	-	2ª pessoa do plural
P6	-	3ª pessoa do plural

Corradello (1997) propõe quatro empregos distintos do *você* com relação à sua referência indeterminada (v. capítulo II). Transpondo-os para a terminologia de Câmara Jr. (2006 [1980]), tem-se³⁵:

Graus de indeterminação do pronome “você” (Corradello, 1997)	Equivalente na terminologia proposta por Câmara Jr. (1980)
(i) referente ao próprio locutor (+1ª – 2ª – 3ª)	P1
(ii) referente ao locutor + grupo específico de indivíduos, no qual o interlocutor não é incluído (+1ª -2ª +3ª)	P4 exclusivo
(iii) referente a uma terceira pessoa (– 1ª – 2ª + 3ª)	P3 ou P6
(iv) referência às pessoas em geral (1ª + 2ª + 3ª)	P4 inclusivo

Quadro 12: Graus de indeterminação da referência do pronome *você/cê* (cf. Corradello, 1997)

No presente estudo, além dos quatro empregos acima, foram também considerados usos específicos ao interlocutor (P2), usos “**híbridos**” (cuja referência é simultaneamente específica e genérica); e os “**expletivos**” (usos do pronome supostamente “sem referência” – cf. Duarte, 2003; Vitral & Ramos,

³⁵ A terminologia P1, P4 exclusivo, P4 inclusivo, P3/P6, embora não apareça em Corradello (1997), é aqui empregada para facilitar a comparação entre os vários trabalhos que tratam dessa questão.

2006 [1999]). Procedeu-se também a um detalhamento dos tipos de referência **P1** (1ª pessoa do singular) e **P4 inclusiva** (1ª pessoa do plural inclusiva). A seguir, descreve-se cada um dos empregos, exemplificando-os.

a) Referência P2 (específica):

Trata-se dos casos em que o pronome *você/cê* refere-se ao interlocutor (2ª pessoa do singular) especificamente:

(85) Que mais, **Van**, que cê que(r) sabe(r)? (Jéssica-F1F-2007)

Considerando-se o processo de gramaticalização desse pronome, pode-se esperar que tais usos desfavoreceriam o uso da variante reduzida *cê* por portarem uma referência específica. Por outro lado, como discutiu-se no capítulo I, vários trabalhos que trataram desse fenômeno variável observam que é a referência específica (e não a genérica) que mais favorece o emprego de *cê* (cf. Andrade, 2004; Peres, 2006; Gonçalves, 2008; dentre outros). No presente estudo, verifica-se se essa tendência também se mantém na comunidade de fala paulistana.

b) Referência P1 (1ª pessoa do singular):

Foram classificadas nessa categoria as ocorrências do pronome em que há “pistas” referenciais no discurso que remetem ao próprio locutor – tais como os pronomes *eu*, *me*, *mim*, *comigo*, *meu* e as terminações verbais na primeira pessoa do singular. Em determinados casos, os indícios referenciais encontram-se implícitos – como, por exemplo, em impressões, julgamentos, pensamentos ou sentimentos do próprio locutor, sem haver, contudo, uma remissão explícita a este.

Essa referenciação, por sua vez, foi dividida em dois subtipos: P1 ‘oculta’ e P1 ‘expandida’. Em ambos os casos, há um discurso em primeira pessoa, que é generalizado pelo emprego do *você/cê*. Tem-se, portanto, uma *embreagem actancial*: emprega-se a segunda pessoa em lugar da primeira (*você/cê*, em lugar de *eu*). A diferença entre esses dois subtipos reside na intenção comunicativa de

preservar da face do locutor, que é nítida apenas nos casos de P1 ‘oculto’, assim denominado por haver uma espécie de “ocultamento do eu” (Corradello, 1997).

Nos exemplos (86-87), vê-se o emprego de “EU” no discurso antecedente seguido de uma generalização com o pronome “você/cê”. Em (86), o locutor provavelmente não quer passar a imagem de alguém que é privilegiado em meio à grande maioria das pessoas, que não possui plano de saúde. Em (87), o fato de “encostar em um local e já dormir” também poderia apresentar uma conotação negativa. Nos exemplos (88-89), embora não exista nenhuma “pista” referencial explícita no contexto, têm-se avaliações e percepção do próprio locutor. O emprego de *você/cê*, nestes últimos exemplos, funciona como um “recurso” utilizado pelo locutor (mesmo que inconsciente) para não se comprometer diretamente com as percepções e opiniões expressas. Nestes casos, portanto, algo particular é generalizado (Particular > Geral).

<p>P1 ‘oculto’ PARTICULAR > GERAL (“ocultamento do eu”)</p>	<p>(86) <i>Doc.: cê tem convênio?</i> <i>Inf.: Eu tenho convênio porque é da empresa... que meu marido trabalha ... se fosse para mim paga(r) não teria</i> <i>Doc.: ahan...mas aí então cê num... cê num passa por isso né?</i> <i>Inf.: não... EU não passo por isso... poRque:... CÊ tem uma outra comodidade... né? (Silvana-F2FI-2007)</i></p> <p>(87) Às vezes EU até parece que tirei um cochilo. Parece que cê encostou e tirou um cochilo! Tô desse jeito, Van. (Jéssica-F1FI-2007)</p> <p>(88) EU falo “Meu Deus do céu!” E os cachorro(s) tudo no meio assim, cê fala “Gente, é digno de dó!” (Marta-F2FI-2008)</p> <p>(89) Tinha um cara... o cara era chinês, mas cê falando/ cê falando/[hes.] você falando com ele... ah você tinha vontade de mata(r) o rapaz. (José-M3SI-2006)</p>
<p>P1 expandido GERAL > PARTICULAR</p>	<p>(90) ... é bom também... pra você estuda(r) ... porque aparece palavra que cê nunca viu na tua vida e (isso) vai te forçar ao quê? pra você forma(r) outra cê vai te que pegar o dicionário é o que EU acostumei. (Silmara-F3FI-2007)</p> <p>(91) Hoje é... EU vou ao mesmo à casa que eu morava com os meus pais e você num encontra uma criança brincando com a outra, não tem mais ninguém na rua, né? (Daniel-M2FI-2008)</p>

Quadro 13: *Você/cê* com referência P1

Na referência denominada **P1 ‘expandida’**, não se nota claramente uma intenção de preservação da face por parte do indivíduo. Neste caso, a remissão à primeira pessoa no discurso indica que o locutor é um exemplo em meio às pessoas em geral (daí a denominação P1 ‘expandido’). No exemplo (90), a locutora apresenta os pontos positivos de se fazer palavras cruzadas, sendo ela apenas um exemplo em meio a uma maioria que gosta desse tipo de atividade. Em (91), a locutora menciona que não encontra mais crianças brincando na rua quando vai visitar a casa dos seus pais. Trata-se de algo visualmente perceptível a qualquer pessoa que passa por aquela região. Nestes casos, o particular é apenas um exemplo em meio ao geral (Geral > Particular).

Espera-se que os empregos em que há uma referenciação ao próprio locutor assemelhem-se àqueles em que a referência é específica ao interlocutor, por ambos referirem-se às pessoas da enunciação (primeira e segunda pessoas: *EU-TU*) – cf. Benveniste (1991). A subdivisão entre “P1 oculto” e “P1 expandido”, por sua vez, visa a analisar se ambos apresentam resultados semelhantes de um ponto de vista quantitativo ou se o segundo subtipo estaria mais próximo a uma referenciação P4, que será tratada nos próximos subitens.

c) Referência P4 (1ª pessoa do plural):

Denomina-se P4 os empregos do pronome *você/cê* que se referem a uma primeira pessoa do plural. Uma primeira subdivisão distingue os empregos cuja referenciação é um conjunto de indivíduos composto pelo locutor e uma terceira pessoa (P4 exclusivo, no sentido de que exclui o interlocutor da referência), daqueles que remetem às “as pessoas em geral” (P4 inclusivo).

P4-exclusiva	<p>(92) <i>Doc.: Cê busca eles na na escola?</i> <i>Inf.: Não. EU levo de manhã cedo, MEU MARIDO busca e a minha filha mais velha de vez em quando ela vem sozinha. Mas A GENTE controla o horário né, deu aquele horário A GENTE já fica perto... CÊ já vai atrás. Então assim, A GENTE controla bem assim... não deixa eles sozinhos na rua não. [CÊ = Eu + Meu marido] (Edith-F2FI-2008)</i></p> <p>(93) ANTIGAMENTE tinha-se educação... VOCÊ dava a benção pro teu pai dava benção pra suas tias... né?... VOCÊ frequentava a igreja ia na missa. [VOCÊ = EU + as pessoas de antigamente] (Marco-M3FI-2007)</p>
---------------------	---

Quadro 14: *Você/cê* com referenciação P4-exclusiva

No exemplo (92), o documentador faz uma pergunta à interlocutora. Esta inicia um discurso em primeira pessoa (“*EU levo de manhã cedo e MEU MARIDO busca*”). O discurso continua com o uso do pronome “a gente”, cuja referência é a locutora (1ª pess. sing.) e o seu marido (3ª pess. sing.). No decorrer do turno conversacional, entretanto, a locutora alterna os pronomes “a gente” e “cê”. Este último, portanto, é equivalente ao “a gente”, cuja referência é a primeira pessoa do plural exclusiva (“você” = “a gente” = “eu e meu marido”). Já em (93), o informante, um senhor de 57 anos, relata ao documentador (um jovem de 20 anos de idade) uma situação de seu passado (“*antigamente*”), época em que era habitual dar a bênção aos mais velhos da família e também frequentar a missa. Os usos do pronome *você/cê*, nesse contexto, remetem ao próprio locutor (1ª pess. sg.) e às pessoas de sua época (3ª pess. pl.) – visto que o documentador, 37 anos mais jovem que o informante, não faz parte daquela geração.

Em ambos os casos, tem-se um processo de *embreagem actancial*: o pronome de segunda pessoa é empregado com valor de primeira pessoa do plural “exclusiva” – o interlocutor não está necessariamente incluso entre as pessoas que buscaram a filha da locutora na escola (92) nem no conjunto das pessoas que davam bênção aos mais velhos e que frequentavam a igreja antigamente (93). Uma vez que o processo de referenciação constrói-se “no e pelo discurso” (Mondada & Dubois, 2003 [1995]; Koch & Marcuschi, 1998), a *embreagem actancial* (Fiorin, 2008) cria um efeito discursivo: aproximar o interlocutor do assunto tratado, incluindo-o “virtualmente” naquele conjunto de pessoas.

A referência “P4 inclusiva”, por sua vez, foi subdividida levando em consideração o traço [$\pm$ definido], como nos exemplos a seguir:

Referências P4 “inclusiva”	
[+ definida]	(94) Tá ficando cada vez pior eu acho né? pra__ se(r) PORTEIRO hoje CÊ tem que te(r) o segundo grau... né? (Fabiana-F1SI-2003)
[- definida]	(95) <i>DOC.: o que você acha que hoje seria prioridade na prefeitura?... O que você sente que VOCÊ precisa e que TODO MUNDO precisa?</i> <i>INF.: [que TODO MUNDO...A saúde. A saúde eu acho que tem que se(r) principal. Porque sem saúde VOCÊ não vive bem, sem médico, sem... a saúde pra mim é o principal. (Marta-F2FI-2008)</i>

Quadro 15: *Você/cê* com referênciação P4 ‘inclusiva’

No exemplo (94), a presença do sintagma nominal (SN) “porteiro” restringe a referência do pronome *você* a “um determinado conjunto de pessoas, dentro de um universo mais amplo” – no caso, aquelas que possam/queiram ser porteiros, pois não são todas as pessoa que pretendem exercer essa profissão. Em (95), por sua vez, o pronome refere-se a todos os indivíduos, indistintamente, uma vez que a saúde é uma necessidade de todos os humanos, e não apenas de uma parcela deles. Dizemos, portanto, que a referência do pronome porta o traço [+ definido] no primeiro caso, e o [- definido] no segundo – embora ambas apresentem uma referênciação de primeira pessoa do plural “inclusiva” (1^a + 2^a + 3^a pessoas).

Se o mesmo padrão depreendido em Minas Gerais (pPeres, 2006; Gonçalves, 2008) e em Brasília (Andrade, 2004) se mantiver na cidade de São Paulo – i.e. o de que o uso específico favorece o emprego da variante reduzida do pronome –, pode-se esperar que os usos com uma referência P4 favoreçam um pouco menos o emprego da variante *cê* do que os usos com referências P2 e P1, mas apresentam um maior favorecimento do que os usos com referência P3 (que referem-se a uma terceira pessoa, i.e. uma pessoa fora da enunciação). De maneira similar, espera-se que os empregos denominados P4 inclusivos (sobretudo os que portam o traço [+ definido]) favoreçam mais o uso de *cê* do que os usos em P4 exclusivos, pois estes últimos não incluem o interlocutor em sua referênciação.

d) Referência P3 / P6 (3ª pess. sing. ou pl.):

Quando as “pistas” referenciais presentes no discurso remetem a uma terceira pessoa (singular ou plural), as ocorrências do pronome foram classificadas como P3/P6³⁶. Em (96), por exemplo, nem o locutor, nem o interlocutor são empresários: emprega-se a segunda pessoa para fazer referência a uma pessoa fora do âmbito da enunciação (uma terceira pessoa), via uma *embreagem actancial*.

<p>Referência P3/P6</p> <p>(3ª pessoa)</p>	<p>(96) Então O EMPRESÁRIO que pensa que eh, aumentando... havendo um aumento coletivo dos salários ele vai diminuir o seu lucro..., eu acho que O EMPRESARIADO BRASILEIRO é inteligente, ele não pensa dessa forma. É claro que eh, TODO BOM, BOM NEGOCIANTE sabe que se o seu eh, cliente tiver eh, boas condições pra adquiri(r) o seu produto, mais pessoas vão adquirir o seu produto e aí <u>você</u> passa a ganha(r) ... te(r) um lucro mais saudável, maior, né? com/ com maior quantidade.</p> <p>(Mauro-M1SI-2003)</p>
---	---

Quadro 16: *Você/cê* com referenciação P3/P6

e) Referência Híbrida (específica e genérica):

Os empregos considerados “híbridos” em relação à (in)determinação de sua referência (Nascimento, 2008a, b) são aqueles em que o pronome refere-se simultaneamente ao interlocutor (referência específica) e a “qualquer pessoa” (referência genérica). O termo “híbrido” parece mais adequado do que o termo “ambíguo”, uma vez que esse último pode implicar uma leitura ou outra; enquanto, “híbrido” indica as duas leituras simultaneamente.

<p>Referência Híbrida</p>	<p>(97) <u>Por exemplo, SE você</u> mora em frente uma pracinha, <u>cê</u> vê a pracinha detonada, ninguém tem a coragem de pega(r) um enxada e i(r) limpa(r). (Marta-F2F-2008)</p> <p>(98) <u>Se você</u> me pergunta(r) hoje se eu tenho candidato, a resposta é “não”. (Jéfferson-M2FI-2008)</p>
----------------------------------	---

Quadro 17: *Você/cê* com referenciação Híbrida

³⁶ Não foi feita uma distinção de número, uma vez que esses dois casos por vezes se confundem: as expressões “um empresário”, “o empresário” e “os empresários”, por exemplo, referem-se aos indivíduos daquela classe profissional, independentemente da morfologia que acompanha esses sintagmas nominais.

Nos exemplos anteriores, as expressões “por exemplo” e “se” possibilitam uma leitura hipotética, que não se aplica unicamente ao indivíduo com quem se fala, mas a qualquer pessoa: “*se uma pessoa mora em frente a uma pracinha*”, “*se alguém perguntar*”.

Dada a gradualidade de qualquer processo de gramaticalização e o fato de que, em fases intermediárias, “ambiguidades” podem ser detectadas (Heine, 1991), a expansão do emprego do pronome *você/cê* a contextos com interpretação genérica pode ter se iniciado a partir de contextos hipotéticos. Consequentemente, pode-se esperar um comportamento intermediários quanto ao uso das variantes nestes casos relativamente aos empregos com referência específica (P2) e genérica (P1, P4 e P3).

f) Expletivos (sem referência):

Foram considerados empregos expletivos do *você/cê* quando este preenche a posição sintática de sujeito em sentenças com o verbo “ter”, cuja omissão do pronome permite uma interpretação existencial – como no exemplo (99).

Expletivos	<p>(99) a) hoje você já tem outras instituições aí tão caminhando, tão indo bem, tão evoluindo né? (Fabiana-F1S-2003)</p> <p>b) hoje ____ já tem outras instituições aí tão caminhando, tão indo bem, tão evoluindo né?</p> <p>c) hoje ____ já existem outras instituições aí tão caminhando, tão indo bem, tão evoluindo né?</p> <p>d) hoje ____ já há outras instituições aí tão caminhando, tão indo bem, tão evoluindo né?</p>
-------------------	--

Quadro 18: Usos expletivos de *você/cê*.

Embora as sentenças (99a-d) acima possam ser consideradas equivalentes e, consequentemente, *você* seria um expletivo nessa construção, há várias controvérsias em torno do PB apresentar ou não empregos expletivos de pronomes pessoais – como será discutido adiante.

À parte essas controvérsias, uma análise que considere estes contextos separados daqueles com referência genérica faz-se necessária a fim de observar se, de um ponto de vista quantitativo, tais usos assemelham-se ou não aos

empregos genéricos do pronome, no que diz respeito ao uso de suas variantes plena e reduzida.

g) Outros casos:

Os casos cuja interpretação da referência do pronome foi prejudicada devido à falta de contexto (100-102) foram considerados em uma categoria à parte, denominada “outros” e, posteriormente, foram excluídos da análise quantitativa desse grupo de fatores³⁷ - totalizando 84 ocorrências (80 na amostra dos anos 2000 e 4 na amostra dos anos 1970).

(100) Acho que antigamente você tinha/... os jogadores eram mais, vamos supor, mais time assim, né? Mais/ mais camisa vamos dize(r) assim. (José-M3SI-2006)

(101) Então esse é o problema... Você faz/... surgiu muitas faculdades, mas não surgiu a especialização para professores, por exemplo... (José-M3SI-2006)

(102) você num tinha muito, vamos supor... é... você vê por exemplo, eu tava lá num tempo (como eu tava comentando pro cê)... (José-M3SI-2006)

Inicialmente, a análise da referência do pronome visa a verificar se haveria correlação entre os processos de “perda de material fônico” (*erosão*) e “perda de conteúdo semântico” (*bleaching*). É razoável esperar que a variante reduzida do pronome tenda a ocorrer nos contextos com referência genérica ou sem referência (usos expletivos).

Além disso, os nove empregos do pronome *você* apresentados anteriormente poderiam corresponder a distintas etapas no processo de gramaticalização desse item. Embora *vossa mercê* seja morfologicamente uma forma de terceira pessoa, e, assim como *você/cê*, leve o verbo a uma conjugação também de terceira pessoa, esta expressão foi originalmente empregada para se referir a um interlocutor, à segunda pessoa do discurso. Pode-se esperar assim, que o emprego das variantes plena e reduzida desse pronome possa estar correlacionado às pessoas envolvidas em sua referência: de um lado, a 1ª e a 2ª

³⁷ Diferentemente dos truncamentos logo após o emprego do pronome, nos truncamentos de sentença é possível se analisar a atuação de outros grupos de fatores, exceto os de natureza semântica, que requerem um contexto maior para sua plena interpretação – daí dados como estes serem excluídos apenas nas análises da referência do pronome.

pessoas (presentes na situação enunciativa) e, de outro, a 3ª pessoa (ausente na enunciação) – ou seja, a oposição *pessoa / não-pessoa* (Benveniste, 1991) pode ser útil para explicar esse fato variável.

Os nove empregos apresentados anteriormente são sumarizados na tabela abaixo:

Referenciação	Pessoas			[referencial]	[exofórico]
	1ª	2ª	3ª		
P2		x		+	+
Ambígua/Híbrida	x	x	x	+	+
P1 ‘oculto’		x		+	+
P1 ‘expandido’	x	x	X	+	+
P4 inclusiva [+def.]	x	x	X	+	+
P4 inclusiva [- def.]	x	x	X	+	+
P4 exclusiva	x		X	+	+
P3/P6			X	+	-
Expletivo	???			-	-

Tabela 12: Referenciações do pronome *você/cê* e traços semântico-referenciais

Sendo assim, pode-se propor a seguinte escala:

P2 > P1 [oculto] > P1 [expandido] > P4 incl. [+def.] > P4 incl. [-def.] > P4 excl. > Expletivo

[+ referencial] ----- [- referencial]

[+exofórico/enunciação] ----- [-exofórico/enunciação]

***Você/cê* em sentenças existenciais: expletivo ou genérico?**

Como já se mencionou, o estatuto expletivo do *você/cê* em sentenças existenciais com o verbo “ter” não é consensual entre os pesquisadores. Nesse tópico, retoma-se essa discussão e analisam-se qualitativamente alguns dados.

Nos exemplos discutidos anteriormente, a omissão do pronome levou a uma interpretação existencial da construção. No entanto, nem toda sentença existencial com o verbo “ter” admite a inserção do *você/cê* na posição de sujeito –

o que coloca em dúvida o estatuto expletivo desse pronome. Avelar (2006:06) considera duas restrições a esse emprego³⁸:

(i) a ausência de um sintagma locativo:

- (103) a) (***você**) *tem E.T.*
b) (**você**) *tem E.T. em diversos filmes de Hollywood.*

(ii) a simultaneidade entre o fato enunciado e a enunciação:

- (104) a) (***você**) *tem um carro buzinando lá na rua.*
b) toda noite (**você**) *tem um carro buzinando lá na rua.*

Para o autor, o pronome apresenta uma referência genérica/indeterminada nos exemplos acima, e teria se movido do sintagma locativo para a posição de sujeito da sentença – daí a ausência de locativos “vetar” o emprego do pronome em tais construções existenciais. Um contra-exemplo a essa proposta, contudo, é a sentença a seguir:

- (105) Então é::... **existem** muitas oportunidades... só que o que eu queria ressaltar é que... apesar de aparentemente **você**... **ter** estas oportunidades... **você** num **tem** uma divulgação... é:: importante... (Cecília-F2SI-2007)

Nesse caso, além da alternância entre “existir” e “você ter” em um mesmo enunciado, verifica-se que o uso do pronome em sentenças existenciais com “ter” ocorre também na ausência de qualquer sintagma locativo. Este não é um caso isolado: muitos outros dados como este foram encontrados no *corpus* aqui analisado.

A impossibilidade de preenchimento do sujeito em determinadas construções com “ter” existencial, portanto, transcende questões de natureza sintática. O próprio Avelar (2007) alerta para questões pragmáticas envolvidas: “*It is possible that there are pragmatic effects of the você insertion on the interpretation of the existential clause*” (Avelar, 2006: 05). Os exemplos dados pelo autor são:

³⁸ Exemplos de Avelar (2006: 06).

(106) a) *você tem muitos castelos na Europa.*

b) *você tem muitos castelos na Europa que você pode ver se você estiver lá.*³⁹

Neste estudo, considera-se que a inserção do *você* em sentenças como (103) acima seria problemática não devido à ausência de um locativo *per se*, mas por se tratar de uma asserção: “a existência de E.T”. A inserção do sintagma locativo “em diversos filmes de Hollywood” imprime um caráter “hipotético” àquele fato, tornando-o replicável a qualquer indivíduo:

(103) c) *qualquer pessoa que veja filmes de Hollywood, pode ver E.T.s*

Essa hipótese justificaria também por que eventos simultâneos à enunciação não admitem o preenchimento do sujeito nessas construções: trata-se de uma situação particular, situada em um determinado ponto do tempo, que não pode ser generalizada.

Viotti (2003) defende a hipótese de que o verbo *ter* é um “verbo leve”, i.e., “um verbo semanticamente esvaziado” e “desprovido de capacidade predicativa”. Segundo a autora, haveria uma única entrada lexical para este item, e os diferentes sentidos que ele assume nos enunciados (“posse alienável”, “posse inalienável”, “existência”, “locação”, etc.) devem-se a um “princípio de composicionalidade”, segundo o qual seu sentido só pode ser definido a partir do sentido de todas as expressões que compõem a sentença.

Essa hipótese pode explicar as restrições ao emprego do *você* nos contextos existenciais aqui investigados. Em primeiro lugar, parte-se do princípio de que construções existenciais com o verbo “ter”, cuja posição do sujeito é preenchida por *você/cê* admitem duas leituras: uma existencial (107b), nítida quando se omite o pronome; e outra indeterminada (107c):

(107) a) numa cidade pequena, **você tem** mais recursos. (Flávio-M2S-2003)

b) numa cidade pequena, ___ **tem/existem** mais recursos. (Leitura Existencial)

c) numa cidade pequena, **você “encontra/dispõe de”** mais recursos.
(Leitura indeterminada)

³⁹ Grifos nossos.

A presença do pronome na oração faz com que o verbo “ter” adquira um sentido semelhante ao de “encontrar”, “dispor de”, “notar”. O mesmo padrão ocorre nos seguintes exemplos.

- (108) a) Hoje **você tem**:: o interior que tá se desenvolvendo muito.
b) Hoje **você vê/nota** o interior que tá se desenvolvendo muito.
- (109) **você encontrava** poucas casas que tinham televisão, de repente todas... quase **todo mundo tinha** televisão.[...] Eu vou falar pra você, em algumas casas **você tinha** televisão. (José-M3SI-2006)

Em (109), “ter” e “encontrar” alternam-se no mesmo enunciado. Além disso, no trecho “*em algumas casas você tinha televisão*” é ambíguo, pois permite três leituras:

1. “ter” = “encontrar”: “*em algumas casas encontrava-se televisão*”;
2. “ter” = “existir”: “*em algumas casas, existia televisão*”;
3. “ter” = “posse alienável”: “*em algumas casas, as pessoas possuíam televisão*”;

O princípio semântico de “composicionalidade” explicaria por que, na ausência de sintagmas locativos ou temporais em algumas construções, o emprego do *você/cê* não ocorre. Uma vez que a composição do significado do verbo “ter” se estabelece pela “soma” do significado de todos os constituintes presentes na oração, o uso de *você/cê* só ocorre quando, ao lado de uma leitura existencial, *ter* adquire outra interpretação (“encontrar”, “notar”, “ver”, etc.) – que pode ser propiciada pela inserção de determinados locativos.

Se o raciocínio que aqui se propõe estiver correto, o emprego de *você/cê* nessas construções, quando possível, assume um caráter híbrido, situando-se entre um item referencial (genérico) e um item desprovido de referência (expletivo). Tal pronome pode corresponder, por conseguinte, a um estágio intermediário em seu processo de gramaticalização, propiciado pelas estruturas com um verbo “leve”, como “ter”. Argumenta-se, assim, que *você/cê*, nestes casos, situar-se-ia entre itens expletivos propriamente ditos (como o *it*, no inglês) e empregos genéricos (dotados de uma referência). Conforme se argumentou em Nascimento (2007:37):

os supostos usos expletivos do “você / cê” configurariam uma fase intermediária no processo de gramaticalização – como evidencia sua leitura ambígua (ou híbrida), estando entre um uso genérico e um possível (futuro) uso expletivo.

No entanto, seja um genérico, um expletivo ou item intermediário entre estas duas instâncias, faz-se necessário averiguar se os empregos genéricos e aqueles considerados expletivos correlacionam-se de maneira diversa ou não ao emprego das variantes plena e reduzida do pronome em estudo. Nesse sentido é que, à parte das discussões a respeito da formação ou não de itens expletivos no PB, analisaram-se separadamente os empregos genéricos dos supostos usos expletivos do *você/cê* – conforme será apresentado no capítulo IV.

Variáveis discursivas e cognitivas

Paralelismo

Esse princípio foi aqui considerado sob a forma de dois grupos de fatores. O primeiro atua no nível do discurso e visa a analisar a influência que o paralelismo das variantes pronominais de segunda pessoa em séries de orações exerce sobre a seleção das ocorrências subsequentes. O segundo atua no nível do sintagma verbal, verificando se a presença de outras formas reduzidas (*num, tá, tava*, etc) adjacentes ao pronome favoreceria o emprego de *cê*.

Paralelismo das formas pronominais (Paralelismo I):

Considera-se “**série**” uma sequência de orações cujos sujeitos são os pronomes de segunda pessoa. Nos exemplos a seguir, entre as duas primeiras ocorrências do pronome, existe uma oração cujo sujeito não é o pronome de segunda pessoa. As orações sublinhadas em (110) e (111) são, ambas, orações encaixadas: uma oração substantiva objetiva direta e uma oração relativa, respectivamente. Em casos como estes, considerou-se que há uma “série” desde que a oração encaixada não ultrapassasse a extensão de uma linha de transcrição ortográfica.

- (110) assim **cê** sente que eles são mais.../ que eles conservam a cultura própria... é (hes.) [RP]é] difícil **você** entra(r) na comunidade se são comunidades muito fechadas é difícil **você** entra(r) na comunidade né? (Fabiana-F1S-SP-2003)

- (111) às vezes cê vê disciplinas sendo dada que tem o mesmo foco que a outra, {...} você num sabe muito bem qual que é a diferença. (Fabiana-F1S-SP-2003)

As sentenças com o pronome de segunda pessoa não foram consideradas como integrando uma série em orações não-encaixadas ou orações encaixadas com extensão superior a uma linha de transcrição entre duas ocorrências do pronome.

As repetições ou truncamentos do pronome no contexto antecedente também foram desconsiderados. Em um dado como (112), apenas a terceira ocorrência foi analisada.

- (112) É, você/ você... de repente você encontrava poucas pessoas que tinham televisão, de repente todas... (José-M3SI-2006)

Sujeitos nulos entre as ocorrências do pronome foram codificados em fatores específicos, com relação ao pronome que antecede o sujeito nulo: séries VOCÊ-Zero ou CÊ-Zero.

Seguindo os critérios acima, definiram-se os seguintes tipos de ocorrências:

a) isolada:

- (113) Sim. É, é o que você disse, o desemprego gera violência, claro, as pessoas ficam aí, não é... completamente desorientadas com a questão do desemprego. (Frederico-M3S-SP-2003)
- (114) *Doc.: e no ramo artístico quais os profissionais que a senhora conhece quais que a senhora dá mais valor ou menos?*
Inf.: ah [hes.] você diz como cantores [esp.] o {[hes.]}/?
Doc.: {o tipo de profissional} (Brígida-F2SI-NURC-SP-DID-228-1970)

b) primeira de uma série:

- (115) ... porque às vezes você pega numa administração você não tem, a pessoa não tem esse perfil né? (Fabiana-F1S-2003)
- (116) não sei se vocês conhecem Cabo Frio ... cê num acha que o lugar assim é muito pobre?... (Luís Marcondes-M1SI-NURC-SP-DID-46-1970)

c) segunda de uma série cuja forma antecedente é VOCÊ:

- (117) ... porque às vezes você pega numa administração você não tem, a pessoa não tem esse perfil né? (Fabiana-F1S-2003)

- (118) ele disse “bom se **você** num que(r) aprende(r) nitro celulose... o [hes.] o lugar certo que **você** vai é o [hes.] Picattini Arsenal. (Eliomar-M3SI-GESOL-SP-1970)

d) segunda de uma série cuja forma antecedente é CÊ:

- (119) Meu pai fala ai “por que que **cê** gasto(u) , por que que **cê** compro(u)?” (Marta-F2F-2008)
- (120) com chuva eu chego sempre atrasado... trânsito é imprevisível né?... **cê** pensa que (es)tá tudo certo... **cê** chega num lugar onde tem uma fila. (Luís Marcondes-M1SI- 1970)

e) terceira (ou mais) de uma série cujas formas antecedentes são VOCÊ:

- (121) ... **você** trabalha, **você** é cidadã, **você**... contribui, mas mesmo assim **você**... tem que pagar o convênio.(Marta-F2F-2008)
- (122) Hoje **você** tem que te(r) um um um convênio caro, que **você** num tem vínculo nenhum com ele, aí **você** parou de paga(r) um mês ou dois **você** perde esse esse esse direito que **você** já vem pagando há há quinze vinte anos. (Daniel-M2FI-2008)

f) terceira (ou mais) de uma série cujas formas antecedentes são CÊ:

- (123) É uma boa. Também é bom... **cê** (forra(r):: é tipo uma) esteira, deitar lá (fica(r)) a tarde inteira, tem estacionamento tem as trilha(s) que **cê** pode para(r) o carro... fica(r) lá... Tem uns lado(s) mais deserto(s), se **cê** quiser fica(r) sozinha lá... (Jéssica-F1F-2007)
- (124) dinheiro faz muita falta... se **cê** num tem dinheiro... **cê** num faz nada na tua vida... se **cê** num tem dinheiro **cê** tem que toma(r) condução... ônibus cheio... se **cê** tem um pouquinho de dinheiro **cê** compra um carrinho nem que se fo(r) velho... (Luís Marcondes-M1SI-1970)

g) terceira (ou mais) de uma série mista, sendo a última antecedente VOCÊ:

- (125) Aí, ou **cê** sobe de carro... Tem uma lotação... que **você** paga... tipo micro ônibus... **cê** pega ele lá embaixo na entrada. (Jéssica-F1F-2007)
- (126) E quando **cê** num tá... trabalhando? O que **você** mais gosta de fazer? Cinema teatro...o que que **você** curte? Balada? (Ellen-F1SD-2007)

h) terceira (ou mais) de uma série mista, sendo a última antecedente CÊ:

- (127) Ah **você** peRcebe que a pessoa fala, né?... **cê** vai cumprimentar alguém, **cê** percebe “Oi, bom dia, boa taRde”. (Marta-F2FI-2008)
- (128) na época de apresentação do filme... **você** chegava pra uma pessoa **cê** falava “**cê** entendeu::? **você** assistiu o filme e gostou?” (Lúcio-M1SI-1970)

i) terceira (ou mais) de uma série mista formada por VOCÊ e ZERO:

- (129) antigamente tinha-se educação... **você** dava a benção pro teu pai Ø dava benção pra suas tias... né?... **você** frequentava a igreja ia na missa. (Marco-M3F-2007)
- (130) “e **você** vai para lá e Ø tem que se vira(r)... **você** aprende a faze(r) nitrocelulose a faze(r) a coisa”. (Eliomar-M3SI-1970)

j) terceira (ou mais) de uma série mista formada por CÊ e ZERO:

- (131) e **você** fica de mãos atadas po[r]que **cê** num tem pra re/ pra quem Ø reclama[r]... **cê** vai reclama[r] pra quem?... (Marco-M3F-2007)
- (132) eu pensei que **cês** vo/ viessem aqui e Ø me fizessem perguntas a respeito da minha profissão... e [hes.] e **vocês** escolheram um tema entende inteiramente assim fora da minha competência viu? (Raquel-F3SI-1970)

Paralelismo de formas reduzidas (Paralelismo II):

De modo análogo à hipótese anterior, investiga-se também se a adjacência de outros itens reduzidos como *num* (redução de *não*) e *ta(r)* (reduções do verbo *estar* em todas as flexões) favorece a ocorrência de outras reduções. O efeito do paralelismo investigado aqui, todavia, estaria em sentido inverso ao anterior: formas reduzidas no contexto seguinte favoreceriam a redução de *você*. Para isso, consideraram-se os seguintes fatores:

a) Presença de *tá* (1 sílaba):

- (133) Eu (es)tu falando aí não sei se **cê tá** entendendo. (Ulisses-M3F-2006)

b) Presença de *tava/távamos/etc.* (2 ou + sílabas):

- (134) Pensou se **cê tive(r)** morrendo, **cê** precis/ **cê** fica quatro horas esperando! (Jefferson-M2F-2008)

c) Presença de *num*:

- (135) ... droga bomba...briga...antigamente **cê num** via essas coisa(s) hoje em dia é tudo:...a violência tá em tudo quanto é lugar né... (Luís-M2F-2007)

d) Presença de duas reduções (*num tá*):

- (136) “Filha não faz assim, **cê num tá** vendo a Sabrina? A Sabrina tá com o material que a mãe comprou no começo do ano”. (Jéssica-F1F-2007)

e) Não existem reduções de elementos adjacentes ao pronome:

- (137) **você** respeitava o professor... o professor... mandava dentro da sala de aula... (Marco-M3F-2007)

Frequência do verbo seguinte

Investigou-se também se a frequência dos verbos que acompanham o pronome *você/cê* favoreceria a redução do pronome. Para tal, calculou-se a frequência *token* dos verbos, dividindo-se o número de ocorrências de cada verbo nas construções cujo sujeito é *você/cê* pelo número total de ocorrências de

“você/cê + verbo” em cada uma das amostras analisadas⁴⁰ (Infs-GESOL-SP-2000, Docs-GESOL-SP-2000⁴¹ e Infs-NURC-SP-1970)⁴².

Considerou-se o item lexical, independentemente de suas flexões verbais. As únicas exceções foram os verbos: *ser* (*é/ser; foi; era*); *ter*: (“*tem + X*”; “*tem/tinha + que*”; outros tempos verbais que não o presente); *ir* (“*ir + infinitivo*” e outros casos); e (*es*)*tar* (seguido ou não de gerúndio). Foram estabelecidos seis frequências (v. anexo 1): até 1%; 1 a 2ª %; 2 a 3%; 3 a 4%; 4 a 6%; e acima de 6%.

O objetivo desse grupo de fatores é verificar se a redução do pronome tende a ocorrer quando este integra orações cujos verbos são empregados com maior frequência na amostra considerada, por uma espécie de “habituação”.

Tipo de discurso

Finalmente, considerou-se também o tipo de discurso em que o pronome ocorre: reportado (138) ou não-reportado.

(138) eu falo: “Gente, como **cês** consegue faze(r) um carnê...Vem com um carro novo e um caRnê junto!” (Marta-F2F-2008)

Como a maioria dos casos de discursos reportados ocorre com referência específica a um interlocutor “virtual” ou “hipotético”, a análise dessa variável visa a observar se a variante plena tende a ser a preferida nos discursos reportados.

Os quadros 19 e 20, nas páginas seguintes, sintetizam os grupos de fatores linguísticos analisados nesta pesquisa.

Para manter a *ortogonalidade*⁴³ entre todos esses grupos de fatores linguísticos, um cuidado especial foi dado às seguintes variáveis independentes:

⁴⁰ A opção por computar separadamente a frequência dos verbos nessas três amostras deve-se à presença de temas diversos, que poderiam propiciar o emprego de um dado verbo nas entrevistas dos anos 1970, mas não nas entrevistas dos anos 2000, por exemplo. A frequência dos verbos na fala dos documentadores e dos informantes também foi analisada separadamente porque apenas os primeiros tinham um roteiro a seguir.

⁴¹ No caso da amostra dos documentadores, para o cálculo da frequência do verbo, além dos 9 documentadores paulistanos, levou-se em consideração também os dados de seis documentadores nascidos no interior do estado de São Paulo.

⁴² Onde: Infs = informantes; Docs = documentadores.

(i) Tonicidade da sílaba seguinte; (ii) Tonicidade da sílaba antecedente; (iii) Coalescência do pronome ao verbo; (iv) Tipo de oração; (v) Flexão do pronome; (vi) Tipo de discurso (reportado *versus* não-reportado); e (vii) Paralelismo II. Durante a apresentação dos resultados, serão destacados os grupos de fatores que foram considerados em cada uma das análises.

A variável (i), por exemplo, interage com as variáveis (iii) e (vii), e somente uma versão concisa destas possibilitam sua análise conjunta. A variável (iv), por sua vez, interage com a (ii), posto que as orações interrogativas *wh-que* sempre terão a mesma sílaba no contexto precedente (*que*). Além disso, o reduzido número de ocorrências obtido para alguns fatores afeta esse requisito. Sendo assim, em determinados casos, fez-se necessário desenvolver análises de variáveis individualmente (*univariadas*), ou em conjunção apenas com os grupos de fatores sociais.

⁴³ De acordo com Guy (2007: 52 [1988]): *os grupos de fatores devem ser “ortogonais” ou quase “ortogonais”, i.e., eles devem co-ocorrer livremente, e não ser sub- ou supercategorias uns dos outros.* (Tradução de: Ana M.S. Zilles). Embora o programa VARBRUL trabalhe bem com um elevado grau de *não-ortogonalidade*, análises em que todos os grupos de fatores considerados na pesquisa foram analisados conjuntamente produziram resultados problemáticos – detectados pela comparação entre as melhores análises quantitativas das fases *step up* e *step down* do Golvarb X (o número de grupos de fatores selecionados nas duas melhores rodadas do programa estatístico e os pesos relativos apresentados não são idênticos).

VARIÁVEIS LINGUÍSTICAS		FATORES
FONÉTICO-FONOLÓGICAS	Tonicidade da sílaba seguinte	Tônica átona mediana (<i>tá/num</i>) *não se aplica (pausa/hesitação)
	Tonicidade da sílaba antecedente	Tônica átona *não se aplica (pausa/hesitação)
	Presença de constituinte /SE/ adjacente ao pronome	antecedendo: /SE/ + pronome seguindo: pronome + /SE/ antecedendo e seguindo: /SE/ + pron. + /SE/ não há
MORFOSSINTÁTICAS e SINTÁTICAS	Coalescência entre o pronome e o verbo	coalescente/adjacente clíticos (<i>se, me, etc.</i>) ' <i>não</i> ' ' <i>num</i> ' (redução de 'não') monossílabo com ditongo nasal (<i>nem</i>) monossílabo ditongado não-nasal (<i>mal, mais</i>) "que" monossílabos tônicos (<i>já, só</i>) advérbios/locuções adverbiais (<i>às vezes</i>) sintagmas (+ de uma palavra; exceto locuções) * não se aplica (pausas, hesitações, s/ verbo)
	Tipo de oração	Declarativa interrogativas simples interrogativa <i>que-que</i> outras interrogativas <i>QU-que</i> (<i>onde que</i>) interrogativas <i>que</i> outras interrogativas <i>QU-</i> (<i>onde, quando</i>) interrogativas de <i>QU- in situ</i> interrogativas <i>QU- é que</i>
	Flexão do pronome	Singular Plural
	Função sintática	Sujeito objeto dupla função (sujeito e objeto) complemento de preposição outras

Quadro 19: Grupos de fatores linguísticos: Fonético-Fonológicos, Morfossintáticos e Sintáticos

VARIÁVEIS LINGÜÍSTICAS		VARIANTES
SEMÂNTICA E DISCURSIVA	Referência / Referencialidade do pronome	<p>P1 “oculto” (1ª p. sg. e “ocultamento do eu”)</p> <p>P1 “expandido” (1ª p. sg. como um exemplo)</p> <p>P2 (2ª p. sg. = específica ao interlocutor)</p> <p>P3 (3ª p. sg. ou pl.)</p> <p>P4 [+ definido] (1ª p. pl. = 1ª +/ou 2ª +/ou 3ª)</p> <p>P4 [- definido] (1ª p. pl.=1ª +/ou 2ª +/ou 3ª)</p> <p>P4 “exclusivo” (1ª p. pl. = 1ª + 3ª – 2ª pess.)</p> <p>P5 (2ª p. pl. específica) → Excluídos</p> <p>Híbrido/ambíguo (específico e genérico)</p> <p>Expletivo?</p> <p>* Outros (Não identificável no contexto) – Excluídos</p>
	Tipo de discurso	Reportado não-reportado
COGNITIVAS E DISCURSIVAS	Paralelismo I (pronomes <i>você/cê</i>)	<p>Isolada</p> <p>1ª de uma série</p> <p>2ª de uma série, sendo a anterior VOCÊ</p> <p>2ª de uma série, sendo a anterior CÊ</p> <p>3ª ou + de uma série, sendo as anteriores VOCÊ</p> <p>3ª ou + de uma série CÊ</p> <p>3ª de uma série mista CÊ-VOCÊ</p> <p>3ª de uma série mista VOCÊ-CÊ</p> <p>3ª de uma série mista VOCÊ-Zero</p> <p>3ª de uma série mista CÊ-Zero</p>
	Paralelismo II (formas reduzidas adjacentes ao pronome: <i>num, tá</i>)	<p><i>tá</i> (redução de <i>estar</i> / 1 sílaba)</p> <p><i>tava/tive</i> (redução de <i>estar</i> / 2 ou + sílabas)</p> <p><i>num</i> (redução de <i>não</i>)</p> <p>2 reduzidos (<i>num</i> + redução de <i>estar</i> : <i>num tá</i>)</p> <p>não há reduções</p>
	Frequência do verbo em estruturas “ <i>você/cê</i> + verbo” no <i>corpus</i>	<p>até 1%</p> <p>de 1 a 2%</p> <p>de 2 a 3%</p> <p>de 3 a 4%</p> <p>de 4 a 6%</p> <p>acima de 6%</p>

Quadro 20: Grupos de fatores linguísticos: Variáveis Semânticas e discursivo-cognitivas

CAPÍTULO IV

Análise Quantitativa

Language is conceived [...] as an abstract pattern, exterior to individual. In, fact, it can be argued that the individual does not exist as a linguistic entity. That is not to say that we do not study individuals [...]. But the individuals we study are conceived of the product of their social histories and social memberships.

(Labov, 2006: 05 [1966])

Introdução

Neste capítulo, interpretam-se os resultados obtidos nas análises quantitativas dos dados, a partir dos grupos de fatores que foram estabelecidos, com base nas hipóteses discutidas no capítulo anterior.

As análises multivariadas foram feitas com o programa estatístico VARBRUL, na versão Goldvarb X (Sankoff, Tagliamonte & Smith, 2005). Primeiramente, são interpretados os resultados da análise da amostra dos anos 2000. Em seguida, interpretam-se os resultados das análises da amostra dos anos 1970, em comparação à amostra anterior.

4.1. Análise sincrônica – Amostra *GESOL-SP-2000*

Distribuição geral dos dados

Considerando-se as ocorrências do pronome *você(s)* provenientes da fala de 45 paulistanos (36 informantes e 9 documentadores) em todos os contextos sintáticos, incluindo-se os casos de repetições e truncamentos (v. capítulo III), obteve-se um total de 4227 ocorrências: 2253 dados (53%) com a variante plena *você*; e 1855 (44%) com a forma reduzida *cê*. Estas duas variantes juntas totalizam 97% das ocorrências do pronome no *corpus* geral. Os 3% restantes

ocorrem com as formas “ocê” (5 ocorrências), “ucê” (3 ocorrências), “vcê” (52 ocorrências) e *cê* cuja pronúncia não pôde ser determinada com segurança¹ (59 ocorrências).

VARIANTES	CÊ						
	GESOL-SP-2000				TOTAL		
	(36 INFs.)		(9 Docs.)				
	Nº	%	Nº	%	Nº	%	Total
VOCÊ	1616	52,7	637	54,8	2253	53,32	4108 (97,18%)
CÊ	1367	44,6	488	42,0	1855	43,87	
[vse] (você?)	<u>37</u>	1,2	<u>15</u>	1,3	<u>52</u>	1,23	119 (2,82%)
CÊ (?)	<u>38</u>	1,2	<u>21</u>	1,8	<u>59</u>	1,39	
OCÊ	05	0,2	-	-	05	0,12	
UCÊ	02	0,1	01	0,1	03	0,07	
Total	3065	100	1162	100	4227	100	4227

Tabela 13: Pronomes de 2ª pessoa empregados pelos paulistanos (GESOL-SP-2000)

O número das ocorrências de “ocê” e “ucê”, conforme já havia sido comentado anteriormente, foi muito pequeno no *corpus* analisado – razão pela qual foram excluídos da análise quantitativa. Também o foram aqueles casos cuja pronúncia do pronome não pôde ser precisamente estabelecida. Dessa forma, o total de dados seria 4108. No entanto, conforme a discussão desenvolvida no capítulo anterior, nem todas as 4108 ocorrências do pronome encontram-se em contextos variáveis (i.e., que permitem a comutação pela outra variante em análise). Na tabela a seguir, tais ocorrências são destacadas, conforme sua função sintática ou o tipo de estrutura sintática que integram (clivada, topicalizada, em coordenação, etc.); aparecem também os dados truncados, repetições e estruturas (quase) cristalizadas.

¹ Os dois últimos casos foram chamados de duvidosos. Trata-se de casos cuja percepção auditiva foi dificultada (por problemas com a qualidade da gravação, pela velocidade da fala do informante, etc).

FUNÇÃO / ESTRUTURA SINTÁTICA		INFs GESOL-2000		DOCs GESOL-2000		Total	
		Cê/Total	%	Cê/Total	%	CÊ	
FUNÇÃO SINTÁTICA	Sujeito simples	1153/2349	49,1	461/944	48,8	1614	3293
	Sujeito em perguntas específicas*	---	---	6/60	10,0	6	60
	Sujeito / objeto em coordenação	0/1	0	0/6	0	0	7
	Sujeito posposto ao V	0/2	0	-	-	0	2
	Objeto direto	0/9	0	0/4	0	0	13
	Complemento de preposição	2/110	1,8	0/44	0	2	154
	Dupla função (SUJ-OBJ)	22/88	25,0	2/8	25,0	24	96
ESTRUTURA SINTÁTICA	Clivadas propriamente ditas	0/3	0	0/1	0	0	4
	Pseudo-clivadas	0/4	0	-	-	0	4
	Clivadas “é-que”	0/3	0	-	-	0	3
	Clivadas Foco-SER	-	-	-	-	-	-
	Clivadas duplo Foco	-	-	-	-	-	-
	Clivadas “Que-foco”	4/9	44,4	-	-	4	9
	Tópico	0/13	0	0/1	0	0	14
	Modificada por advérbio	0/12	0	0/1	0	0	13
OUTROS	<i>“marcadores conversacionais”</i>	76/88	86,4	2/2	100	78	90
	<i>Expressões (quase) cristalizadas</i>	54/77	70,1	1/1	100	55	78
	Repetições / hesitações	23/60	38,3	7/17	41,2	30	77
	Repetições de SVs	9/23	39,1	0/5	0	9	28
	Truncamentos	8/90	8,9	7/20	35,0	15	110
	Truncamentos (v. subentendido)	16/29	55,2	2/6	33,3	18	35
	Isolados	0/9	0	0/3	0	0	12
	Usos enfáticos	0/4	0	0/2	0	0	6
TOTAL Geral		1367/2983	45,8	488/1125	43,4	1855	4108
CONTEXTOS VARIÁVEIS		1195/2475	48,3	465/958	48,5	1660	3433

Tabela 14: Emprego de CÊ conforme a função e a estrutura sintática

Conforme foi apontado no capítulo II, repetições do pronome, truncamentos e os casos que integram construções “quase cristalizadas” poderiam inflacionar o número total de ocorrências do pronome e levar resultados menos precisos a respeito das tendências de uso das variantes. Tais casos, assim como

aqueles que se mostraram categóricos ou quase categóricos nas amostras foram desconsiderados, para os fins da análise quantitativa. Além disso, foram descartadas as ocorrências do pronome-sujeito em perguntas previamente elaboradas (que faziam parte de um roteiro) nas falas dos documentadores, que apresentaram uma alta taxa de emprego da forma plena do pronome. Cabe, então, rever o envelope de variação, definido no capítulo anterior:

O envelope de variação deste estudo define-se: pelas variantes *você* e *cê* nos contextos em que estas desempenham a função sintática de sujeito na ordem canônica do PB (SV, i.e. sujeito-verbo, ainda que o verbo esteja subentendido) ou uma dupla função (objeto-sujeito) – exceto em estruturas topicalizadas ou clivadas (a única exceção são as clivadas do tipo *Que-Foco*). Foram excluídas também as ocorrências do pronome em uma coordenação, sob o escopo de um advérbio, enfáticas, os truncamentos, as primeiras ocorrências em repetições e as construções “(quase) cristalizadas”.

Dessa forma, as ocorrências de fato incluídas na análise multivariada totalizam 3433 dados, assim distribuídos: 1773 (52%) com a variante plena *você* e 1660 (48%) com a variante foneticamente reduzida *cê*. Em relação a informantes e documentadores, esses dados estão distribuídos do modo como mostra a tabela abaixo:

VARIANTES	GESOL-SP-2000 (36 INFs.)		GESOL-SP-2000 (9 Docs.)		TOTAL	
	Nº	%	Nº	%	Nº	%
<i>Você</i>	1280	51,7	493	51,5	1773	51,65
<i>Cê</i>	1195	48,3	465	48,5	1660	48,35
Total	2475	100	958	100	3433	100

Tabela 15: *Você* e *Cê* em contextos variáveis – informantes e documentadores (GESOL-SP-2000)

Em princípio, as percentagens de *você* e *cê* mostram-se muito semelhantes entre os documentadores e os informantes. Deve-se, contudo, atentar para o fato de que a maioria dos documentadores é constituída de mulheres da primeira faixa etária, ao passo que a amostra dos informantes traz um número equilibrado de homens e mulheres em três faixas etárias².

² Embora a amostra dos documentadores seja bastante desequilibrada, optou-se por considerar esses dados em algumas análises, devido à fala do documentador ser propícia ao aparecimento de determinadas construções – como é o caso das orações interrogativas e da referência específica ao interlocutor. Na maioria das análises sincrônicas, foram considerados “somente” 2475 dados – aqueles provenientes da fala de 36 informantes.

Diferentemente daquilo que se observou em outras comunidades de fala, tais como a mineira (na qual *cê* parece ser mais frequente do que *você/ocê*³), o emprego dessas duas variantes mostra-se bastante equilibrado entre os paulistanos. A tabela a seguir traz a distribuição geral das variantes *você/ocê/cê/(zero)* nesta pesquisa (dados dos informantes) e em estudos realizados em Minas Gerais e em Brasília.

VARIANTES	Belo Horizonte (MG)				Arcos (MG)		DF (Brasília)		Cidade de SP	
	Ramos (1997) ⁴		Peres (2006) ⁵		Gonçalves (2008) ⁶		Andrade (2004) ⁷		Nascimento – (Anos2000)	
	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
CÊ	189	55	1055	72,6	277	54	48	29	1195	48,2
OCÊ	41	12	56	3,9	120	24	9	5,5	5	0,2
VOCÊ	112	33	342	23,5	113	22	108	65,5	1280	51,6
Total	342	100	1453	100	510	100	165*	100	2480	100

*Total: 169 - 24 zeros = 169

Tabela 16: Distribuição de *você/ocê/cê* em diferentes comunidades de fala: BH, DF e SP

Entre estes estudos, apenas Andrade (2004) verifica uma baixa frequência da variante *cê* (25%) na fala de crianças brasilienses entre 10 e 14 anos de idade. Nos estudos realizados em Belo Horizonte, Ramos (1997) observa uma maior frequência de *cê* (55%) comparativamente a *ocê* (12%) e *você* (33%), da mesma forma que Peres (2006), em uma amostra maior da mesma comunidade, na qual foram incluídas crianças e adolescentes. Em outra comunidade de fala mineira (Arcos), Gonçalves (2008) também depreende uma frequência mais elevada de *cê* (54%), por oposição a *você* (22%) e *ocê* (24%). No presente estudo, a análise da

³ Cf. Ramos (1997); Peres (2006); Gonçalves (2008).

⁴ Ramos (1997) trabalha com uma amostra de 12 informantes adultos naturais de Belo Horizonte.

⁵ Peres (2006) considera em sua amostra a fala de 47 informantes de Belo Horizonte, divididos em cinco faixas etárias, entre as quais há crianças (de 8 a 11 anos de idade) e adolescentes (de 12 a 16 anos). Estes últimos empregam mais frequentemente a variante reduzida do pronome – o que explicaria a elevada frequência de *cê*.

⁶ O *corpus* analisado por Gonçalves (2008) compreende 40 informantes da região de Arcos (Minas Gerais) (20 da zona rural e 20 da zona urbana) em narrativas orais espontâneas. As faixas etárias compreende informantes entre 15 e 30 anos; 31 a 59 anos e 60 anos ou mais.

⁷ Andrade analisa a fala de crianças do Distrito federal (Brasília) entre 10 e 14 anos. A autora também analisa a fala da documentadora – não incluída aqui nos resultados aqui apresentados.

fala de indivíduos paulistanos adultos (acima de 20 anos) revela equilíbrio na distribuição das duas variantes.

Variáveis sociais⁸

A tabela a seguir traz os resultados obtidos para os grupos de fatores sexo/gênero, faixa etária e nível de escolaridade na fala de 36 informantes da amostra *GESOL-SP2000*⁹.

SEXO / GÊNERO	VOCÊ		CÊ			Total
	Nº	%	Nº	%	Peso	
Feminino	677	49,6	687	50,4	.52	1364
Masculino	603	54,3	508	45,7	.47	1111
Range: .05						
FAIXA ETÁRIA	Nº	%	Nº	%	Peso	Total
20-32	307	41,5	432	58,5	.59	739
35-45	445	62,1	272	37,9	.40	717
50 ou mais	528	51,8	491	48,2	.51	1019
Range: .19						
ESCOLARIDADE	Nº	%	Nº	%	Peso	Total
Fundamental/Médio	557	42,7	748	57,3	.59	1305
Superior	723	61,8	447	38,2	.39	1170
Range: .20						
Total	1280	51,7	1195	48,3		2475
<i>p</i> < 0.03 <i>ll.</i> = -1403.699 <i>input:</i> .445						

Tabela 17: Correlação de fatores sociais ao emprego de *você/cê* (Infs.-GESOL-SP-2000)¹⁰

As mulheres apresentam uma pequena tendência ao emprego da variante reduzida *cê* (.52), que é algo desfavorecida entre os homens (.47). Apesar disso, não se pode dizer que elas estariam liderando uma possível mudança nessa

⁸ Diferentemente dos grupos de fatores sociais, os interacionais não apresentam um número equilibrado de informantes em todos os perfis (v. Capítulo II). Uma análise conjunta desses grupos de fatores produziu resultados não confiáveis (detectados pela comparação entre as duas melhores rodadas estatísticas apresentadas pelo Golvarb X: estas, que deveriam ser idênticas, apresentavam pesos relativos e valores de *log likelihood* distintos). Em consequência, nas análises em que os grupos de fatores interacionais foram considerados, os sociais não o são (e vice-versa).

⁹ A variante colocada em foco (estabelecida como valor de aplicação) foi sempre *cê*.

¹⁰ Os resultados apresentados nesta seção advieram de análises em que, além de três grupos de fatores sociais, foram incluídos seis grupos de fatores linguísticos: referência do pronome; coalescência; paralelismo I; tonicidade da sílaba seguinte; tonicidade de sílaba antecedente; frequência do verbo – todos analisados em versões mais concisas do daquelas apresentadas no capítulo II. Todos os grupos de fatores foram selecionados pelo programa estatístico como estatisticamente relevantes.

comunidade de fala, pois os pesos relativos obtidos encontram-se muitos próximos ao ponto neutro, o que torna o *range*¹¹ muito pequeno.

A análise da faixa etária dos informantes revela que a forma *cê* é favorecida entre mais jovens (.59) e desfavorecida entre os falantes da faixa etária intermediária (.40); entre os mais velhos, *cê* não é nem favorecida nem desfavorecida (.51).

O gráfico a seguir indica que os mais jovens lideram o emprego da forma pronominal inovadora. Não se verifica, contudo, um padrão típico de mudança em progresso, uma vez que os mais velhos não desfavorecem o uso da variante *cê*, mas sim aqueles de uma faixa etária intermediária. O padrão curvilíneo deste gráfico pode ser explicado pela atuação do “mercado linguístico” no emprego das variáveis – cf. Paiva & Duarte (2003).

Os indivíduos de idade intermediária são aqueles que se encontram mais ativos e já estabelecidos no mercado de trabalho – por oposição aos mais jovens (muitas vezes ainda reticentes quanto à profissão que vão exercer) e aos mais velhos (mais próximos de deixar o mercado de trabalho, ou já aposentados). Nesse sentido, a variante reduzida *cê*, ainda que não seja uma forma socialmente estigmatizada, parece ser mais consistentemente evitada entre os informantes de segunda faixa etária, talvez por estar associada a um estilo de fala menos cuidado. Essa generalização é possível, ainda que possa haver indivíduos que estejam fora

¹¹ O *Range* corresponde à diferença entre os valores do maior e do menor pesos relativos obtidos para um determinado grupo de fatores (cf. Guy & Zilles, 2007).

do mercado profissional, entre aqueles cuja faixa etária é mediana. Além disso, o tipo de atividade profissional exercida pelo indivíduo, o assunto e o grau de formalidade da situação podem influenciar suas falas. Ainda que os assuntos das entrevistas sociolinguísticas tenham sido os mesmos para todos os informantes da amostra GESOL-SP-2000, parece ser natural que indivíduos mais velhos tenham mais histórias de vida a relatar do que indivíduos jovens. Assim, os informantes mais velhos, sobretudo as mulheres, demonstraram grande disponibilidade e disposição ao diálogo na ocasião da gravação de suas entrevistas e reportaram, mais vezes que os homens, situações relacionadas a assuntos pessoais e à sua infância – proporcionando um clima informal em torno da gravação das entrevistas e, possivelmente, favorecendo o emprego do pronome reduzido.

Tais considerações, contudo, carecem de investigações futuras. Um estudo que construa suas amostras de fala com o controle de variáveis como “profissão”, “assunto”, “formalidade da situação” e “estilos de fala em trechos de entrevistas” poderá formular conclusões mais precisas do que estas que o presente estudo pôde oferecer.

O cruzamento dos grupos sexo/gênero e faixa etária aponta uma distinção entre os falantes mais velhos no que diz respeito ao seu gênero e à frequência de uso da variante *cê*. Análises quantitativas separadas dos dados provenientes da fala dos homens e dos dados provenientes da fala das mulheres confirmam a distinção de gênero em relação às diferentes faixas etárias¹².

FAIXA ETÁRIA	CÊ					
	Mulheres			Homens		
	Nº / Total	%	p.r.	Nº / Total	%	p.r.
20-32	270 / 471	57,3	.58	162 / 268	60,4	.61
35-45	156 / 428	36,4	.38	116 / 289	40,1	.44
50 ou +	261 / 465	56,1	.54	230 / 554	41,5	.48
Total	687 / 1364	50	Range: .20	508 / 1111	46	Range: .17
Input: 0.483 l.l.= -768.005 p < 0.02						Input: 0.402 l.l.= -643.344 p = 0.03

Tabela 18: Emprego de CÊ conforme a Faixa etária do Informante e seu Sexo/Gênero (Inf-GESOL-SP-2000)

¹² Em ambas as análises, consideraram-se os grupos de fatores faixa etária; nível de escolaridade; referência do pronome (versão concisa); coalescência (versão concisa); paralelismo I (versão concisa); tonicidade da sílaba seguinte (versão concisa); tonicidade de sílaba antecedente (versão concisa); frequência do verbo. Todos os grupos de fatores foram selecionados como estatisticamente relevantes.

As mulheres da terceira faixa etária têm um comportamento linguístico semelhante às da primeira: ambas tendem a empregar a variante reduzida do pronome (pesos relativos .58 e .54, respectivamente). Os homens da terceira faixa etária, ao contrário, aproximam-se daqueles de faixa etária intermediária, desfavorecendo a variante reduzida do pronome (pesos relativos .44 e .48). Depreende-se, portanto, uma gradação etária, correlacionada ao gênero/sexo do falante – conforme ilustra o gráfico a seguir.

Quanto ao sexo/gênero associado à idade dos informantes, os resultados anteriores equiparam-se àqueles obtidos por Peres (2006) em Belo Horizonte. No entanto, se se comparar a variante *cê* e o pronome *tu*, considerando-os formas de “solidariedade” (Gonçalves, 2008), os resultados aqui depreendidos afasta-se daqueles obtidos por Lucca (2005) e Dias (2007) em Brasília. Nesses dois últimos estudos, o *tu* tende a ser empregado pelos homens mais jovens.

Com relação ao nível de escolarização, o emprego de *cê* é favorecido entre os falantes que cursaram até o ensino Fundamental ou Médio¹³ (.59) e

¹³ De 18 informantes, 16 estudaram da 4ª à 8ª séries do ensino Fundamental, 1 começou o primeiro ano do ensino Médio, mas não o concluiu e uma concluiu o ensino Médio, mas não deu continuidade aos estudos no nível superior. O comportamento linguístico dos dois informantes com ensino médio não difere dos demais com ensino Fundamental, razão pela qual foram mantidos na amostra.

desfavorecido entre aqueles que têm nível superior de escolaridade (.39) – tabela 17, p. 134).

Considerando-se que os informantes têm consciência de que se trata de uma entrevista para fins acadêmicos, é possível que, no sentido de preservar “sua face” de escolarizados, tais falantes adotem mais consistentemente um estilo “mais cuidado” na situação de entrevista, evitando, com maior frequência, o emprego de formas associadas a uma fala geralmente avaliada como “descuidada” ou “espontânea”. Da mesma forma que nas variáveis anteriores, este trabalho não procedeu a uma análise de percepção das variantes, nem pôde controlar certas variáveis que permitissem encaminhar conclusões referentes a estilos em trechos de entrevistas (a depender do assunto, da situação etc). Ainda assim, os resultados obtidos para escolaridade parecem confirmar a hipótese de que, embora *cê* não seja exatamente uma forma estigmatizada, “ainda” está associada a padrões que não são encorajados na escola.

Variáveis interacionais

A tabela seguinte traz as correlações entre o “tipo de relação entre os interlocutores” e o emprego das variantes *você* e *cê* na fala de 24 informantes.

INTERLOCUTORES (INFs)	VOCÊ		CÊ			Total
	Nº	%	Nº	%	p.r.	Nº
Desconhecidos (3)	142	37,9	233	62,1	.63	375
Parentes (6)	153	43,2	201	56,8	.56	354
Amigos (5)	208	55,5	167	44,5	.47	375
Amigos de amigos (3)	230	57,9	167	42,1	.46	397
Conhecidos/colegas (7)	255	64,7	139	35,3	.39	394
Subtotal (24)	988	52,1	907	47,9		1895*
Total Geral (36)	1280	51,7	1195	48,3		2475

Input: 0.449 ll.= - 1456.743 p < 0.02

Range: .24

* Exceto dados não disponíveis de 12 informantes.

Tabela 19: Emprego de *você/cê* e relação entre os interlocutores (Informantes-GESOL-SP-2000)¹⁴

¹⁴ Além do grupo de fatores “relação entre os interlocutores”, foram analisadas as seguintes variáveis independentes: sexo/genro dos interlocutores, referência do pronome; coalescência; paralelismo I; tonicidade da sílaba seguinte; tonicidade de sílaba antecedente; e frequência do verbo. Todos os grupos de fatores foram selecionados como estatisticamente relevantes.

Por um lado, quando o documentador e o informante têm algum grau de parentesco, o informante tende a empregar a variante reduzida do pronome (.56); quando são conhecidos/colegas, é desfavorecido o uso de *cê* – ou seja, “*você é a forma que tende a ser empregada nesse caso*”. Em se tratando de interlocutores que são amigos um do outro, ou que foram apresentados por um amigo em comum, observa-se um desfavorecimento muito pequeno ao *cê* (.47 e .46 – pesos bastante próximos do ponto neutro, evidenciando que o emprego de uma ou outra variante se aproxima de indiferente, nesses casos).

No entanto, contraditoriamente, justamente os informantes que não conheciam seus entrevistadores antes da entrevista apresentam maior tendência de emprego de *cê* (.63). A fim de compreender melhor esse fato, observe-se a tabela abaixo:

Relação entre os interlocutores	Informante	Perfil	Ano	CÊ/Total	%
PARENTE	Zoraide	F3SI	2003	23/54	42,6
	Roberta	F2FI	2008	25/71	35,2
	Clara	F3FI	2007	58/76	76,3
	Bruno	M1FI	2008	40/45	88,9
	Pedro	M1SI	2007	33/76	43,4
	Fabrício	M1SI	2007	22/32	68,8
AMIGO	Camila	F1SI	2006	5/37	13,5
	Cecília	F2SI	2007	13/76	17,1
	Denise	F3SI	2009	56/93	60,2
	Marta	F2FI	2008	63/120	52,5
	Gustavo	M1SI	2006	30/49	61,2
AMIGO DE AMIGO	Fabíola	F1SI	2007	53/64	82,8
	José	M3SI	2006	72/225	32,0
	Jéfferson	M2FI	2008	42/108	38,9
COLEGA / CONHECIDO	Fabiana	F1SI	2003	58/159	36,5
	Ingrid	F2SI	2007	15/41	36,6
	Silvia	F3SI	2006	25/58	43,1
	Rogério	M2SI	2007	5/8	62,5
	Cláudia	F1FI	2007	8/22	36,4
	Maria	F3FI	2008	12/61	19,7
	Daniel	M2FI	2008	16/45	35,6
DESCONHECIDO	Paloma	F1MI	2009	73/88	83,0
	Silmara	F3FI	2007	87/123	70,7
	Marco	M3FI	2007	73/164	44,5

Tabela 20: Distribuição dos informantes conforme sua relação com o interlocutor (GESOL-SP- 2000)

Dos três informantes que não se conheciam previamente à entrevista, dois apresentam alta frequência da variante reduzida: duas mulheres, uma da primeira faixa etária (Paloma) e outra da terceira faixa etária (Silmara) – que, como se verificou anteriormente, estão entre os perfis sociais que mais tendem a empregar a variante reduzida do pronome. Dessa forma, excluíram-se os dados desses três informantes em uma análise que permitisse melhor avaliação do impacto da variável “relação entre informante e documentador” no caso de variação *você/cê*:

INTERLOCUTORES	VOCÊ		CÊ			Total
	Nº	%	Nº	%	p.r.	Nº
Parente (6)	153	43,2	201	56,8	.59	354
Amigo (5)	208	55,5	167	44,5	.51	375
Amigo de amigo (3)	230	57,9	167	42,1	.48	397
Conhecido/colega (7)	255	64,7	139	35,3	.42	394
Subtotal (21)	846	55,7	674	44,3		1520*
Input: 0.419 l.l.= - 1460.716 p < 0.02					Range: .17	
* Exceto dados não disponíveis e dois informantes desconhecidos.						

Tabela 21: Emprego de *Você/Cê* conforme a relação do INF com o DOC (GESOL-SP-2000)¹⁵

Esses resultados indicam tendência de empregar *cê* quando há uma relação de parentesco entre os indivíduos. Nos demais casos, os pesos relativos obtidos aproximam-se do ponto neutro. Estes resultados aplicam-se tanto aos empregos indeterminados/genéricos quanto àqueles cuja referência é específica ao interlocutor. Considerando-se a possibilidade de que sejam estes últimos aqueles que tendam a ser influenciados mais nitidamente pela variável “relação entre interlocutores”, procedeu-se a uma análise em que se incluísem apenas os dados de referência específica, nas falas dos 9 documentadores e dos 24 informantes anteriores¹⁶.

¹⁵ Nesta análise, foram considerados os grupos de fatores: referência do pronome; coalescência; paralelismo I; tonicidade da sílaba seguinte; tonicidade de sílaba antecedente; e frequência do verbo – todos demonstraram significância estatística.

¹⁶ Mantiveram-se excluídos os dados provenientes dos três informantes que não conheciam os documentadores antes da gravação. Dentre os documentadores que não conheciam os informantes previamente à entrevista, nenhum é paulistano – razão pela qual não tiveram seus dados incluídos na análise.

RELAÇÃO ENTRE OS INTERLOCUTORES	CÊ (REF Específica)					
	Informantes (21)			Documentadores (9)		
	Nº	%	p.r.	Nº	%	p.r.
Parente (6+2)	41/72	56,9	.57	145/289	50,2	.56
Amigo (5+3)	51/90	56,7	.55	209/303	69,0	.73
Amigo de amigo (3+1)	25/48	52,1	.51	78/138	56,5	.63
Conhecido/colega (7+3)	33/76	43,4	.38	35/271	12,9	.16
Subtotal (21+9)	150/286*	52,4	Range: .19	467/1001	46,7	Range: .57
Input: 0.554 ll.= - 318.449 p < 0.05						Input: 0.439 ll.= - 570.407 p < 0.02
* Exceto dados não disponíveis de 12 informantes e 3 informantes sem relação prévia com o DOC.						

Tabela 22: Uso de CÊ específico e relação entre os interlocutores (Infs e Docs – GESOL-SP-2000)¹⁷

Nos dados dos informantes, apenas para “parentes” e “amigos” observa-se tendência de emprego de *cê*; de fato, amigos e parentes se assemelham nos resultados e diferem de “colegas” e “conhecidos”.

De uma maneira geral, entre os informantes, a variante reduzida é favorecida quando há uma relação de parentesco (.57) ou de amizade (.55); entre os documentadores, *cê* também tende a ser empregado quando há uma relação de parentesco entre os interlocutores (.56); e é ainda mais desfavorecida entre os documentadores (.16) do que entre os informantes (.38), quando a relação entre os interlocutores é a menos próxima.

Portanto, o grau de intimidade entre os interlocutores condiciona o emprego das variantes em estudo. De um lado, situam-se os parentes e amigos; de outro, amigos de amigos e colegas/conhecidos. Os resultados da tabela 23 a seguir indicam que o grau de intimidade entre os informantes atua mais fortemente na subamostra de documentadores (.63) do que na de informantes (.56).

¹⁷ Foram realizadas duas análises: uma apenas com os dados dos informantes; outra, apenas com os dados dos documentadores. Além da variável “relação entre os interlocutores”, foram incluídos os grupos: “tipo de oração” (interrogativa; declarativa); “frequência do verbo” (versão concisa) e “paralelismo I” (versão concisa), dentre os quais apenas a frequência do verbo não se mostrou estatisticamente relevante na análise multivariada dos dados dos documentadores.

Grau de intimidade entre os Interlocutores	CÊ (REF Específica)					
	Informantes (21)			Documentadores (9)		
	Nº	%	p.r.	Nº	%	p.r.
+ próximo (11+5)	92/162	56,8	.56	354/592	59,8	.63
- próximo (10+4)	58/124	46,8	.43	113/409	27,6	.32
Subtotal (21+9)	150/286	52,4	Range: .13	467/1001	46,7	Range: .31
Input: 0.554 ll.= - 319.398 p < 0.02						Input: 0.458 ll.= - 619.506 p = 0.000
* Exceto dados não disponíveis de 12 informantes; exceto 3 dados de informantes que não se conheciam.						

Tabela 23: Uso de CÊ com referência específica de acordo com o grau de intimidade entre os interlocutores (Infs e Docs – GESOL-SP-2000)¹⁸

Finalmente, numa análise em que se recodificaram os dados de acordo a interrelação de interlocutores baseada no seu sexo/gênero, evidencia-se que somente entre mulheres é favorecida a variante *cê* (.54)¹⁹.

SEXO/GÊNERO DOC-INF	VOCÊ		CÊ			Total
	Nº	%	Nº	%	p.r.	Nº
F-F	473	49,0	492	51,0	.54	965
M-M	397	55,2	322	44,8	.46	719
F-M	206	52,6	186	47,4	.48	392
M-F	204	51,1	195	48,9	.50	399
Total	1280	51,7	1195	48,3		2475
p < 0.02 ll. = - 1615.270 input: 0.446						Range: .08

Tabela 24: *Você/Cê* em relação ao sexo/gênero dos interlocutores (Infs – GESOL-SP-2000)²⁰

Nota-se, todavia, que a simetria do sexo/gênero dos interlocutores exerce um papel extremamente reduzido: comparando-se estes resultados àqueles da tabela 17 (p. 134), vê-se que há um ligeiro favorecimento das mulheres empregarem a variante reduzida do pronome (.52), este índice eleva-se um pouco mais quando a relação é simétrica quanto ao gênero (.54) e reduz-se um pouco quando a relação é assimétrica (.50). A distinção, portanto, é de apenas 0.02. No

¹⁸ Mantiveram-se os mesmos grupos de fatores das análises expressas nas tabelas 21 e 22.

¹⁹ Esse resultado é interessante, pois em Brasília são os homens que mais favorecem o uso de *tu* em interações simétricas (Lucca, 2005; Dias, 2007), e ambas as formas (*tu* e *cê*) são consideradas variantes que indicam maior “informalidade” e “solidariedade” para com o interlocutor (v. Gonçalves, 2008).

²⁰ Além da variável desta variável, foram incluídas também na análise quantitativa: “referência do pronome”, “coalescência do pronome ao verbo”, “Tonicidade da sílaba seguinte”, “Tonicidade da sílaba antecedente”, “frequência do verbo” e “paralelismo I”. Todos se mostraram estatisticamente significativos.

caso dos homens, viu-se na tabela 17 (p. 134) um pequeno desfavorecimento ao uso de *cê* (.47). Contrariamente ao observado nas mulheres, esse índice cai 0.01 em relações simétricas (.46) e sobe 0.01 nas relações de assimetria quanto ao gênero (.48).

Variáveis linguísticas

A tabela 25 a seguir sintetiza os principais resultados obtidos para as variáveis de natureza linguística. Todos os grupos de fatores foram selecionados pelo Goldvarb como estatisticamente significativos, com exceção da “flexão do pronome” (singular ou plural) e do “tipo de discurso” (reportado *versus* não-reportado), cujos resultados aqui apresentados são provenientes de uma análise *One Level*.

CÊ							
REFERÊNCIA	Nº/Total*	%	P.R.	PARALELISMO I	Nº/Total	%	P.R.
P2 (Específica)	160/249	64	.70	Anterior CÊ	383/489	78	.78
P1 (1ª p. sg./pl. excl.)	300/552	54	.56	Isolado/1º	625/1341	46,5	.47
P4 (1ª p. pl. incl.)	378/778	48,5	.46	Anterior VOCÊ	187/645	29	.33
Híbrido	82/212	38,5	.42	Total	1195/2475	48	
P3 /P6 (3ª p. sg./pl.)	74/213	34,5	.41				
Expletivos	13/65	20	.22				
Subtotal*	1007/2069	Range: .48					
* Exceto: 2ª p. pl. (52); disc. report.(274); e “outros” (80)				* Exceto pausas e hesitações.			
TONICIDADE ANTECEDENTE	Nº/Total	%	P.R.	TONICIDADE SEGUINTE	Nº/Total	%	P.R.
Tônica	228/431	67	.61	Tônica	951/1769	54	.53
Átona	527/1286	41	.46	Átona	233/632	37	.41
Subtotal*	755/1717	44		Subtotal*	1184/2401	49	
		Range: .15				Range: .12	
* Exceto pausas e hesitações.				* Exceto pausas e hesitações.			
FREQUÊNCIA VERBO	Nº/T	%	P.R.	COALESCÊNCIA	Nº/T	%	P.R.
+ de 6 %	614/1073	57	.58	Coalescente	990/1968	50	.51
Até 6 %	581/1402	41,5	.44	Interpolação	187/406	46	.45
		Range: .14				Range: .06	
TOTAL Geral	1195/2475	48					
p < 0.05 ll. = - 1403.699 input: 0.445							

Tabela 25: Correlações de natureza linguística ao emprego de *CÊ* (Infs.-GESOL-2000)²¹

²¹ Nas análises em se obtiveram esses resultados foram incluídos também os grupos de fatores sociais Sexo/Gênero, Faixa etária e Nível de escolarização. A distinção entre os resultados dos grupos de fatores que integraram mais de uma análise foi muito reduzida (com uma diferença de aproximadamente 0.01 em seus pesos relativos).

PRESENÇA DE /SE/	Nº/T	%	P.R.	FUNÇÃO SINTÁTICA	Nº/T	%	P.R.
Não	1146/2229	51,5	.54	Sujeito	1173/2387	49	.51
Sim	49/246	20	.20	dupla função	22/88	25	.27
Range: .34				Range: .24			
TIPO DE ORAÇÃO	Nº/T	%	P.R.	PARALELISMO II	Nº/T	%	P.R.
Interrogativas qu-que	38/54	70	.66	Sim	203/364	55,5	.63
Outras interrogativas	150/230	65	.59	Não	992/2111	47	.48
Declarativas	1007/2191	46	.49	Range: .15			
Range: .17							
FLEXÃO DO PRONOME	Nº/T	%	P.R.	TIPO DE DISCURSO	Nº/T	%	P.R.
Singular	1175/2423		(.50)	Reportado	141/275	51	(.53)
Plural	20/52		(.40)	Não-report.	1054/2200	48	(.50)
(Análise One Level)				(One Level)			
Range: .10				Range: .10			
TOTAL Geral	1195/2475	48					

Tabela 26: Emprego de CÊ conforme a atuação de Fatores linguísticos II (Infs.-GESOL-2000)²²

“Referência do pronome” e “paralelismo I” foram os grupos de fatores que demonstraram maior correlação com o emprego das variantes *você* e *cê* (ranges .48 e .45, respectivamente) – tabela 25. Outros grupos de fatores, como a “flexão do pronome” e o “tipo de discurso (tabela 26) provavelmente não foram selecionados em análises multivariadas em razão do desequilíbrio no número de dados entre os fatores que compõem cada uma destas variáveis independentes.

Função Sintática do pronome

Considerando-se apenas os contextos incluídos no envelope de variação (sujeito e dupla função), verifica-se que esta última desfavorece bastante o emprego da variante reduzida do pronome (.27)²³ – tabela 26 acima. Esse fato provavelmente está associado à presença de uma preposição antecedendo o pronome na maioria dos dados (79 casos em 88 ocorrências do pronome) em

²² Os resultados apresentados na tabela 25 referem-se a seis análises distintas, nas quais se incluíram os grupos de fatores sociais (sexo/gênero; faixa etária e escolaridade) e um desses grupo de fatores linguístico de cada vez.

²³ Apesar de desfavorecerem bastante o uso da variante reduzida do pronome, a exclusão dos dados com dupla função sintática não revelou diferença significativa; decidiu-se, então, mantê-los nas análises.

que este apresenta uma dupla função sintática (*o jeito de cê falar; o motivo de você tá comprando*). Nas nove ocorrências restantes – ou seja, naquelas em que o pronome exerce uma dupla função sintática, mas não é antecedido por uma preposição (e.g. *eu vou manda(r) cês morar junto com o pai...; o cara tira uma radiografia... manda cê pôr tala*) – a variante reduzida é empregada em quatro ocorrências do pronome (44%). O uso reduzido da variante reduzida nesses contextos corroboram as hipóteses de cliticização de *cê(s)* e sua especialização com a função sintática de sujeito.

Referência do pronome

Uma análise da referência do pronome *você/cê* considerando-se apenas dois fatores revela que a variante reduzida tende a ocorrer quando a referência é específica ao interlocutor (.62), e é ligeiramente desfavorecida quando sua referência é genérica/ indeterminada (.47).

REFERÊNCIA	VOCÊ		CÊ			Total
	Nº	%	Nº	%	p.r.	Nº
Específica	198	40,8	287	59,2	.62	485
Genérica	987	53,7	851	46,3	.47	1838
Total	1185	51,0	1138	49,0		2323*
Total geral	1280	51,7	1195	48,3		2475
Input: 0.443 l.l.= -1433.902 p = 0.05						Range: .15
* Exceto 80 dados cuja referência foi classificada como “Outros” (v. p. 116) e 2ª p. pl.						

Tabela 27: *Você/Cê* de acordo com a sua referência (específica x genérica) – (Infs – GESOL-SP-2000)²⁴

O padrão observado confirma o que outros estudos depreenderam em outras comunidades de fala, como a mineira (Alves, 1998; Peres, 2006; Gonçalves, 2008). No entanto, contraria a hipótese de Ramos (1997), para quem a variante reduzida do pronome tenderia a ser empregada com referência genérica.

²⁴ O número de ocorrências do pronome com referência específica é superior àquele apresentado na tabela 24 (p. 142) porque os dados de discursos reportados foram excluídos apenas naquela análise. Os dados com referência genérica (tabela 26), por sua vez, abarcam os fatores P1, P4, P3, híbrido e expletivo apresentados na tabela 24 (p. 142). Além da “referência” também foram considerados nesta análise os grupos de fatores sociais (sexo/gênero, faixa etária e escolaridade) e os linguísticos “Coalescência do pronome ao verbo”, “Paralelismo I”, “Frequência do verbo”, “Tonicidade seguinte” e “Tonicidade antecedente” – todos em versões concisas, a fim de se manter a ortogonalidade entre eles.

Em entrevistas sociolinguísticas, são menos frequentes os casos em que o informante se dirige diretamente ao documentador – daí o número de dados com referência específica corresponder a apenas 21% (485/2475) dos dados analisados no presente estudo (tabela 27). Adicionalmente, esses resultados incluem tanto ocorrências do pronome em discurso reportado (ou discurso relatado) quanto aquelas de discurso não-reportado (tabela 28). A grande maioria dos dados de discursos reportados ocorre com referência específica ao interlocutor (255/274).

DISCURSO	CÊ				
	REF Específica		REF Genérica		Total
	Nº / Total	%	Nº / Total	%	
Reportado	136/255	53,0	4/19	21,0	140 / 274
Não-reportado	171/282	61,0	847/1819	47,0	1018 / 2101
Total	307/537	57,0	851/1838	46,0	1158 / 2375*

Tabela 28: Emprego de *CÊ* conforme o tipo de discurso e a referência do pronome (Infs-GESOL-SP-2000)²⁵

Essa distribuição dos dados propicia não-ortogonalidade entre os grupos de fatores “referência do pronome” e “tipo de discurso”. Sendo assim, a análise desta da variável “Tipo de discurso” deve aplicar-se apenas aos usos cuja referência é específica. Na tabela anterior, vê-se que o emprego da variante reduzida *cê* é menos frequente nos casos de discursos reportados (53%) do que nos discursos não-reportados (61%), quando se trata de referência específica. A fim de se manter maior simetria em relação aos usos específicos e genéricos desse pronome, excluíram-se os dados de discursos reportados da análise da “referência”, conforme mostra a tabela 29. Nela, verifica-se um aumento do peso relativo da variante *cê* com referência específica – de .62 a .70 – em relação à tabela 28 anterior.

²⁵ Exceto 80 ocorrências cuja referência não pode ser depreendida com precisão, aqui denominados “outros”.

REFERÊNCIA	VOCÊ		CÊ			Total
	Nº	%	Nº	%	p.r.	Nº
Específica	89	35,7	160	64,3	.70	249
Genérica	973	53,5	847	46,5	.47	1820
Subtotal	1062	51,3	1007	48,7		2069*
Total	1280	51,7	1195	48,3		2475

Input: 0.447 ll.= -1429.323 p < 0.02

Range: .23

* Exceto: "Outros" (80), 2ª p. pl. (52) e discursos reportados (274).

Tabela 29: Referência do pronome em discursos não-reportados (oposição binária) (Infs.-GESOL-2000)²⁶

Com objetivo de verificar a relação entre “perda de material fônico” (erosão) e “perda de conteúdo semântico” (*bleaching*), as ocorrências com referência genérica foram subdivididas entre as indeterminadas de fato e aquelas que integram construções existenciais com o verbo “ter” (que, supostamente, não têm referência e funcionam como expletivos, segundo Vitral & Ramos (2006 [1999]) e Duarte (2003)).

Os resultados demonstram que não se pode estabelecer uma relação direta e indiscutível entre processos de “erosão fonética” e *bleaching*: a variante reduzida do pronome tende a ser mais empregada quando a referência é específica ao interlocutor (.70) e é fortemente desfavorecida nos casos considerados expletivos (.23). Além disso, de um ponto de vista quantitativo, os empregos supostamente expletivos demonstram um comportamento distinto dos usos genéricos.

REFERÊNCIA	VOCÊ		CÊ			Total
	Nº	%	Nº	%	p.r.	Nº
Específica	89	35,7	160	64,3	.70	249
Genérica	921	52,5	834	47,5	.48	1755
Expletivo	52	80,0	13	20,0	.23	65
Subtotal	1062	51,3	1007	48,7		2069*
Total	1280	51,7	1195	48,3		2475

Input: 0.445 ll.= -1423.353 p = 0.05

Range: .39

* Exceto: "Outros" (80), 2ª p. pl. (52) e discursos reportados (274).

Tabela 30: Correlação da referência ao emprego de *você/cê* (3 fatores) – (Infs – GESOL-SP-2000)

²⁶ Nas análises apresentadas nas tabelas 29 e 30, foram considerados os grupos de fatores sociais (sexo/gênero, faixa etária e escolaridade) e os linguísticos “Coalescência do pronome ao verbo”, “Paralelismo I”, “Frequência do verbo”, “Tonicidade seguinte” e “Tonicidade antecedente” – todos em versões concisas.

As ocorrências do pronome cuja referência é interpretada como genérica/indeterminada podem ser subdivididas, de acordo com as discussões desenvolvidas no capítulo III, a partir de pistas referenciais presentes no discurso ou a partir de inferências pragmáticas. Esses casos foram detalhadamente analisados neste trabalho, a fim de constatar correlação entre eles e o emprego das variantes *você* e *cê*. A tabela 31 a seguir e o gráfico 3 trazem os resultados dessa análise.

REFERÊNCIA	VOCÊ		CÊ			Total
	Nº	%	Nº	%	p.r.	Nº
P2 (específica)	89	35,7	160	64,3	.70	249
P1 'ocultamento do eu'	29	31,2	64	68,8	.65	93
P1 'expandido'	105	42,2	144	57,8	.55	249
P4 'exclusivo'	118	56,2	92	43,8	.54	210
P4 'incl.' [+definido]	274	49,6	278	50,4	.47	552
P4 'incl.' [-definido]	126	55,8	100	44,2	.43	226
Híbridos/ambíguos	130	61,3	82	38,7	.42	212
P3/P6 (3ª pessoa)	139	65,3	74	34,7	.42	213
Expletivos (?)	52	80,0	13	20,0	.23	65
Subtotal	1062	51,3	1007	48,7		2069*
Total	1280	51,7	1195	48,3		2475

Input: 0.446 ll: -1412.819 $p < 0.05$

Range: .47

* Exceto "Outros"; 2ª p. pl. e discurso reportado.

Tabela 31: Emprego de *você/cê* de acordo com sua referência no discurso (Infs.-GESOL-SP-2000) (Análise expandida da referência I)²⁷

²⁷ Para esta análise consideraram-se os mesmos grupos de fatores das análises precedentes (v. nota 26, p. 147).

Observa-se aqui a mesma polaridade apreendida na análise com três fatores: de um lado, os empregos com referência específica ao interlocutor favorecem o uso da variante reduzida (.70); de outro lado, os empregos (supostamente) expletivos desfavorecem-na (.23). Já os distintos empregos genéricos do pronome *você/cê* podem ser agrupados de acordo com fatos de natureza qualitativa e quantitativa:

- P1 e P4 ‘exclusivo’ favorecem o emprego da variante reduzida *cê*, sobretudo quando o emprego da segunda pessoa em lugar da primeira (embreagem actancial) visa a preservar a face do locutor (P1 ‘oculto’);
- P4 ‘inclusivo’ desfavorecem um pouco o emprego da variante reduzida do pronome; tal desfavorecimento é um pouco maior quando a referência genérica engloba número [-definido] de indivíduos (.43) do que nos casos em que remete a um grupo [+ definido] (.47);
- P3 / P6 e Híbridos/Ambíguos: desfavorecem o emprego da variante reduzida do pronome (.42).

Amalgamando-se os fatores semelhantes quantitativa e qualitativamente; mantendo-se separados os empregos denominados P3/ P6 e os Híbridos (por suas distinções qualitativas), demonstra-se que, quando a referência do pronome se situa no âmbito da enunciação (1ª e 2ª pessoas = P1 e P2), é maior a tendência de se empregar a variante reduzida (tabela 32)²⁸.

REFERÊNCIA	VOCÊ		CÊ			Total
	Nº	%	Nº	%	p.r.	Nº
P2 (específica)	89	35,7	160	64,3	.70	249
P1 (1ª p. sg/pl."excl.)	252	45,7	300	54,3	.56	552
P4 ('inclusivo')	400	51,4	378	48,6	.46	778
Híbridos/ambíguos	130	61,3	82	38,7	.42	212
P3/P6 (3ª pessoa)	139	65,3	74	34,7	.42	213
Expletivos (?)	52	80,0	13	20,0	.23	65
Total	1062	51,3	1007	48,7		2069*
Total Geral	1280	51,7	1195	48,3		2475

Input: 0.445 l.l.= -1414.515 p < 0.05

Range: .47

* Exceto: "Outros" (80), 2ª p. pl. (52) e discurso reportado (274).

Tabela 32: Emprego de *você/cê* de acordo com sua referência no discurso II (Inf.-GESOL-SP-2000)

²⁸ Para esta análise consideraram-se as mesmas variáveis das análises precedentes (v. nota 26, p. 147).

Contrariamente ao que se previa, os empregos híbridos não se mostram intermediários entre os usos específicos e genéricos, mas aproximam-se dos empregos do pronome cujas “pistas referenciais” ou inferências pragmáticas remetem a uma terceira pessoa. Como explicar este resultado?

De acordo com a discussão desenvolvida no capítulo III, os empregos denominados híbridos são aqueles que remetem simultaneamente ao interlocutor e às pessoas em geral. Em alguns casos, tais usos são depreendidos pelo emprego de expressões que indicam situações hipotéticas, nas quais o interlocutor (mas não apenas ele) poderia se inserir. Dentre as expressões de caráter hipotético, destacam-se duas: “por exemplo” e a conjunção “se”. Uma das hipóteses aventadas foi a de que haveria desfavorecimento ao uso *cê*, quando elementos semelhantes (caso da conjunção *se*) ocorressem adjacentes ao pronome. O cruzamento da referência com o grupo “presença de /SE/” revela que os dados em que /SE/ aparece adjacente ao pronome concentram dois tipos de referência: P4 ‘inclusiva’ e híbrida. Nestes, a presença de /SE/ faz baixar consideravelmente a frequência da variante reduzida no *corpus* (tabela 33).

REFERÊNCIA	CÊ				
	Sem /SE/		Com /SE/		TOTAL
	Nº / Total	%	Nº / Total	%	
P2 (específico)	150/227	66	10/22	45	160/249
P1 ‘oculto’	63/90	70	1/3	33	64/93
P1 ‘expandido’	140/242	58	4/7	57	144/249
P4 ‘exclusivo’	89/195	46	3/15	20	92/210
P4 ‘inclusivo’	365/703	52	13/75	17	378/778
P3/P6 (3ª pess.)	73/201	36	1/12	8	74/213
Híbridos/Ambíguos	70/136	51	12/76	16	82/212
Expletivos	13/65	20	0/0	0	13/65
Total	963/1859	52	44/210	21	1007/2069

Tabela 33: Emprego de *CÊ* conforme a sua referência e a presença de constituintes /SE/ adjacentes ao pronome (InfS.-GESOL-SP-2000)

Essa restrição fonética poderia estar se sobrepondo aos resultados obtidos para a referência²⁹. Com vistas a eliminar tal possibilidade, optou-se por excluir também os dados com /SE/ adjacente ao pronome. Os resultados obtidos confirmam a suspeita: o peso relativo dos usos híbridos aumenta, posicionando-se entre os usos P2 e P4, sendo superados apenas pelos usos genéricos do tipo P1.

REFERÊNCIA	VOCÊ		CÊ			Total
	Nº	%	Nº	%	p.r.	Nº
P2 (específica)	77	33,9	150	66,1	.69	227
P1 (1ª p. sg/pl.) + P4 excl.	235	44,6	292	55,4	.55	527
Híbridos/ambíguos	66	48,5	70	51,5	.50	136
P4 ('inclusivo')	338	48,1	365	51,9	.46	703
P3/P6 (3ª pessoa)	128	63,7	73	36,3	.41	201
Expletivos (?)	52	80,0	13	20,0	.21	65
Subtotal	896	48,2	963	51,8		1859
Total	1083	48,6	1146	51,4		2229*

Input: 0.492 ll.= -1250.783 p < 0.05

Range: .48

Exceto: "Outros", 2ª p. pl. e discurso reportado.

*Exceto dados com /SE/ adjacentes ao pronome (2475 - 246 = 2229).

Tabela 34: Emprego de *você/cê* de acordo com sua referência no discurso (Infs.-GESOL-SP-2000) (Análise expandida da referência II)³⁰

Dessa forma, os resultados obtidos para a referência até aqui indicam que:

- não se observa correspondência entre “perda de substância fonológica” e “perda de conteúdo semântico”;
- quando a referência do pronome se encontra no âmbito da enunciação, remetendo direta ou indiretamente às pessoas presentes na situação enunciativa, é maior a tendência de emprego da variante reduzida *cê*; ao contrário, quando há referência no discurso a uma terceira pessoa – uma *não-pessoa* (Benveniste, 1991) – o uso dessa variante é desfavorecido;

²⁹ Parte-se do pressuposto de que alguns grupos de fatores, por apresentarem uma forte correlação ao fenômeno investigado, acabariam “ofuscando” a atuação de outros grupos de fatores. Alguns estudos têm demonstrado, por exemplo, que determinadas variáveis independentes somente apresentam uma significância estatística ou esta se torna mais nítida quando se elimina a atuação do paralelismo (considerando-se apenas as ocorrências isoladas e primeiras de uma série). Propõe-se aqui um efeito semelhante no caso da variável “Presença de /SE/”.

³⁰ Para esta análise consideraram-se os mesmos grupos de fatores das análises precedentes (v. nota 26, p. 147).

- a referência às pessoas em geral desfavorece um pouco o emprego da variante reduzida (.46), situando-se entre os empregos com uma referência híbrida (.50) e aqueles que remetem a uma terceira pessoa (.41).

Tais resultados podem ser assim esquematizados:

P2 > P1 / P4 ‘exclusivo’ > Híbrido > P4 ‘inclusivo’ > P3 / P6 > Expletivo [± preservação da face] [± definido]	
[+ exofórico] ‘EU-TU/VOCÊ’	[- exofórico] ‘ELE’
[+ referencial]	[- referencial]
+	-
CÊ	

Quadro 21: Uso de *você/cê* conforme sua referência no discurso

De acordo com o esquema proposto, *cê* tende a ser empregado nos casos mais à esquerda, e desfavorecido quanto mais se avança à sua direita. Os empregos P2 (específico), P1 (1ª p. sg.) e P4 ‘exclusivo’ favorecem o uso da variante reduzida do pronome por portarem os traços [+ referencial] e [+ exofórico], situando-se no âmbito da enunciação. Os usos híbridos mostraram-se intermediários com relação ao uso de *cê*, situando-se entre os usos específicos e a maioria dos usos genéricos (exceto aqueles que remetem ao próprio locutor). A referência P4 ‘inclusiva’, por sua vez, refere-se a “qualquer pessoa”, “as pessoas em geral”, “todo mundo”, englobando o locutor e o interlocutor em sua referência [+ exofórico], embora tenham um caráter mais abrangente que os empregos anteriores. Quando há referência a uma terceira pessoa [-exofórico], há desfavorecimento do emprego de *cê*, pois se trata de alguém ou um grupo externo à situação enunciativa. Os casos expletivos – que supostamente portam o traço [-referencial] – desfavorecem ainda mais a variante reduzida do pronome.

Uma análise separada dos dados com referência genérica reflete o mesmo padrão observado nas análises anteriores: os empregos que remetem ao âmbito da enunciação favorecem o uso de *cê*; aqueles que remetem a uma terceira pessoa

(P3) desfavorecem-no. Os dados de referência híbrida e aqueles que apontam para as pessoas de uma maneira geral (P4 ‘inclusivo’) – em que não há nítida oposição “pessoa” *versus* “não-pessoa” – situam-se próximos ao ponto neutro (.52 e .47).

REFERÊNCIA (Genérica)	CÊ		
	Nº/Total	%	Peso
P1 ‘oculto’	63/90	70,0	.65
P4 ‘exclusivo’	89/195	45,6	.56
P1 ‘expandido’	140/242	57,9	.53
Híbrido	70/136	51,5	.52
P4 ‘inclusivo’ [+ definido]	265/486	54,5	.48
P4 ‘inclusivo’ [- definido]	100/217	46,1	.44
P3 /P6 (3ª p. sg./pl.)	73/201	36,3	.42
Subtotal	800/1567	51,1	Range: .23
Input: 0.477 l.l. = -1270.917 p < 0.05			
REFERÊNCIA (Genérica)	Nº/Total	%	Peso
P1 + P4 ‘exclusivo’	292/527	55,4	.56
Híbrido	70/136	51,5	.52
P4 ‘inclusivo’ [± definido]	365/703	51,9	.47
P3 /P6 (3ª p. sg./pl.)	73/201	36,3	.42
Subtotal	800/1567	51,1	Range: .14
Input: 0.477 l.l. = -1272.631 p < 0.05			

Tabela 35: Usos genéricos de CÊ em relação à sua referenciação no discurso (Inf.-GESOL-SP-2000)³¹

As análises anteriores demonstram a importância de um detalhamento na análise da referência em sua correlação ao uso de *você* e *cê*. A fim de justificar estatisticamente uma análise desta natureza, por oposição a análises com dois ou três fatores nesta variável independente, comparam-se os resultados obtidos em análises idênticas mais expandidas e mais concisas da referência. A tabela abaixo apresenta estes resultados. Logo em seguida, apresentam-se os testes de *qui-quadrado*.

³¹ A fim de comparar esses resultados aos anteriores, mantiveram-se na análise os mesmos grupos de fatores investigados nas análises precedentes: “sexo/gênero”, “faixa etária”, “escolaridade”, “coalescência do pronome ao verbo”, “paralelismo I”, “frequência do verbo seguinte”, “Tonicidade antecedente” e “Tonicidade seguinte” – todos em versões concisas.

9 fatores	REFERÊNCIA (Análise A)	CÊ		
		Nº/Total	%	Peso
	P2 (Específica)	150/227	66,1	.69
	P1 ‘oculto’	63/90	70,0	.64
	P1 ‘expandido’	140/242	57,9	.52
	P4 ‘exclusivo’	89/195	45,6	.54
	Híbrido	70/136	51,5	.50
	P4 ‘inclusivo’ [+ definido]	265/486	54,5	.47
	P4 ‘inclusivo’ [- definido]	100/217	46,1	.43
	P3 /P6 (3ª p. sg./pl.)	73/201	36,3	.41
	Expletivos	13/65	20,0	.21
	Subtotal	963/1859	51,8	Range: .48
Input: 0.492 l.l. = - 1248.900 p < 0.05				
6 fatores	REFERÊNCIA (Análise B)	Nº/Total	%	Peso
	P2 (Específica)	150/227	66,1	.69
	P1 + P4 ‘exclusivo’	292/527	55,4	.55
	Híbrido	70/136	51,5	.50
	P4 ‘inclusivo’ [± definido]	365/703	51,9	.46
	P3 /P6 (3ª p. sg./pl.)	73/201	36,3	.41
	Expletivos	13/65	20,0	.21
	Subtotal	963/1859	51,8	Range: .48
	Input: 0.492 l.l. = - 1250.783 p < 0.05			
3 fatores	REFERÊNCIA (Análise C)	Nº/Total	%	Peso
	Específica	150/227	66,1	.69
	Genérica	800/1567	51,1	.48
	Expletivos	13/65	20,0	.21
	Subtotal	963/1859	51,8	Range: .48
	Input: 0.492 l.l. = - 1256.695 p < 0.05			
2 fatores	REFERÊNCIA (Análise D)	Nº/Total	%	Peso
	Específica	150/227	66,1	.69
	Genérica (+ Expletivos?)	813/1632	49,8	.47
	Subtotal	963/1859*	51,8	Range: .22
	Input: 0.494 l.l. = - 1264.394 p < 0.05			
* Exceto: 2ª p. pl. , “outros”, discursos reportados e dados com/SE/ adjacentes ao pronome.				

Tabela 36: Diferentes abordagens da referência do *você/cê* (Infs – GESOL-SP-2000)³²

³² Para estas análises, mantiveram-se os mesmos grupos de fatores descritos na nota 31.

Teste de *qui-quadrado*:

Análise A: $\log \text{likelihood} = -1248.900$ (9 fatores)

Análise B: $\log \text{likelihood} = -1250.783$ (6 fatores)

$$(1250.783) - (1248.900) = 1,883 \times 2 = \mathbf{3,766} (= \text{qui-quadrado})$$

Graus de liberdade³³: $9 - 6 = \mathbf{3 \text{ graus de liberdade}}$

Valor esperado para uma análise significativa: $p = 0,05 \rightarrow 7,815 (> 3,766)$

Portanto: O *qui-quadrado* obtido equivale a uma significância estatística maior que 0.05 (o *qui-quadrado* de $p = 0,05$, considerando-se 3 graus de liberdade, é de **7,815** – um valor mais elevado do que **3,766**). Portanto, a diferença da análise A (que opera um número mais elevado de distinções quanto à referência do pronome) em relação à análise B (que diminui as distinções de 9 para 6, com base em semelhanças qualitativas e quantitativas entre elas) **NÃO** é significativa.

Resultado: A **análise B** (6 distinções) é melhor de um ponto de vista estatístico do que a análise A (9 fatores).

Teste de *qui-quadrado*:

Análise B: $\log \text{likelihood} = -1250.783$ (6 fatores)

Análise C: $\log \text{likelihood} = -1256.695$ (3 fatores)

$$(1256.695) - (1250.783) = 5,912 \times 2 = \mathbf{11,824} (= \text{qui-quadrado})$$

Graus de liberdade: $6 - 3 = \mathbf{3 \text{ graus de liberdade}}$

Valor esperado para uma análise significativa: $p = 0,01 \rightarrow 11,345 (< 11,824)$

Portanto: O *qui-quadrado* obtido equivale a uma significância estatística menor que 0.01 (o *qui-quadrado* de $p = 0,01$, considerando-se 3 graus de liberdade, é de **11,345** – um valor um pouco abaixo de **11,824**). Portanto, a diferença da análise B (mais detalhada) em relação à análise C (3 fatores) é significativa. Logo, de um ponto de vista estatístico, a análise detalhada é a melhor.

Resultado: A **análise B** (mais detalhada: 6 fatores) é a melhor comparativamente à análise C (3 fatores).

³³ Uma tabela dos valores de *qui-quadrado* em relação a diferentes graus de abertura (cf. Guy & Zilles, 2007) é apresentada no anexo 2 (p. 217).

Teste de *qui-quadrado*:

Análise B: $\log \text{likelihood} = -1250.783$ (6 fatores)

Análise D: $\log \text{likelihood} = -1264.394$ (2 fatores)

$$(1264.394) - (1250.783) = 13,611 \times 2 = \mathbf{27,222} (= \text{qui-quadrado})$$

Graus de liberdade: $6 - 2 = \mathbf{4 \text{ graus de liberdade}}$

Valor esperado para uma análise significativa: $p = 0,01 \rightarrow 13,277 (< 27,222)$

Portanto: O *qui-quadrado* obtido equivale a uma significância estatística menor que 0.01 (o *qui-quadrado* de $p = 0,01$, considerando-se 4 graus de liberdade, é de **13,277** – um valor abaixo de **27,222**). Portanto, a diferença da análise B (mais detalhada) em relação à análise D (2 fatores) é significativa. Logo, de um ponto de vista estatístico, a distinção entre 6 tipos de referências genéricas se mostrou mais significativa do que a oposição “específica” versus “genérica”.

Resultado: A **análise B** (mais detalhada: 6 oposições da referência) é a melhor de um ponto de vista estatístico do que uma análise com apenas duas oposições (análise D).

Teste de *qui-quadrado*:

Análise C: $\log \text{likelihood} = -1256.695$ (3 fatores)

Análise D: $\log \text{likelihood} = -1264.394$ (2 fatores)

$$(1264.394) - (1256.695) = 7,699 \times 2 = \mathbf{15,398} (= \text{qui-quadrado})$$

Graus de liberdade: $3 - 2 = \mathbf{1 \text{ grau de liberdade}}$

Valor esperado para uma análise significativa: $p = 0,01 \rightarrow 6,635 (< 15,398)$

Portanto: O valor de *qui-quadrado* obtido (15,398) é maior do que o esperado para uma análise significativa.

Resultado: Comparadas as análises concisas, a **análise C** (3 fatores) é a melhor de um ponto de vista estatístico do que a análise D (2 fatores).

De acordo com os testes realizados, as análises detalhadas da referência do pronome são estatisticamente melhores do que análises mais concisas. Por outro lado, uma análise que divide a referência em seis tipos (P2, P1/P4 ‘exclusivo’, P3/P6, P4 ‘inclusivo’, Híbrido, Expletivo) é mais significativa do que uma análise da referência em nove tipos.

Referência e grupos de fatores sociais

As tabelas a seguir trazem os resultados de duas análises quantitativas em que se consideraram separadamente os dados de referência específica e de referência genérica, a fim de verificar se a atuação dos fatores sociais difere a depender da referência do pronome.

SEXO/ GÊNERO	CÊ					
	REF Específica			REF Genérica		
	Nº / Total	%	p.r.	Nº / Total	%	p.r.
Mulheres	69/89	77,5	.64	480/913	52,6	.52
Homens	81/138	58,7	.40	333/719	46,3	.47
Total	150/227	66,1	Range: .12	813/1632	49,8	Range: .07
Input: 0.692			l.l.= -121.374 p < 0.02	Input: 0.465		
				l.l.= -909.954 p < 0.05		

Tabela 37: Emprego de *você* e *cê* conforme a referência do pronome e o Sexo/Gênero dos informantes (GESOL-SP-2000)³⁴

Quando a referência do pronome é genérica, há um ligeiro favorecimento ao emprego da variante reduzida pelas mulheres (resultado idêntico ao observado na amostra geral). No entanto, tanto homens como mulheres empregam mais frequentemente a variante reduzida do pronome quando a referência é específica ao interlocutor³⁵ (77% entre as mulheres; e 59% entre os homens). Além disso, observa-se forte tendência entre as mulheres de empregar a variante reduzida nos contextos cuja referência do pronome é específica (.64) – resultado que era omitido em uma análise conjunta dos dois tipos de referência.

³⁴ Mantiveram-se, nesta análise, os mesmos grupos de fatores analisados nas análises precedentes (v. nota 31, p. 153).

³⁵ A fim de manter um paralelo com as análises anteriores, excluíram-se, aqui também, os dados de discursos reportados e aqueles em que /SE/ aparecia adjacente ao pronome.

No caso da faixa etária, o padrão observado no *corpus* geral é mantido para ambas as referências: os falantes mais jovens e mais velhos favorecem o emprego da variante reduzida, e os da faixa etária intermediária desfavorecem-no. Na análise da referência específica, provavelmente devido ao número reduzido de dados, a faixa etária não foi selecionada como estatisticamente relevante. Nota-se, contudo, que a frequência de *cê* com referência específica é superior a 50% entre as três faixas etárias e é ainda mais frequente entre os mais velhos (71%).

CÊ						
IDADE	REF <u>Específica</u>			REF <u>Genérica</u>		
	Nº / Total	%	p.r.	Nº / Total	%	p.r.
20-32	51/77	66,2	(.55)	309/503	61,4	.61
35-45	27/49	55,1	(.36)	192/491	39,1	.37
50 ou +	72/101	71,3	(.54)	312/638	48,9	.51
Total	150/227	66,1	<i>Range:19</i>	813/1632	49,8	<i>Range: .24</i>
<i>(One Level)</i>				<i>Input: 0.465 l.l.= -909.954 p < 0.05</i>		
<i>* Exceto dados com /SE/ adjacente ao pronome; 2ª p. pl.; discursos reportados e aqueles classificados como “outros”.</i>						

Tabela 38: Emprego de CÊ de acordo com a referência (específica x genérica) e a Faixa etária (Inf.-GESOL-SP-2000)

Ramos (1997) também verifica maior frequência de *cê* genérico (indefinido) entre os informantes mais velhos (65%) do que entre os informantes medianos (40%) ou mais jovens (33%) – v. tabela 39 a seguir. A autora encontra exatamente o inverso para a referência genérica (indefinida): são os informantes mais jovens e de faixa etária mediana que mais utilizam a variante reduzida (56% e 63%, respectivamente), ao passo que os mais velhos empregam-na com menor frequência (33%). Esses resultados podem, segundo a autora, ser entendidos como: “*uma migração do sistema definido para o indefinido*” – de modo semelhante ao uso do clítico indefinido *on* em Montreal (Laberge & Sankoff, 1976; apud. Ramos, 1997).

IDADE	Referência Definida						Referência Indefinida					
	Cê		Ocê		Você		Cê		Ocê		Você	
	Nº.	%	Nº.	%	Nº.	%	Nº.	%	Nº.	%	Nº.	%
Jovem	2	33	-	-	4	67	47	56	16	19	21	25
Mediano	29	40	9	13	34	47	35	63	6	11	14	26
Velho	71	65	8	7	31	28	5	33	2	13	8	53
Total	102		17		69		87		24		43	

Tabela 39: Uso de *você*, *ocê* e *cê* de acordo com a referência e a faixa etária em BH (Ramos, 1997)

Os resultados de Ramos (1997) se assemelham aos da presente pesquisa relativamente ao uso do pronome reduzido com referência específica entre os informantes mais velhos. No entanto, não se verifica na comunidade de fala paulistana o mesmo padrão depreendido pela autora em Minas Gerais. Podem ser diversas as razões para isso: desde questões diatópicas (Belo Horizonte *versus* São Paulo) a questões sociais (diferenças nas faixas etárias dos informantes nas duas pesquisas) e diacrônicas (os *corpora* são de épocas diferentes).

Relativamente à Escolaridade, embora os pesos relativos nas análises dos dois tipos de referência sejam muito semelhantes na cidade de São Paulo, as percentagens novamente indicam que, quando a sua referência é específica, a variante reduzida do pronome é mais frequente que a sua forma plena – tanto entre os informantes com nível superior quanto entre aqueles que não cursaram uma faculdade (tabela 40).

ESCOLARIDADE	CÊ					
	REF <u>Específica</u>			REF <u>Genérica</u>		
	Nº / Total	%	p.r.	Nº / Total	%	p.r.
Fundamental/Médio	75/100	75,0	.61	528/878	60,1	.61
Superior	75/127	59,1	.41	285/754	37,8	.37
Total	150/227	66,1	Range: .20			Range: .24
Input: 0.692			LL= -121.374	Input: 0.465		
			p < 0.02			

Tabela 40: Uso de CÊ de acordo com a referência e o nível de escolaridade (Infs.-GESOL-SP-2000)³⁶

³⁶ Analisaram-se aqui também os mesmos grupos de fatores das análises precedentes (cf. nota 31, p. 153).

Referência específica e tipo de discurso

A fim de desenvolver uma análise da referência específica com um maior número de ocorrências do pronome, foram investigados também os dados provenientes da fala dos documentadores e aqueles em discursos reportados, incluindo-se o “tipo de discurso” como um grupo de fatores.

Além dos usos específicos à segunda pessoa do singular, incluíram-se também 52 ocorrências de segunda pessoa do plural (antes não analisadas justamente em virtude de sua referência específica). Também foram consideradas, nestas análises, aqueles casos em que havia um /SE/ adjacente ao pronome. Os resultados obtidos são apresentados na tabela a seguir³⁷:

	VOCÊ		CÊ			
SEXO / GÊNERO	Nº	%	Nº	%	Peso	Total
Feminino	129	43,9	165	56,1	-	294
Masculino	101	41,6	142	58,4	-	243
FAIXA ETÁRIA	Nº	%	Nº	%	Peso	Total
20-32	70	42,7	94	57,3	.50	164
35-45	71	52,6	64	47,4	.40	135
50 ou mais	89	37,4	149	62,6	.55	258
					Range: .15	
ESCOLARIDADE	Nº	%	Nº	%	Peso	Total
Fundamental/Médio	90	34,2	173	65,8	.59	263
Superior	140	51,1	134	48,9	.41	274
					Range: .18	
DISCURSO	Nº	%	Nº	%	Peso	Total
não reportado	111	39,4	171	60,6	.55	282
Reportado	119	46,7	136	53,3	.45	255
					Range: .10	
FLEXÃO	Nº	%	Nº	%	Peso	Total
Singular	198	40,8	287	59,2	.52	485
Plural	32	61,5	20	38,5	.29	52
Total	230	42,8	307	57,2	Range: .23	537
<i>p < 0.05 ll. = -340.016 input: .578</i>						

Tabela 41: Uso de *Você/Cê* específico conforme o sexo, a idade, a escolaridade, o discurso e a flexão do pronome (Infs. – GESOL-SP-2000)

³⁷ O número absoluto de dados com referência específica é bastante superior àqueles descritos até aqui porque foram incluídos aqui também os dados de discursos reportados e de 2ª p. pl. Além dos cinco grupos de fatores expressos na tabela 41, incluiu-se nesta análise também a “frequência do verbo”.

Com relação á faixa etária, vê o mesmo padrão curvilíneo apresentado na análise geral dos dados (tabela 17, p. 134). Chama a atenção, contudo, o fato de que os falantes mais velhos são aqueles que demonstram maior tendência de emprego da variante reduzida do pronome quando a sua referência é específica, ultrapassando mesmo os da primeira faixa etária (que se mostram neutros) e os da segunda faixa etária (que desfavorecem essa variante).

O discurso reportado, por sua vez, desfavorece um pouco o emprego da forma reduzida (.45) e o plural do pronome (*vocês/cês*) desfavorece bastante tal forma (.29). Deve-se ressaltar, contudo, o reduzido número de casos de segunda pessoa do plural (52 em um total de 537), o que compromete generalizações acerca deste último resultado para a comunidade de fala paulistana.

O grupo sexo/gênero não foi selecionado como relevante – contrariamente ao que se observou na análise em que se desconsideraram os dados sem discursos reportados e sem a adjacência a um elemento /SE/³⁸.

Para ampliar o número de dados com referência específica e refinar a análise, incluíram-se as ocorrências do pronome provenientes da fala dos nove documentadores paulistanos – sete mulheres (entre 21 e 32 anos) e dois homens (com 37 e 52 anos), todos cursando o nível superior de ensino.

A distribuição desequilibrada dos dados, não possibilitou uma análise dos grupos de fatores de natureza social. A tabela a seguir apresenta como as variantes com referência específica se distribuem entre os documentadores e os informantes.

VOCÊ/CÊ (REF Específica)					
VARIANTES	Informantes (36)		Documentadores (9)		TOTAL
	Nº	%	Nº	%	Nº %
VOCÊ	230	42,8	534	53,3	764 49,7
CÊ	307	57,2	467	46,7	774 50,3
Total	537	100	1001	100	1538 100

Tabela 42: Uso de CÊ com referência específica de acordo com o papel do falante na entrevista (Inf-Doc – GESOL-SP-2000)

³⁸ Com relação aos demais grupos de fatores considerados na análise, nota-se um padrão semelhante àquele observado nas análises precedentes (seja para a referência específica, seja para as ocorrências genéricas); por isso não são reportados aqui.

Os informantes empregam a variante reduzida mais frequentemente do que os documentadores. Isto provavelmente se deve ao fato de que os documentadores seguiram um roteiro para as entrevistas: instruções escritas podem ter influenciado suas falas, além da possibilidade de que, no papel de condutores da entrevista, podem ter desempenhado monitoramento mais constante de suas falas. Trata-se de um problema metodológico que convém enfatizar: a análise da fala dos documentadores, por um lado, eleva o número de ocorrências do pronome com referência específica; mas, por outro, pode ser menos espontânea e enviesar, por receber influência da escrita, as análises dos dados de fala.

Variáveis de natureza discursiva e cognitiva - paralelismo e frequência

Uma análise detalhada do grupo de fatores “paralelismo das variantes pronominais” (que se optou por chamar “paralelismo I” neste trabalho) revela resultados bastante interessantes e estatisticamente relevantes. Uma vez empregada a variante reduzida *cê*, o falante tende a continuar repetindo o emprego da mesma forma ao longo de seu texto oral.

PARALELISMO I	VOCÊ		CÊ			Total
	Nº	%	Nº	%	p.r.	Nº
3ª ou + de uma série CÊ	24	15,0	136	85,0	.85	160
2ª de uma série CÊ	45	20,8	171	79,2	.79	216
3ª série mista VOCÊ-CÊ	32	33,3	64	66,7	.67	96
3ª série mista CÊ-ZERO	5	29,4	12	70,6	.61	17
3ª série mista CÊ-VOCÊ	38	50,7	37	49,3	.52	75
Isolado	391	48,9	408	51,1	.50	799
3ª série mista ZERO-VOCÊ	10	50,0	10	50,0	.46	20
1ª de uma série	325	60,0	217	40,0	.42	542
2ª de uma série VOCÊ	246	73,4	89	26,6	.30	335
3ª de uma série VOCÊ	164	76,3	51	23,7	.29	215
Total	1280	51,7	1195	48,3	Range: .56	2475

Input: 0.447 l.l.= -1387.492 p = 0.02

Tabela 43: Uso de *você* e *cê* quanto ao Paralelismo das variantes no discurso – (Inf.-GESOL-SP-2000)³⁹

³⁹ Nesta análise, foram controlados os seguintes grupos de fatores: “Sexo/Gênero”, “Faixa etária”, “Escolaridade”, “Referência”, “Coalescência do pronome ao verbo”, “Frequência do verbo”, “Tonicidade seguinte”, “Tonicidade antecedente” e “Paralelismo I”. Todos selecionados como estatisticamente relevantes.

A tabela anterior mostra que há um grande favorecimento ao emprego da variante reduzida quando esta integra séries de orações em que há duas ou mais formas antecedentes idênticas a ela (.85) – em um total de 160 ocorrências do pronome antecedidas por duas ocorrências da variante reduzida, apenas 24 (15%) não seguiram esse padrão. No caso de haver uma única forma antecedente *cê*, a tendência desta variante continuar a ser empregada no contexto seguinte persiste (.79).

Em séries mistas, em que se alternam as variantes *você* e *cê*, nesta ordem, a última continua a ser empregada (.67), ainda que com uma probabilidade menor do que nos casos anteriores. Mesmo quando a sequência pronominal é quebrada por um zero (séries mistas CÊ-Zero), se a última forma pronominal pronunciada for *cê*, há grandes chances de que a seja retomada após o zero (.61).

De maneira similar, a tendência ao emprego de *cê* reduz-se consideravelmente se nas orações que a antecedem o falante emprega a variante plena *você* (pesos relativos entre .29 e .30); ou seja, tais contextos favorecem a manutenção do uso da forma plena do pronome.

Os casos sem antecedentes pronominais de segunda pessoa nas orações precedentes mostraram-se neutros ou próximos ao ponto neutro. Os empregos “isolados” apresentam um peso relativo neutro (.50), enquanto os primeiros de uma série desfavorecem um pouco o uso da variante reduzida (.42). Esse último resultado pode ser explicado pela atuação de outros fatores que não o paralelismo.

A fim de tornar a análise mais sucinta como recomenda a conhecida *navalha de Occam* (Guy & Zilles, 2007), outra análise dessa variável foi desenvolvida, amalgamando-se os empregos do pronome conforme figurassem: sem antecedentes (isolado ou primeiro de uma série); anteceditos por *cê*; anteceditos por *você*. Os resultados dessa análise são os seguintes:

PARALELISMO I	VOCÊ		CÊ			Total
	Nº	%	Nº	%	p.r.	Nº
Sem antecedente	716	53,4	625	46,6	.47	1341
Antecedido por CÊ	106	21,7	383	78,3	.78	489
Antecedido por VOCÊ	458	71,0	187	29,0	.33	645
Total	1280	51,7	1195	48,3	Range: .45	2475
Input: 0.445 ll.= -1403.699 p = 0.02						

Tabela 44: Emprego de *você/cê* e paralelismo das variantes – Análise concisa (Inf.-GESOL-SP-2000)⁴⁰

Scherre (1998) considera o “paralelismo” como uma sorte de princípio universal de processamento paralelo, cuja natureza é cognitivo-discursiva. Os resultados acima confirmam essa hipótese, na medida em que, uma vez empregada uma determinada variante pronominal, o falante tende a continuar empregando-a nas orações seguintes.

Outro grupo de fatores se propôs a verificar se o emprego de itens reduzidos adjacentes ao pronome estaria correlacionado à sua pronúncia reduzida. Consideraram-se os itens *num* (redução de “não”) e *tá/tava/tavam* (reduções do verbo “estar”). Para diferenciar este tipo de paralelismo do anterior, ele foi denominado “paralelismo de formas reduzidas” (ou “paralelismo II”).

PARALELISMO II	VOCÊ		CÊ		Total
	Nº	%	Nº	%	Nº
<i>Num tá / num tava...</i>	3	33,3	6	66,7	9
<i>Num</i>	108	49,1	112	50,9	220
<i>tá</i> (1 síl.)	47	36,2	83	63,8	130
<i>tava...</i> (2+ síl.)	3	60,0	2	40,0	5
Não há reduções	1119	53,0	992	47,0	2111
Total	1280	51,7	1195	48,3	2475

Tabela 45: Distribuição de *você* e *cê* conforme o paralelismo de formas reduzidas (Infs – GESOL-SP-2000)

Na presença de itens reduzidos, é mais frequente o emprego da variante *cê*. Por outro lado, o número muito reduzido de ocorrências do pronome em alguns dos casos acima exigiu uma análise mais sucinta desse grupo de fatores:

⁴⁰ Mantiveram-se os mesmos grupos de fatores investigados na análise descrita anteriormente (v. nota 39, p. 162).

PARALELISMO II	VOCÊ		CÊ			Total
	Nº	%	Nº	%	p.r.	Nº
Reduções adjacentes	161	44,2	203	55,8	.63	364
Não há reduções adjacentes	1119	53,0	992	47,0	.48	2111
Total	1280	51,7	1195	48,3		2475
Input: 0.445 ll.= -1419.884 p < 0.02						Range: .15

Tabela 46: Uso de *você* e *cê* conforme a presença de formas reduzidas adjacentes ao pronome – análise concisa (Infs – GESOL-SP-2000)⁴¹

Este grupo de fatores foi selecionado pelo Goldvarb como estatisticamente relevante, o que permite concluir a atuação de mais esse caso de paralelismo no emprego de *cê*. Os gráficos a seguir ilustram melhor este fato.

⁴¹ Além do Paralelismo II, foram controlados os grupos de fatores sociais (Sexo/gênero, Faixa etária e Escolaridade) e os linguísticos “Referência”, “Coalescência do pronome ao verbo” (adjacente x não adjacente), “Tonicidade antecedente”, “Paralelismo I” e “Frequência do verbo”. A Tonicidade seguinte não foi controlada, pois se detectou interação desta com o grupo de fatores Paralelismo II.

Finalmente, no que diz respeito ainda a grupos de fatores de natureza cognitivo-discursiva, a variante reduzida *cê* também tende a ser mais empregada quando o verbo que a acompanha é altamente frequente no *corpus* (.58), ao passo que é um pouco desfavorecida com verbos infrequentes (.44). Ainda que tais sequências [pronome + (X) + verbo] não constituam construções cristalizadas na língua, pode-se prever maior automatização no emprego de formas reduzidas nos contextos mais previsíveis do que naqueles menos utilizados.

FREQUÊNCIA DO V	VOCÊ		CÊ			Total
	Nº	%	Nº	%	p.r.	Nº
+ freq. (acima de 3%)	459	42,8	614	57,2	.58	1073
- freq. (até 3%)	821	58,6	581	41,4	.44	1402
Total	1280	51,7	1195	48,3		2475

Input: 0.445 l.l.= -1403.699 p = 0.02 Range: .14

Tabela 47: Emprego de *você/cê* conforme a frequência do verbo (Infs – GESOL-SP-2000)⁴²

Variáveis fonético-fonológicas e a atuação do *Princípio do Contorno Obrigatório*

Os resultados⁴³ apresentados a seguir indicam que: (i) sílabas tônicas ou de tonicidade mediana (as formas reduzidas *num* e *tá*) favorecem o uso da variante reduzida do pronome; (ii) sílabas átonas desfavorecem-na (.41); e (iii) pausas no contexto seguinte desfavorecem ainda mais o emprego de *cê*.

TONICIDADE SEGUINTE	VOCÊ		CÊ			Total
	Nº	%	Nº	%	p.r.	
Mediana	152	43,4	198	56,6	.64	350
Tônica	666	46,9	753	53,1	.52	1419
Átona	399	63,1	233	36,9	.41	632
Pausas/hesitações	63	85,1	11	14,9	.21	74
Total	1280	51,7	1195	48,3		2475

Input: 0.440 l.l.= -1391.726 p < 0,02 Range: .43

Tabela 48: Uso de *Você* e *cê* conforme o contexto seguinte ao pronome (Infs – GESOL-SP-2000)

⁴² A análise incluiu os grupos de fatores “Sexo/gênero”, “Faixa etária”, “Escolaridade”, “Referência”, “Coalescência do pronome ao verbo”, “Tonicidade antecedente”, “Tonicidade seguinte” e “Frequência do verbo”.

⁴³ A análise apresentada na tabela 48 incluiu os mesmos grupos de fatores expressos na nota 42.

A divisão desse grupo em quatro fatores, todavia, acaba por levar à análise de distintas questões: (i) sílabas tônicas no contexto seguinte favorecem ligeiramente o uso de *cê* (.52); (ii) itens reduzidos (*num/tá*) no contexto seguinte favorecem ainda mais o emprego de *cê* (.64) – o que remete à atuação do Paralelismo de formas reduzidas, comentado anteriormente; (iii) pausas desfavorecem muito o uso da variante reduzida (.21). Este último resultado emparelha-se com a hipótese de cliticização de *cê* no PB (Vital, 2006a): na falta de um constituinte no qual possa se “apoiar”, *cê* é desfavorecido. Embora todas estas questões sejam importantes para o entendimento da variação *você/cê*, o fato de que sejam todas elas agrupadas em uma mesma variável independente não corresponde ao que se recomenda na metodologia de análise quantitativa de dados linguísticos.

Para solucionar este problema, procedeu-se a três análises⁴⁴: (i) excluindo-se os dados com pausa no contexto seguinte; (ii) amalgamando-se a tonicidade mediana aos contextos tônicos e excluindo-se os dados com pausa; e (iii) excluindo-se tanto os dados com pausa, quanto aqueles com itens de tonicidade mediana.

TONICIDADE SEGUINTE	VOCÊ		CÊ			Total
	Nº	%	Nº	%	p.r.	
Mediana	152	43,4	198	56,6	.64	350
Tônica	666	46,9	753	53,1	.51	1419
Átona	399	63,1	233	36,9	.40	632
Total	1217	50,7	1184	49,3		2401*
Input: 0.443 ll.= -1396.892 p < 0,02 Range: .24						
Tônica + Mediana	818	46,2	951	53,8	.53	1769
Átona	399	63,1	233	36,9	.41	632
Total	1217	50,7	1184	49,3		2401*
Input: 0.446 ll.= -1403.669 p < 0.05 Range: .12						
Tônica	666	46,9	753	53,1	.53	1419
Átona	399	63,1	233	36,9	.43	632
Total	1065	51,9	986	48,1		2051**
Input: 0.435 ll.= -1401.661 p < 0.02 Range: .10						
*Exceto pausas e hesitações. **Exceto pausas e hesitações e tonicidade mediana.						

Tabela 49: Uso de *você/cê* e Tonicidade da sílaba seguinte ao pronome (Inf – GESOL-SP-2000)

⁴⁴ Grupos de fatores investigados: “Sexo/Gênero”, “Faixa etária”, “Escolaridade”, “Referência”, “Coalescência”, “Paralelismo I”, “Tonicidade antecedente”, “Tonicidade seguinte” e “Frequência do verbo”.

A tonicidade mediana é o fator que mais favorece o emprego da variante reduzida (.64), ao passo que a presença de uma sílaba tônica no contexto seguinte é neutra para o emprego de *cê* (.51). De início, pode-se pensar que os itens *num* e *tá* seriam, na verdade, tônicos (e não intermediários entre sílabas tônicas e átonas). No entanto, isto não explica por que estes contextos favoreceriam ainda mais o emprego de *cê* do que os demais contextos tônicos. A análise conjunta destes fatores revela que contextos tônicos favorecem um pouco o uso de *cê* (.53), ao passo que os átonos desfavorecem-no (.41).

Com o objetivo de analisar o efeito da tonicidade do contexto seguinte, sem a sobreposição do Paralelismo II (a presença de formas reduzidas no contexto adjacente ao pronome), excluíram-se da análise deste grupo os dados com itens reduzidos no contexto seguinte (cuja tonicidade foi inicialmente considerada intermediária).

Os resultados obtidos, porém, são muito próximos aos anteriores: sílabas tônicas no contexto seguinte favorecem ligeiramente o emprego de *cê* (.53), sílabas átonas desfavorecem essa variante (.43). Atua, portanto, o *Princípio do Contorno Obrigatório*: concebendo-se *cê* como um item que funciona como clíticos, conforme sugerem Vitral (2006a) e Ciriaco et. al. (2004), o falante tende a evitar o encontro de duas sílabas com acento tônico (choque acentual), ao mesmo tempo em que evitam a adjacência de muitas sílabas átonas (lapso acentual). Este último caso é mais nítido do que o primeiro. A atuação da tonicidade da sílaba seguinte no emprego das variantes pronominais em estudo, todavia, mostra-se bastante sutil (o *range* é de apenas 0.10). A análise de outras questões rítmicas (como, por exemplo, a possível retração do acento em determinados empregos da variante plena *você*) poderá trazer futuras contribuições a esta discussão.

Por outro lado, os contextos antecedentes átonos não favorecem o uso de *cê*, como se havia inicialmente cogitado (v. tabela 50 abaixo). Ao contrário, tal ambiente desfavorece o emprego dessa variante (.41), enquanto sílabas tônicas antecedentes ao pronome favorecem-na ligeiramente (.56). É interessante observar que, no contexto que antecede o pronome, pausas favorecem o uso da variante reduzida (.61).

TONICIDADE ANTECEDENTE	VOCÊ		CÊ			TOTAL
	Nº	%	Nº	%	p.r.	
Pausas/hesitações	318	42,0	440	58,0	.61	758
Tônica	203	47,1	228	52,9	.56	431
Átona	759	59,0	527	41,0	.41	1286
Total	1280	51,7	1195	48,3		2475
Input: 0.491 l.l.= -1404.488 p < 0,02						Range: 20

Tabela 50: *Você/cê* conforme o contexto antecedente ao pronome (Infs – GESOL-SP-2000)⁴⁵

Gonçalves (2008), em uma análise do contexto fônico precedente (consoante, vogal ou pausa), também verifica maior frequência de uso da variante reduzida *cê* (72%) quando há uma pausa no contexto antecedente ao pronome. O mesmo observa Coelho (1999; apud. Gonçalves, 2008).

A fim de refinar a análise da tonicidade do contexto antecedente, excluíram-se os dados que apresentavam uma pausa antes do pronome:

TONICIDADE ANTECEDENTE	VOCÊ		CÊ			TOTAL
	Nº	%	Nº	%	p.r.	
Tônica	203	47,1	228	52,9	.61	431
Átona	759	59,0	527	41,0	.46	1286
Total	962	56,0	755	44,0		1717*
Input: 0.445 l.l.= -1403.699 p < 0,05						Range: 15
*Exceto pausas e hesitações.						

Tabela 51: Uso de *você* e *cê* conforme a Tonicidade da sílaba antecedente ao pronome (Infs – GESOL-SP-2000)

Observa-se tendência ao uso da variante *cê* quando precedida de uma sílaba tônica (.61). Tal resultado é semelhante àquele obtido para o contexto seguinte ao pronome e confirma a hipótese de que tal forma encontra-se mais próxima do funcionamento de um item átono do que de um itemônico.

Ainda no que diz respeito a variáveis fonético-fonológicas, analisou-se o impacto de constituintes /SE/ no emprego da variante reduzida do pronome. Em princípio, o grupo de fatores proposto para esse fim distinguia estes constituintes de acordo com sua posição (seguinte ao pronome, antecedente ou ambos).

⁴⁵ Para as análises apresentadas nas tabelas 50 e 51, foram investigados os mesmos grupos de fatores descritos na nota 44 (p. 167).

PRESENÇA DE /SE/	VOCÊ		CÊ		TOTAL
	Nº	%	Nº	%	
/SE/ antecedendo o pronome	178	82,8	37	17,2	215
/SE/ seguindo o pronome	16	57,1	12	42,9	28
/SE/ antes e após o pronome	3	100	-	0	3
Não h	1083	48,6	1146	51,4	2229
Total	1280	51,7	1195	48,3	2475

Tabela 52: Distribuição de *você* e *cê* quanto à presença de constituintes /SE/ adjacentes ao pronome – (Infs – GESOL-SP-2000)

Tal detalhamento da análise (tabela 52), contudo, não permite uma análise multivariada, devido à baixa frequência de algumas estruturas na amostra. Assim, essa hipótese foi reformulada, de maneira binária: presença ou ausência de constituintes /SE/ adjacentes. Verificou-se que, embora a presença de elementos /SE/ – [sɪ], [sɐ], [sê] – tenha sido pouco frequente no *corpus* (246/2475), é notável que a variante reduzida *cê* seja consideravelmente desfavorecida na presença de algum desses elementos – confirmando, mais uma vez, a atuação do *Princípio do Contorno Obrigatório*.

PRESENÇA DE /SE/	VOCÊ		CÊ			TOTAL
	Nº	%	Nº	%	p.r.	
Sim	197	80,1	49	19,9	.20	246
Não	1083	48,6	1146	51,4	.54	2229
Total	1280	51,7				2475
Input: 0.491 ll.= -1397.302 p < 0,05 Range: .34						

Tabela 53: Uso de *você/cê* relativamente à presença de /SE/ adjacente (Inf – GESOL-SP-2000)⁴⁶

Variáveis morfossintáticas

A tabela a seguir traz uma distribuição geral dos dados de acordo com os fatores do grupo “coalescência do pronome ao verbo”:

⁴⁶ Além do grupo de fatores “Presença de /SE/”, foram investigados nesta análise as seguintes variáveis independentes: “Referência” (específica x genérica), “Coalescência do pronome ao verbo” (adjacente x não-adjacente), “Paralelismo II” (sim x não), “Tonicidade seguinte” (tônica + mediana x átona), “Tipo de oração”, “Função sintática” e “Frequência do verbo”. Em análises com um número menor de grupo de fatores linguísticos ou apenas com os grupos de fatores sociais, os pesos relativos para a “Presença x ausência de /SE/” mantiveram-se idênticos.

COALESCÊNCIA	VOCÊ		CÊ		TOTAL
	Nº	%	Nº	%	
Adjacente	978	49,7	990	50,3	1968
“que”	4	50,0	4	50,0	8
Clíticos	18	58,1	13	41,9	31
“num”	106	48,4	113	51,6	219
“não”	22	100	-	0	22
ditongo não-nasal	1	50,0	1	50,0	2
ditongo nasal ‘nem’	2	28,6	5	71,4	7
monossílabos tônicos CV	29	46,0	34	54,0	63
Advérbios ou loc. adv.	28	62,2	17	37,8	45
2 ou + palavras	9	100	-	0	9
Não se aplica*	83	82,2	18	17,8	101
Total	1280	51,7	1195	48,3	2475

* Pausas, hesitações, materiais discursivos (eh, ah, ähn, etc.), verbo subentendido.

Tabela 54: Distribuição de *você* e *cê* conforme o tipo de constituinte interveniente entre o pronome e o verbo (Inf – GESOL-SP-2000)

Como se nota, a grande maioria das ocorrências do pronome, seja reduzido ou pleno, aparece adjacente ao verbo. Nota-se que os constituintes *num* e *não*, quando intervenientes entre o pronome sujeito e o verbo, apresentam um comportamento bastante diferenciado em relação ao emprego da variante reduzida *cê*: quando há um “não” pré-verbal, o uso da variante plena *você* foi categórico na amostra analisada; ao contrário, quando esse advérbio de negação é pronunciado de forma reduzida (*num*), há uma taxa de quase 52% de emprego de *cê*. Essa distinção corrobora a hipótese de Corrêa & Ramos (2006) de que *num* pode ser considerado um clítico.

No entanto, uma distinção minuciosa do tipo de material interveniente entre o pronome e o verbo não se mostra exequível em uma análise multivariada, dado o reduzido número de dados obtido em algumas células. Sendo assim, procedeu-se a três análises dessa variável: (i) considerando-se a mobilidade ou não do constituinte interpolado; (ii) considerando-se o número de sílabas do constituinte interpolado; (iii) distinguindo-se a existência ou não de interpolação de constituintes entre o pronome e o verbo. Trata-se de mais uma confirmação da hipótese de que a variante reduzida do pronome estaria passando por um processo de cliticização.

Para a primeira dessas análises, foram considerados como constituintes “fixos” no sintagma: *não, num, que, nem* e clíticos (*se, me, lhe*, etc.)⁴⁷. Como constituintes “móveis” no sintagma, foram classificados os monossílabos tônicos com ou sem ditongo (*só, já, mal*), advérbios, locuções adverbiais ou sintagmas mais extensos.

Quando adjacente ao verbo ou quando há interveniência de constituintes fixos entre o pronome e o verbo, os pesos relativos obtidos encontram-se muito próximos entre si e muito próximos do ponto neutro (.51 e .49). Já quando os elementos intervenientes são móveis, é desfavorecido o emprego da variante reduzida do pronome (.41).

COALESCÊNCIA	VOCÊ		CÊ			TOTAL
	Nº	%	Nº	%	p.r.	
Adjacente	978	49,7	990	50,3	.51	1968
Constituintes fixos	152	53,0	135	47,0	.49	287
Constituintes móveis	67	56,3	52	43,7	.41	119
Total	1197	50,4	1177	49,6		2374*
Input: 0.446 ll.= -1427.092 p< 0.02					Range: .10	
* Exceto: pausas, hesitações, materiais discursivos, verbo subentendido.						

Tabela 55: Uso de *Você* e *cê* de acordo com a mobilidade do constituinte interpolado
(Infs – GESOL-SP-2000)⁴⁸

No que diz respeito à extensão fônica do elemento interveniente, todos os fixos analisados anteriormente são também monossilábicos, de modo que a análise do peso silábico do constituinte interveniente leva a resultados semelhantes aos da análise anterior: quando o pronome se encontra adjacente ao verbo ou quando há a interpolação de constituintes monossilábicos, observa-se neutralidade para o emprego de *cê*; mas quando o elemento interveniente tem duas ou mais sílabas, é desfavorecido o uso da variante reduzida (.34):

⁴⁷ Embora os clíticos sejam móveis, a ênclise é rara no PB falado atualmente.

⁴⁸ Resultados de uma análise multivariada em que consideraram os grupos de fatores sociais (Sexo/Gênero, Faixa etária e Escolaridade) e os linguísticos “Referência” (específica x genérica), “Coalescência”, “Paralelismo I”, “Tonicidade antecedente” e “Frequência do verbo”.

COALESCÊNCIA	VOCÊ		CÊ			TOTAL
	Nº	%	Nº	%	p.r.	
Adjacente	978	49,7	990	50,3	.51	1968
1 sílaba	182	51,7	170	48,3	.48	352
2 ou + sílabas	37	68,5	17	31,5	.34	54
Total	1197	50,4	1177	49,6		2374*

Input: 0.447 LL=-1427.197 p<0.02 Range: .17

* Exceto: pausas, hesitações, materiais discursivos, verbo subentendido.

Tabela 56: Uso de *Você/cê* e extensão fônica do constituinte interveniente (Infs-GESOL-SP-2000)⁴⁹

Finalmente, numa análise que contrapõe os dados em que o pronome é adjacente ao verbo àqueles em que se observa interpolação, a variante reduzida do pronome é desfavorecida em estruturas interpoladas (.46)⁵⁰. No entanto, essa análise mais geral, que não leva em consideração a natureza ou extensão fônica dos constituintes que se situam entre o pronome e o verbo, revela-se menos significativa do que as anteriores – como é possível observar pelo *range* muito reduzido (0.05)⁵¹.

COALESCÊNCIA	VOCÊ		CÊ			TOTAL
	Nº	%	Nº	%	p.r.	
Adjacente	978	49,7	990	50,3	.51	1968
Não-adjacente	219	53,9	187	46,1	.46	406
Total	1197	50,4	1177	49,6		2374*

Input: 0.447 LL=-1429.323 p<0.02 Range: .05

* Exceto: pausas, hesitações, marcadores discursivos, verbo subentendido.

Tabela 57: Uso de *você/cê* e coalescência do pronome ao verbo (Infs – GESOL-SP-2000)

⁴⁹ Nas análises apresentadas nas tabelas 56 e 57, foram controlados os mesmos grupos de fatores da análise anterior (v. tabela 56 e nota 48, p. 172).

⁵⁰ Além das três análises aqui reportadas, em uma quarta, excluíram-se os dados com “não” interveniente (contexto cujo emprego da variante plena *você* foi categórico na amostra). Para esta última análise, os constituintes intervenientes foram reclassificados em três categorias: monossílabos tônicos, monossílabos átonos e outros constituintes. Os resultados demonstram um comportamento semelhante dos monossílabos átonos (clíticos e *num*) e dos contextos em que o pronome e o verbo encontram-se adjacentes – ambos mostraram-se neutros (pesos relativos: .52 e .50, respectivamente). Os monossílabos tônicos, por sua vez, desfavoreceram muito pouco a variante reduzida (.48). Tal desfavorecimento foi acentuado apenas no caso de constituintes intervenientes com duas ou mais sílabas (.34).

⁵¹ Um teste de *qui-quadrado* revela que uma análise que distingue os tipos de constituintes intervenientes entre móveis e fixos é preferível a uma análise que não opera essa distinção. O valor do *qui-quadrado* obtido comparando-se as análises representadas nas tabelas 54 e 56 é de 4,462 – um valor maior do que o *qui-quadrado* correspondente para uma significância estatística de 0,05 (3,841), considerando-se um grau de liberdade. O mesmo se verifica entre uma análise concisa da Coalescência e aquela que distingue os constituintes interpolados por sua extensão: o *qui-quadrado* obtido, nesse caso, é de 4,252.

Finalmente, ainda no âmbito das variáveis morfossintáticas, a tabela a seguir apresenta uma distribuição geral das variantes de acordo com os tipos de sentenças.

TIPO DE ORAÇÃO	VOCÊ		CÊ		TOTAL
	Nº	%	Nº	%	
interrogativas s/ const. interr.	45	27,4	119	72,6	164
interrogativas <i>Q-in-situ</i>	3	30,0	7	70,0	10
outras interrogativas <i>QU-que</i>	2	25,0	6	75,0	8
interrogativas ' <i>que-que</i> '	10	30,3	23	69,7	33
outras <i>QU-(é)-que</i>	4	30,8	9	69,2	13
interrogativas 'que'	8	40,0	12	60,0	20
Declarativas	1184	54,0	1007	46,0	2191
outras interrogativas <i>QU</i> ⁵²	24	66,7	12	33,3	36
Total	1280	51,7	1195	48,3	2475

Tabela 58: Distribuição das variantes *você* e *cê* de acordo com o tipo de oração (Infs – GESOL-SP-2000)

A grande maioria dos pronomes ocorreu em orações declarativas. As interrogativas que possuem complementizador⁵³ *que* (*que-que*, *é-que* e outras *QU-que* e demais interrogativas de constituinte) foram extremamente infrequentes na amostra. Para tornar possível sua análise, as interrogativas *que-que*, *é-que* e outras *qu-que* foram amalgamadas (*interrogativas qu-(é)que*).

TIPO DE ORAÇÃO	VOCÊ		CÊ			TOTAL
	Nº	%	Nº	%	p.r.	
Interrogativas qu-(é)-que	16	29,6	38	70,4	.68	54
Outras interrogativas	80	34,8	150	65,2	.67	230
Declarativas	1184	54,0	1007	46,0	.48	2191
Total	1280	51,7	1195	48,3		2475
Input: 0.482 ll.= -1608.128 p = 0.02						Range: .20

Tabela 59: Emprego de *Você* e *cê* conforme o tipo de oração (Infs – GESOL-SP-2000)⁵⁴

⁵² 'Onde', 'quando', 'quem', etc.

⁵³ Núcleo da projeção CP (Miotto, 2004).

⁵⁴ Nas análises apresentadas nas tabelas 59 e 60, foram considerados apenas os grupos de fatores sociais: Sexo/Gênero, Faixa etária e Escolaridade. Todos foram selecionados pelo Goldvarb como estatisticamente relevantes.

As estruturas interrogativas favorecem o emprego da variante *cê*. Ao contrário da hipótese aventada por Ramos (1997), portanto, parece que não se trata de uma questão sintática (o preenchimento ou não da posição de complementizador), mas de uma questão prosódica: estruturas declarativas *versus* estruturas interrogativas. As primeiras mostram-se neutras ao emprego de *cê* (.48), e as interrogativas (amalgamadas) favorecem-na (.67) – tabela 60.

TIPO DE ORAÇÃO	VOCÊ		CÊ			TOTAL
	Nº	%	Nº	%	p.r.	
Interrogativas	96	33,8	188	66,2	.67	284
Declarativas	1184	54,0	1007	46,0	.48	2191
Total	1280	51,7	1195	48,3		2475
<i>Input: 0.482 ll.= -1608.139 p = 0.02</i>						<i>Range: .19</i>

Tabela 60: Uso de *Você* e *cê* em orações declarativas e interrogativas (Infs – GESOL-SP-2000)

Um resultado semelhante a esse é obtido por Queiroz Andrade (2010) em uma análise das variantes *você*, *cê* e *tu* em Vila Planalto (Brasília). A autora também depreende que as estruturas com uma entoação ascendente (como é o caso das orações interrogativas) favorecem o uso das variantes *cê* e *tu*.

No entanto, conforme mostra a distribuição a seguir, na maioria das orações interrogativas, a referência do pronome é específica.

TIPO DE ORAÇÃO	VOCÊ / CÊ					
	REF Específica		REF Genérica/outros		TOTAL	
	Nº	%	Nº	%	Nº	%
Interrogativas qu-que	33	61,1	21	38,9	54	2,2
Outras interrogativas	164	71,3	66	28,7	230	9,3
Declarativas	340	15,5	1851	84,5	2191	88,5
Total	537	21,7	1938	78,3	2475	100

Tabela 61: Emprego do pronome de segunda pessoa conforme a referência do pronome e o tipo de oração (Infs – GESOL-SP-2000)

A fim de verificar se os resultados anteriores se devem ou não devido à questão da referência do pronome, analisou-se o impacto do grupo de fatores

“Tipo de oração” apenas nos dados com referência específica, provenientes da fala dos informantes e dos documentadores (tabela 62).

CÊ (REF Específica)						
TIPO DE ORAÇÃO	Informantes (36)			Documentadores (9)		
	Nº/Total	%	p.r.	Nº/Total	%	p.r.
Interrogativas qu-que	25/33	75,8	.72	19/60	31,7	.40
Outras interrogativas	118/164	72,0	.66	303/625	48,5	.54
Declarativas	164/340	48,2	.40	145/316	45,9	.44
Total	307/537	57,2	Range: .32	467/1001	46,7	Range: .14
<div> <div>Input: 0.555 ll.= -318.236 p < 0.05</div> <div>Input: 0.439 ll.= -568.839 p < 0.02</div> </div>						

Tabela 62: Emprego de CÊ específico conforme o tipo de oração (Infs e Docs: GESOL-SP-2000)⁵⁵

A análise do tipo de oração apenas com os dados cuja referência do pronome é específica, na fala dos informantes, demonstra um padrão semelhante à análise anterior: as interrogativas favorecem o emprego da variante reduzida do pronome (.72 e .66). Já na fala dos documentadores, os resultados são bastante diferentes: as interrogativas *Qu-que* e as declarativas desfavorecem o uso de *cê* (.40 e .44, respectivamente), enquanto as demais interrogativas favorecem ligeiramente o emprego desta variante (.54). Este resultado provavelmente devido à natureza distinta das perguntas feitas por documentadores e informantes: enquanto os primeiros geralmente visam a solicitar uma informação do interlocutor, pode-se questionar se as perguntas realizadas pelos informantes teriam o mesmo caráter, ou seriam perguntas retóricas⁵⁶. Além disso, deve-se considerar o fato de que as perguntas dirigidas aos informantes foram previamente formuladas e ensaiadas; além disso, o documentador poderia estar portando anotações no momento da entrevista – fato que pode ter influenciado sua fala.

⁵⁵ Foram considerados nesta análise também os grupos “relação entre os interlocutores”, a “frequência do verbo” e “paralelismo I”. Todos se mostraram estatisticamente relevantes em uma análise multivariada. Em uma análise dos dados apenas com referência genérica, na qual foram controladas as variáveis sociais e o “Tipo de oração”, esta última variável não selecionada pelo Goldvarb X como estatisticamente significativa. Além disso, na fala dos documentadores, apenas 7 ocorrências do pronome apresentam referência genérica – o que inviabilizou um estudo quantitativo desses dados.

⁵⁶ Oushiro (2011), em um estudo variacionista das interrogativas *QU-*, investiga a sinceridade pragmática da pergunta como um grupo de fatores. A autora considera três tipos de perguntas: retóricas, “sinceras” e de estruturação do discurso. Uma oposição como esta poderá ser útil também para explicar o comportamento distinto de informantes e documentadores quanto ao uso de *você* e *cê* nas orações interrogativas.

4.2. Análise diacrônica – dos anos 1970 aos anos 2000

Distribuição dos dados nos anos 1970 e nos anos 2000

Embora tenha sido considerado o mesmo número de informantes (18 em cada amostra), foi obtido um número muito reduzido de dados nos anos 1970: apenas 410 ocorrências do pronome (computando-se todos os contextos – variáveis ou não). Na amostra dos anos 2000, o total foi quase quatro vezes maior: 1490 dados. Isto se deve, em parte, ao fato de que as entrevistas do NURC-SP-1970 têm uma duração bastante menor em comparação às dos anos 2000 (as primeiras duram em torno de 40 minutos, enquanto as segundas têm aproximadamente 60 minutos). Conforme se apontou no capítulo II, trata-se de uma análise que compara a fala apenas dos paulistanos com nível superior de escolaridade, uma vez que não se dispõe de *corpora* da década de 1970 com paulistanos menos escolarizados.

Excluindo-se os casos que não integram o envelope de variação aqui definido, o total de ocorrências obtidas na amostra dos anos 1970 reduz-se para apenas 340 dados; o da amostra recente cai para 1201 dados. Neste total, porém, incluem-se os dados de *vcê* e *cê* cuja pronúncia não pôde ser ouvida com precisão (“dados duvidosos”) e raras ocorrências de *ocê* e *ucê*.

VARIANTES (Contextos variáveis)	NURC-SP-1970 (Superior)			GESOL-SP-2000 (Superior)		
	Nº	%	Total	Nº	%	Total
VOCÊ	213	62,6	334	723	60,2	1170
CÊ	121	35,6	(98,2%)	447	37,2	(97,4%)
[vse] (duvidosos)	<u>01</u>	0,3	05	16	1,3	28
CÊ (duvidosos)	<u>04</u>	1,2	(0,5%)	12	1,0	(2,3%)
OCÊ	01	0,3	01	01	0,1	03
UCÊ	-	-	(0,3%)	02	0,2	(0,3%)
Total	340	100	340	1201	100	1201

Tabela 63: Distribuição Geral das variantes em 1970 e nos anos 2000

Os mesmos critérios adotados na análise sincrônica anteriormente apresentada foram adotados aqui: consideraram-se apenas as ocorrências de *você* e *cê* (as variantes majoritárias nos *corpora*) cuja pronúncia pôde ser determinada:

334 ocorrências provenientes da amostra dos anos 1970 e 1170 ocorrências na amostra mais recente.

Variáveis sociais

SEXO/GÊNERO ⁵⁷	CÊ					
	NURC-SP-1970 ⁵⁸			GESOL-SP-2000 ⁵⁹		
	Nº	%	p.r.	Nº	%	p.r.
Feminino	36/122	29,5	(.46)	265/660	40,2	.53
Masculino	85/212	40,1	(.52)	182/510	35,7	.45
(One Level)	Range: .06			Range: .08		
FAIXA ETÁRIA						
20-32	79/166	47,6	.64	175/383	45,7	.55
35-45	26/84	31,0	.40	70/296	23,6	.34
50 ou mais	16/84	19,0	.31	202/491	41,1	.56
TOTAL	121/334	36,2	Range: .33	447/1170	38,2	Range: .22
Input: 0.319 ll.= -189.221 p < 0.05				Input: 0.337 ll.= -658.104 p < 0.05		

Tabela 64: Correlação de variáveis sociais ao uso de *você* e *cê* (1970-2000)

O grupo sexo/gênero não foi selecionado como estatisticamente relevante em nenhuma das análises dos dados dos anos 1970. Uma análise univariada revela que as mulheres desfavorecem ligeiramente o emprego da variante reduzida do pronome (.46), ao passo que os homens favorecem um pouco o uso dessa variante (.52). Contrastando estes resultados com aqueles obtidos para os informantes dos anos 2000, vê-se um comportamento inverso: as mulheres é quem favorecem ligeiramente o uso da forma pronominal reduzida (.53), enquanto os homens desfavorecem-na (.45). No entanto, os valores obtidos encontram-se muito próximos do ponto neutro e próximos entre si – o que revela que o gênero do falante não se mostra fortemente correlacionado ao emprego das formas *você* e *cê* nem na década de 1970, nem atualmente, nos anos 2000.

⁵⁷ Em ambas as análises foram considerados também os grupos de fatores linguísticos: Referência do pronome; Coalescência do pronome ao verbo; Paralelismo I; Tonicidade da sílaba seguinte; Tonicidade da sílaba antecedente; Frequência do verbo – todas em versões concisas.

⁵⁸ Em ambos os períodos, foram analisados os grupos de fatores “Sexo/Gênero, Faixa etária, Referência (específica x genérica), Coalescência, Paralelismo I, Tonicidade antecedente, Tonicidade seguinte e Frequência do verbo. Na década de 1970, com exceção das variáveis Sexo/Gênero e Coalescência do pronome ao verbo, as demais foram selecionadas pelo Goldvarb como significativas de um ponto de vista estatístico.

⁵⁹ Todos os grupos de fatores considerados na análise (v. nota 58) demonstraram significância estatística.

No caso da faixa etária, como o gráfico a seguir revela, os resultados obtidos na amostra de 1970 indicam uma típica mudança em progresso em tempo aparente.

No entanto, esse indicativo de mudança não se confirma com a análise em tempo real. Após trinta anos, os resultados obtidos para a faixa etária dos informantes revela que a variante reduzida continua a ser favorecida entre os mais jovens (.55), embora com uma tendência um pouco menor do que aquela depreendida nos anos 1970; os informante mais velhos, por outro lado, assemelham-se aos mais jovens (.56); e são os indivíduos de uma faixa etária intermediária aqueles que mais desfavorecem essa variante (.34).

O mesmo padrão observado no *corpus* geral dos anos 2000 (análise sincrônica apresentada em 4.1) manteve-se na análise da subamostra de falantes escolarizados desse período: as mulheres mais velhas apresentam tendência próxima daquela apresentada pelas mais jovens; e o resultado para homens mais velhos se assemelha ao daqueles com idade mediana. Nos anos 1970, encontra-se a mesma gradação etária entre homens e mulheres: quanto mais jovens, maior a frequência de *cê*; quanto mais velhos, menor a sua frequência.

Estes resultados favorecem a interpretação de que os usos de *você* e *cê* encontram-se em variação estável entre os indivíduos escolarizados na comunidade de fala paulistana. No decorrer do tempo, não houve um aumento no emprego da variante reduzida do pronome entre os homens mais jovens, mas sim

entre os mais velhos – o que indica uma gradação etária e uma mudança geracional, pois tanto o indivíduo quanto a comunidade mudam no decorrer do tempo. Além disso, vale reiterar que o emprego dessas variantes pode estar correlacionado a outras variáveis que este estudo não pôde considerar.

O reduzido número de ocorrências do pronome na amostra dos anos 1970 não permitiu uma análise multivariada expandida da referência desse pronome. A tabela 65 a seguir apresenta as frequências da variante reduzida na amostra dos anos 1970.

REFERÊNCIA	CÊ					
	NURC-1970		GESOL-2000 (+Escarizados)		GESOL-2000 (-Escarizados)	
	Nº/Total	%	Nº/Total	%	Nº/Total	%
P2 (específico)	33/79	41,8	78/136	57,4	82/113	72,6
P1 'oculto'	9/17	52,9	43/91	47,3	54/69	78,3
P1 'expandido'	16/38	42,1	10/24	41,7	101/158	63,9
P4 'exclusivo'	11/43	25,6	60/144	41,7	32/66	48,5
Híbrido	7/15	46,7	27/98	27,6	55/114	48,2
P4 'inclusivo' [+ definido]	21/69	30,4	78/194	40,2	199/357	55,7
P4 'inclusivo' [- definido]	9/19	47,4	27/78	34,6	73/148	49,3
P3/P6 (3ª p. sg./pl.)	0/6	-	42/144	29,2	32/69	46,4
Expletivos	0/1	-	11/56	19,6	2/9	22,2
"outros"	0/4	-	04/35	11,4	13/27	48,1
Total	106/287	36,9	380/1000	38,0	643/1130	56,9

* Exceto discurso reportado, 2ª pess. pl.

Tabela 65: Distribuição de *você* e *cê* quanto à sua referenciação – (NURC-SP-1970)

Na amostra de 1970, *cê* é empregado com maior frequência (53%) quando tem uma referenciação ao próprio locutor e ocorre uma tentativa de “ocultamento do eu” (P1 ‘oculto’). Os empregos em P4 distinguem-se pelo traço [$\pm$ definido]: quanto mais abrangente é a referência, maior a frequência de *cê* (47%) – por oposição a 30% na referência P4 [+definida]. Do ponto de vista da frequência, em 30 anos, vê-se um aumento no uso de *cê* entre os informantes mais escolarizados quando sua referência é específica (P2) e nos usos genéricos referentes a uma primeira pessoa do plural exclusiva (P4 exclusivo = 1^a + 3^a pessoas) ou inclusiva [+definida]. Além disso, são raras as ocorrências do pronome com uma referenciação a uma terceira pessoa (P3/P6) nos anos 1970 – 6 dados; e o uso expletivo ocorre uma única vez na amostra. Nestes dois últimos casos, a variante plena do pronome foi categórica na amostra dos anos 1970.

Pode-se, em princípio, aventar a hipótese de que esses dois contextos de uso do pronome (P3/P6 e expletivos) poderiam ser mais recentes na língua e, por tal razão, o emprego da variante reduzida é mais restrito nestes casos atualmente. No entanto, outras questões podem ter contribuído para a ausência do pronome em tais contextos na amostra NURC-SP analisada – como o assunto da entrevista, seu menor tempo de duração e maior grau de formalidade por parte dos informantes. Relativamente aos anos 2000, vê-se que os informantes menos escolarizados empregam mais frequentemente que os escolarizados a variante reduzida em todos os tipos de referenciação considerados.

Devido à distribuição dos dados dos anos 1970 (apresentados na tabela 65) não permitir uma análise quantitativa multivariada que distinguísse os vários tipos de referenciações propostos, fez-se uma análise mais concisa da referência desse pronome, comparando a amostra dos anos 1970 à subamostra dos anos 2000 composta por informantes escolarizados. A tabela 66 a seguir traz os resultados de uma oposição entre referência específica e referência genérica nas duas épocas consideradas⁶⁰ – todos os informantes têm nível superior de escolaridade.

⁶⁰ Em ambas as amostras, além da Referência, foram analisados os grupos de fatores Sexo/Gênero, Faixa etária, Coalescência, Paralelismo I, Tonicidade seguinte, Tonicidade antecedente e Frequência do verbo – todos em versões concisas.

REFERÊNCIA	CÊ					
	NURC-SP-1970			GESOL-SP-2000 (+ Escolarizados)		
	Nº	%	p.r.	Nº	%	p.r.
Específica	33/79	41,8	.63	78/136	57,4	.69
Genérica	73/208	35,1	.45	298/829	35,9	.47
Subtotal	106/287*	36,9	Range: .18	376/965*	39,0	Range: .22
Total	121/334	36,2		447/1170	38,2	
Input: 0.319 ll.= -189.221 p < 0.05						Input: 0.337 ll.= -658.104 p < 0.05
* Exceto quando a referência não pôde ser apreendida ("Outros") e plurais.						

Tabela 66: Emprego de CÊ em relação à sua referência (específica x genérica) – (1970-2000)

Houve um ligeiro aumento nos pesos relativos em ambos os tipos de referência no decorrer de 30 anos, mas manteve-se a correlação: o emprego da variante reduzida continua a ser favorecida quando a referência do pronome é específica (.63 na década de 1970; .69 nos anos 2000), ao passo que os contextos genéricos mostram um pequeno favorecimento do uso dessa variante – conforme se observa no gráfico 9.

O cruzamento da “faixa etária” e da “referência do pronome” nas duas épocas consideradas (tabela 67) demonstra uma sistematicidade da variação: nos anos 1970, quanto mais jovem, maior a frequência de emprego do pronome reduzido, seja com referência genérica (44%), seja com referência específica (77%), sobretudo nesta última; nos anos 2000, os informantes da segunda faixa etária desfavorecem o uso de *cê*, independentemente de sua referência. Em ambas as décadas, *cê* é mais frequente com referência específica do que com referência genérica em quase todas as faixas etárias – a única exceção é a terceira faixa etária nos anos 1970.

CÊ							
Década	REF Específica				REF Genérica		
Referência	Anos 1970		Anos 2000		Anos 1970		Anos 2000
IDADE	Nº	%	Nº	%	Nº	%	Nº %
20-32 anos	17/22	77,0	31/56	55,0	56/128	44,0	129/295 44,0
35-45 anos	13/37	35,0	5/18	28,0	8/35	23,0	47/212 22,0
50 anos ou +	3/20	15,0	42/62	68,0	9/45	20,0	122/322 38,0
Subtotal	33/79	42,0	78/136	57,0	73/208	35,0	298/829 36,0

* Exceto quando a referência não pôde ser apreendida ("Outros") e 2ª p. pl.

Todos os informantes têm nível superior de escolaridade.

Tabela 67: Emprego da variante CÊ na década de 1970 e nos anos 2000 em relação à referência do pronome

Variáveis discursivo-cognitivas :

Em virtude do número de dados, desenvolveu-se apenas uma análise mais concisa do grupo “paralelismo I”⁶¹:

PARALELISMO I	CÊ					
	NURC-SP-1970			GESOL-SP-2000		
	Nº	%	p.r.	Nº	%	p.r.
Sem antecedentes	69/192	35,9	.49	237/644	36,8	.48
Antecedente CÊ	26/44	59,1	.70	134/183	73,2	.82
Antecedente VOCÊ	26/98	26,5	.42	76/343	22,2	.34
Total	121/334	36,2	Range: .28	447/1170	38,2	Range: .48
Input: 0.319 ll.= -189.221 p < 0.05						Input: 0.337 ll.= -658.104 p < 0.05

Tabela 68: Efeito do Paralelismo das variantes no uso de CÊ em 1970 e nos anos 2000

⁶¹ Grupos de fatores investigados: Sexo/Gênero, Faixa etária, Referência, Coalescência, Paralelismo I, Tonicidade antecedente, Tonicidade seguinte e Frequência do verbo. Todos os informantes possuem nível superior de escolaridade.

O paralelismo de formas pronominais atua de maneira semelhante nas duas épocas, mas apresenta um *range* maior na amostra mais recente. No caso do “paralelismo II” (de formas reduzidas), os resultados também são semelhantes em ambos os períodos considerados:

PARALELISMO II	CÊ					
	NURC-SP-1970			GESOL-SP-2000		
	Nº	%	p.r.	Nº	%	p.r.
<i>num/tá/tava</i>	22/37	59,5	.72	83/156	53,2	.65
Sem reduções	99/297	33,3	.47	364/1014	35,9	.47
Total	121/334	36,2	Range: .25			Range: .18
Input: 0.11 = -349 203.266 p < 0.05						Input: 0.376 11 = -750.305 p < 0.02

Tabela 69: Paralelismo de formas reduzidas adjacentes e emprego de *CÊ* – (1970-2000)⁶²

De maneira semelhante ao que se verifica para os grupos de fatores anteriores, a frequência do verbo mostra também uma correlação semelhante em ambas as amostras (tabela 70).

FREQUÊNCIA do VERBO	CÊ					
	NURC-SP-1970			GESOL-SP-2000		
	Nº	%	p.r.	Nº	%	p.r.
+ freq. (acima de 3%)	56/118	47,5	.60	215/469	45,8	.56
- freq. (até 3%)	65/216	30,1	.44	232/701	33,1	.46
Total	121/334	36,2	Range: .16	447/1170	38,2	Range: .10
Input: 0.319 11 = -189.221 p < 0.05						Input: 0.337 11 = -658.104 p < 0.05

Tabela 70: Emprego de *CÊ* relativamente à frequência do verbo que acompanha o pronome (1970-2000)⁶³

Todos esses resultados apontam, consistentemente, para a conclusão de que não se trata de um caso de mudança, mas de variação estável.

⁶² Grupos de fatores analisados: Sexo/Gênero, Faixa etária, Referência, Coalescência, Paralelismo I, Paralelismo II, Tonicidade antecedente e Frequência do verbo. Todos os informantes apresentam nível superior de escolaridade.

⁶³ Grupos de fatores analisados: Sexo/Gênero, Faixa etária, Referência, Coalescência, Paralelismo I, Tonicidade antecedente, Tonicidade seguinte e Frequência do verbo.

Variáveis fonético-fonológicas

Da mesma forma que na análise sincrônica desenvolvida em 4.1, faz-se necessária uma análise apenas dos contextos em que há uma sílaba no contexto seguinte (tônica ou átona). É notável, também na amostra dos anos 1970, que a tonicidade “mediana” favorece o emprego da variante reduzida mais do que os contextos tônicos (tabela 71). A fim de analisar a oposição entre os contextos tônicos e átonos, foram desenvolvidas duas análises: amalgamando-se os casos de tonicidade mediana e tônica; e excluindo-se os dados nos quais a sílaba seguinte apresenta uma tonicidade mediana.

TONICIDADE SEGUINTE	CÊ					
	NURC-SP-1970			GESOL-SP-2000		
	Nº	%	p.r.	Nº	%	p.r.
Mediana	21/36	58,3	.72	78/147	53,1	.65
Tônica	72/180	40,0	.55	264/652	40,5	.53
Átona	27/109	24,8	.37	100/326	30,7	.42
Pausas	01/09	11,1	.17	5/45	11,1	.19
	Range: .55			Range: .46		
Input: 0.343 ll.= -199.045 p < 0.02 Input: 0.371 ll.= -741.080 p < 0.02						
tônica + mediana	93/216	43,1	.57	342/799	42,8	.54
Átona	27/109	24,8	.36	100/326	30,7	.41
Subtotal	120/325	36,9		442/1125*	39,3	
Total	121/334	36,2	Range: .21	447/1170	38,2	Range: .13
Input: 0.350 ll.= -202.268 p < 0.02 Input: 0.379 ll.= -748.743 p < 0.05						
Tônica	72/180	40,0	.57	264/652	40,5	.53
Átona	27/109	24,8	.39	100/326	30,7	.43
Subtotal	99/289	34,3		364/978**	37,2	Range: .10
Total	121/334	36,2	Range: .18	447/1170	38,2	
Input: 0.333 ll.= -203.298 p < 0.05 Input: 0.367 ll.= -751.412 p < 0.05						
* Exceto pausa; ** Exceto pausas e tonicidade mediana.						

Tabela 71: Emprego de CÊ conforme a Tonicidade da sílaba seguinte (1970-2000)⁶⁴

Quanto à tonicidade da sílaba antecedente, observa-se um padrão inverso nas duas épocas: nos anos 1970, quando o contexto antecedente éônico, é

⁶⁴ Grupos de fatores analisados: Sexo/Gênero, Faixa etária, Referência, Coalescência (sim x não), Paralelismo I, Tonicidade antecedente, Tonicidade seguinte e Frequência do verbo. Todos os informantes apresentam nível superior de escolaridade.

desfavorecido o uso de *cê* (.32), o qual deixa de ser o caso nos anos 2000 (.52) – v. tabela 72. Este pode ser um indício de que este pronome ainda tinha uma proeminência acentual na década de 1970, que foi se desgastando até a atualidade. No entanto, o mesmo não se verifica no caso da tonicidade seguinte.

TONICIDADE ANTECEDENTE	CÊ					
	NURC-SP-1970			GESOL-SP-2000		
	Nº	%	p.r.	Nº	%	p.r.
Tônica	13/55	23,6	.32	87/212	41,0	.52
Átona	70/193	36,3	.52	193/605	31,9	.43
Pausas	38/86	44,2	.57	167/353	47,3	.61
Subtotal	121/334	36,2	Range: .25	447/1170	38,2	Range: .18
Input: 0.348 l.l.= - 204.030 p < 0.05						Input: 0.373 l.l.= - 742.437 p < 0.05
Tônica	13/55	23,6	.35	87/212	41,0	.57
Átona	70/193	36,3	.54	193/605	31,9	.47
Subtotal	83/248	33,5	Range: .19	280/817	34,3	Range: .10
Total	121/334	36,2		447/1170	38,2	
Input: 0.329 l.l.= -204.164 p < 0.02						Input: 0.332 l.l.= -742.570 p < 0.05

Tabela 72: Emprego de *CÊ* conforme a Tonicidade da sílaba antecedente (1970-2000)⁶⁵

O gráfico 12 ilustra o último resultado apresentado na tabela anterior:

⁶⁵ Grupos de fatores analisados: Sexo/Gênero, Faixa etária, Referência, Coalescência (sim x não), Paralelismo I, Tonicidade antecedente, Tonicidade seguinte e Frequência do verbo. Todos os informantes apresentam nível superior de escolaridade;

A análise da presença de constituintes semelhantes à forma reduzida do pronome demonstra a atuação do OCP em ambas as épocas, mas tal efeito é ainda mais saliente nos anos 2000 (tabela 72). É preciso ponderar, entretanto, que nos anos 1970 houve uma pequena quantidade de dados e que os resultados a seguir são provenientes de uma análise univariada.

PRESENÇA DE /SE/	CÊ					
	NURC-SP-1970			GESOL-SP-2000		
	Nº	%	p.r.	Nº	%	p.r.
Sem /SE/	109/295	36,9	(.50)	428/1067	40,1	.52
Com /SE/	12/39	30,8	(.44)	19/103	18,4	.27
Total	121/334	36,2	Range: .07	447/1170	38,2	Range: .25
(One Level)			Input: 0.373 LL= - 746.526 p < 0.05			

Tabela 73: Emprego de CÊ e presença de constituintes /SE/ adjacentes ao pronome (1970-2000)⁶⁶

O escasso número de ocorrências de pronomes de segunda pessoa do plural em ambas as amostras também prejudicou a análise da variável independente “Flexão do pronome”. As frequências da variante reduzida encontradas para o singular e o plural são muito próximas em ambas as décadas (tabela 74).

FLEXÃO DO PRONOME	CÊ(s)			
	NURC-SP-1970		GESOL-SP-2000	
	Nº	%	Nº	%
Singular	107/294	36,4	437/1140	38,3
Plural	14/40	35,0	10/30	33,3
Total	121/334	36,2	447/1170	38,2

Tabela 74: Emprego da variante reduzida do pronome quanto à sua flexão (1970-2000)

⁶⁶ Na análise da subamostra dos anos 2000, foram considerados também os grupos de fatores sociais Sexo/Gênero e Faixa etária. Na análise dos dados dos anos 1970 em que esses dois grupos de fatores foram também considerados, apenas a faixa etária demonstrou significância estatística. Todos os informantes apresentam nível superior de escolaridade;

Variáveis Morfossintáticas

A interpolação de elementos entre o pronome e a variante reduzida era mais frequente na década de 1970 (45%) do que atualmente (36%). Além disso, embora não se observem grandes diferenças em relação ao contexto em que pronome e verbo encontram-se adjacentes (em ambos os períodos, os pesos relativos estão próximos ao ponto neutro), a interpolação era favorecida na década de 1970 (.60), o que deixa de ocorrer nos anos 2000 (.46). Esses resultados podem ser tomados como indício da cliticização de *cê* nessa comunidade de fala.

COALESCÊNCIA	CÊ					
	NURC-SP-1970			GESOL-SP-2000		
	Nº	%	p.r.	Nº	%	p.r.
Adjacente	93/265	35,1	.48	373/926	40,3	.51
Interpolação	26/57	45,6	.60	65/178	36,5	.46
Subtotal	119/322	37,0		438/1104	39,7	
Total	121/334	36,2	Range: .12	447/1170	38,2	Range: .05
One Level			Input: 0.383 ll.= -753.719 p < 0.02			

Tabela 75: Coalescência do pronome ao verbo em 1970 e nos anos 2000⁶⁷

⁶⁷ Grupos de fatores considerados na análise: Sexo/Gênero, Faixa etária, Referência, Coalescência, Paralelismo I, Tonicidade antecedente, Tonicidade seguinte e Frequência do verbo.

Finalmente, quanto ao tipo de oração, independente da época considerada, a variante reduzida *cê* ocorre com maior frequência nas estruturas interrogativas do que nas declarativas (tabela 76). Não foi possível, todavia estabelecer uma oposição confiável entre as estruturas sintáticas com a posição do complementizador preenchida (interrogativas *qu-que*) e as demais interrogativas, pois as primeiras foram muito infrequentes nas amostras.

TIPO DE ORAÇÃO	CÊ			
	NURC-SP-1970		GESOL-SP-2000	
	Nº	%	Nº	%
Declarativa	93/284	32,7	380/1055	36,0
interrogativa <i>qu-que</i>	1/2	50,0	7/18	38,9
outras interrogativas	27/48	56,2	60/97	61,9
Total	121/334	36,2	447/1170	38,2

Tabela 76: Emprego de *Cê* conforme o tipo de oração (1970-2000)⁶⁸

⁶⁸ Apenas informantes com nível superior de escolaridade.

Conclusão

Nesta pesquisa, desenvolveu-se um estudo sociolinguístico variacionista do pronome de segunda pessoa na cidade de São Paulo, considerando-se dois períodos de tempo (a década de 1970 e os anos 2000). Na comunidade de fala paulistana, as formas alternativas majoritariamente empregadas são: a variante plena *você(s)* e a forma foneticamente reduzida *cê(s)* – o que a situa entre as comunidades que apresentam um subsistema pronominal do tipo *V* (cf. classificação de Scherre, 2010-2013).

Em uma análise multivariada, o grupo de fatores “flexão do pronome” não apresentou significância estatística – que provavelmente se deve ao número de ocorrências do pronome de segunda pessoa do plural ser extremamente reduzido (52 dados) nas amostras analisadas. Da distribuição das variantes, nota-se que a reduzida é menos frequente quando se trata do plural. No entanto, não é possível fazer maiores generalizações a respeito do pronome plural. Considerando-se que uma análise separada, apenas dos dados singulares desse pronome, não se revelou distinta daquela em que se incluíram os dados de plural, ambos foram analisados em conjunto em todas as análises quantitativas, exceto no estudo da referência, visto que a segunda pessoa do plural sempre possui um referente específico.

Os discursos reportados, por sua vez, também ocorrem em número reduzido e, geralmente, apresentam referência específica. Com referentes específicos, discursos reportados desfavorecem o uso de *cê(s)*. O controle desse grupo de fatores na pesquisa revelou-se importante na medida em que havia interação entre o tipo de discurso e a referência, sendo necessário a exclusão dos dados de discursos reportados para uma análise mais precisa da referência do pronome.

O estudo sincrônico da amostra mais recente (GESOL-SP-2000), por apresentar um número mais robusto de ocorrências do pronome, permitiu uma análise mais profícua das hipóteses aqui aventadas. Nesta análise, as variantes distribuíram-se de maneira bastante equilibrada entre os paulistanos: 52% de *você(s)* e 48% de *cê(s)*. Esse resultado indica que o processo de gramaticalização

deste pronome encontra-se “mais avançado” nas comunidades mineiras de Arcos (Gonçalves, 2008) e Belo Horizonte (Peres, 2006) – nas quais a variante reduzida é mais frequente do que a plena.

Ainda com relação à distribuição das variantes entre comunidades de fala, verifica-se que São Paulo aproxima-se do padrão observado por Ramos (2000) em Ouro Preto (OP), que se mostrou mais “conservadora” do que Belo Horizonte (BH): enquanto em BH, a forma reduzida foi mais frequente, em Ouro Preto, houve maior frequência de *você*. O avanço do processo de gramaticalização desse pronome, portanto, parece não ter correlação direta com as características sócio-econômicas da região, uma vez que, neste sentido, São Paulo estaria mais próxima a BH do que a OP, sobretudo por se tratar de uma investigação desenvolvida dez anos após o estudo empreendido por Ramos (2000) em Minas Gerais.

Uma vez que a amostra dos anos 1970 (NURC-SP-1970) é composta apenas por informantes com nível superior de escolaridade, uma comparação entre as duas épocas foi realizada com uma subamostra dos anos 2000 (apenas com os informantes que apresentam nível superior de escolaridade). Em ambos os períodos, a forma pronominal plena se mostrou mais frequente: 64% (contra 36% de *cê*) na década de 1970; e 62% (contra 38% de *cê*) em 2000. Houve, portanto, um aumento de apenas 2% no uso da variante reduzida em trinta anos¹. Estes resultados, portanto, evidenciam estabilidade dessa variável na cidade de São Paulo entre os informantes mais escolarizados.

Um estudo sincrônico dos dados dos anos 1970 demonstrou um típico padrão de mudança em progresso detectada pelo constructo do tempo aparente: *cê* tendia a ser empregada entre os indivíduos mais jovens e era desfavorecido quanto maior a faixa etária do informante. A análise em tempo real, contudo, não confirmou tal predição, uma vez que: não só o uso das variantes manteve-se relativamente idêntico em ambos os períodos, como a faixa etária do informante nos anos 2000 demonstrou um padrão distinto daquele observado nos anos 1970.

¹ Na década de 1970, o *input* da variante reduzida foi de 0.319, valor muito próximo ao obtido nos anos 2000 (0.337).

Conclusão

Independentemente da escolaridade do indivíduo, nos anos 2000, os mais jovens favorecem o emprego da variante reduzida, os medianos desfavorecem essa variante, enquanto os mais velhos aproximam-se da neutralidade. Estes últimos informantes, porém, diferem entre si relativamente ao sexo/gênero: as mulheres acima de 50 anos de idade aproximam-se dos mais jovens, favorecendo o uso da variante pronominal reduzida; enquanto os homens dessa mesma faixa etária aproximam-se daqueles de faixa etária mediana (entre 35 e 45 anos de idade), desfavorecendo o uso dessa variante.

O mesmo padrão também foi detectado por Peres (2006) em Belo Horizonte – o que remete a uma gradação etária que tem uma relação também com o sexo/gênero do indivíduo, e que pode ser explicada pela noção de mercado linguístico (cf. Paiva & Duarte, 2003). De um lado, indivíduos de faixa etária mediana, por estarem ativos no mercado de trabalho, poderiam assumir uma postura mais monitorada durante a gravação das entrevistas, desfavorecendo, assim, a redução do pronome, visto que, ainda que nenhum prestígio ou estigma seja associado a nenhuma das variantes em estudo, a forma reduzida tem um caráter mais “informal”, ativado em situações em que a auto-monitoração da fala pode ser menor. Com relação às diferenças de gênero detectadas na terceira faixa etária, o mercado linguístico também pode revelar atuação: muitas das mulheres nesta faixa etária, por exemplo, sempre exerceram a função de donas-de-casa, não sofrendo pressão direta do mercado linguístico. Além disso, os indivíduos mais velhos (sobretudo as mulheres) geralmente reportaram vários fatos relacionados à sua infância e juventude, o que pode ter propiciado maior informalidade na ocasião da entrevista, neutralizando o chamado “paradoxo do observador” (Labov, 2006 [1966]). Conforme foi mencionado no capítulo IV, estudos posteriores que considerem variáveis tais como “grau de formalidade”, “assunto da conversa” e “ocupação profissional” poderão trazer maiores elucidções a este respeito.

No estudo que ora se apresenta, observa-se que o uso de *você* e *cê* na cidade de São Paulo é controlado tanto por questões sociais e interacionais (extralinguísticas), como por questões linguísticas. Com relação ao encaixamento

social desse fenômeno variável na comunidade de fala paulistana, todos os grupos de fatores de natureza social investigados mostraram-se significativos. Além da faixa etária, as mulheres favorecem o emprego da variante inovadora *cê* com referência específica, e demonstram ligeiro desfavorecimento quando a referência desse pronome é genérica. A escolaridade também se mostrou atuante no uso variável desse pronome entre os paulistanos: falantes com nível superior de escolaridade desfavorecem o emprego de *cê*, que é favorecido entre aqueles que cursaram até o nível Fundamental/Médio.

Esses resultados vão ao encontro de nossas expectativas iniciais: há maior tendência de emprego da forma mais recente *cê* (sobretudo quando a referência é específica) entre as mulheres da primeira e da terceira faixas etárias – tal fato, porém, só é evidenciado em análises separadas dos dados provenientes da fala de homens e mulheres. Em relação à variável escolaridade, podem-se interpretar os resultados da seguinte maneira: por não se tratar de uma forma veiculada pela escola e pelos compêndios gramaticais, *cê* tende a ser evitado pelos informantes quanto maior o seu nível de escolaridade.

Dentre as questões interacionais, verifica-se ligeiro favorecimento no emprego de *cê* quando os interlocutores são mulheres – o que pode estar relacionado ao fato de que estas tendem a assumir uma atitude mais “solidária” do que os homens em interações simétricas com relação ao gênero. Esse resultado vai em direção oposta ao observado em Brasília em relação ao *tu* (considerada uma forma de “solidariedade entre os interlocutores”), que é mais empregada pelos homens em relações simétricas. O grau de intimidade entre os interlocutores também se mostrou correlacionado à seleção das variantes, sobretudo quando o pronome tem referência específica. Os informantes tendem a empregar *cê* na interação com parentes e amigos, desfavorecendo-o nas interações com colegas e conhecidos – resultado semelhante ao que Martins (2010) obtém para o *tu* em Tefé (Amazonas). O seguinte padrão pôde ser estabelecido: parentes > amigos > amigos de amigos > colegas/conhecidos². Os documentadores demonstraram um

² As únicas exceções são observadas em três entrevistas com pessoas que não se conheciam previamente ao contato para a gravação da entrevista. Estes demonstraram grande favorecimento ao uso de *cê*, superando os

padrão semelhante, mas a tendência ao emprego da variante reduzida foi maior entre amigos e amigos de amigos do que na interação com familiares. Uma análise dos dados com referência específica contrastando a fala de 9 documentadores com a fala de 36 informantes, por sua vez, demonstra que a redução do pronome tende a ocorrer entre os informantes, provavelmente devido ao maior monitoramento da fala por parte dos documentadores e à interferência da escrita (caso estes portassem anotações na ocasião da gravação da entrevista sociolinguística).

No que diz respeito ao encaixamento linguístico dessa variável, há tanto restrições de caráter (quase) categórico com relação ao emprego das variantes, quanto restrições variáveis. Como os diversos estudos já realizados demonstraram, a função sintática de sujeito é a única que não oferece restrições à alternância *você/cê*. As demais funções sintáticas (objeto direto, complemento de preposição), entretanto, impedem ou inibem fortemente o emprego da variante reduzida desse pronome: há apenas duas ocorrências de *cê* objeto de preposição (*p'cê*) em mais de cem ocorrências do pronome nesse ambiente sintático. O mesmo ocorre com construções de tópico e a maioria das clivadas, ou ainda quando o pronome encontra-se isolado, sob o escopo de um advérbio, posposto ao verbo, em coordenação ou é pronunciado de maneira enfática. Tais restrições ao emprego de *cê* corroboram a hipótese de que este estaria se cliticizando no PB atualmente (cf. Vitral, 1996, 2006b).

As restrições linguísticas variáveis que foram estudadas visavam a analisar quatro questões fundamentais:

- (i) questões de natureza fonético-fonológicas, explicadas pelo *Princípio do Contorno Obrigatório (OCP)* – analisado aqui como uma regra variável (Guy & Boberg, 1997);
- (ii) a hipótese de que *cê* estaria se cliticizando;

valores obtidos para os familiares e amigos. Como dois destes informantes eram mulheres da primeira e terceira faixas etárias, respectivamente, e como não há um equilíbrio entre as entrevistas, tal fato pode se dever à sobreposição dos grupos de fatores faixa etária e sexo/gênero ou a questões idiossincráticas. Por essa razão as três entrevistas com informantes que não se conheciam previamente foram descartadas, posteriormente, no estudo desse grupo de fatores.

(iii) a atuação de questões de natureza discursivo-cognitiva na seleção das variantes (tais como o Paralelismo e a Frequência); e

(iv) a referência do pronome (analisada aqui de um ponto de vista semântico, discursivo e pragmático) e sua relação com o processo de gramaticalização desse pronome.

Em relação à questão (i), uma análise da tonicidade das sílabas antecedente e seguinte ao pronome demonstrou que a variante reduzida *cê* tende a ser empregada quando é antecedida ou seguida por uma sílaba tônica, sendo desfavorecida quando há sílabas átonas adjacentes. Esses resultados confirmam os testes acústicos efetuados por Vitral (2006a) e Ciríaco et al. (2004), para quem *cê* aproxima-se de elementos átonos de um ponto de vista fonético. Se, de fato, *cê* for uma forma menos proeminente, ela tenderia a ocorrer em adjacência a sílabas tônicas e ser desfavorecida quando há sílabas átonas circundantes, evitando-se tanto “choques de acento”, quanto “lapsos de acento” – cf. *Princípio de Alternância Ritmica* (Selkirk, 1984; apud. Abousalh, 1997). Tal processo mostrou-se mais atuante no contexto fônico seguinte, do que em relação ao contexto antecedente, uma vez que a primeira sílaba do pronome pleno também é átona, tais resultados não explicam a preferência pela variante reduzida quando ao pronome antecede-se uma sílaba tônica.

Além disso, quando sílabas adjacentes ao pronome são foneticamente idênticas ou semelhantes à variante reduzida *cê*, esta última tende a ser fortemente evitada. Todos estes resultados indicam a atuação do *Princípio do Contorno Obrigatório (OCP)* – segundo o qual sequências idênticas ou semelhantes seriam “vetadas” (McCarthy, 1986; apud. Suzuki, 1998). A exemplo do que Guy & Bober (1997) propõem para o estudo do apagamento de /-t, -d/ em final de palavra no inglês da Filadélfia, a análise desse princípio como uma regra variável também se mostrou adequada ao estudo aqui desenvolvido.

Com relação à hipótese de cliticização de *cê*, além das questões de caráter categórico, *cê* tende ser evitado quando há interpolação de constituintes entre o pronome e o verbo, sobretudo se tais constituintes têm um caráter móvel dentro do

sintagma e/ou apresentam maior extensão fônica. No entanto, contrariamente à hipótese aventada por Ramos (1997), as interrogativas *wh-que* não se correlacionam ao emprego das variantes em estudo. Uma oposição entre sentenças declarativas e interrogativas, por sua vez, mostrou-se mais explicativa: as interrogativas favorecem o uso da variante reduzida.

Dentre as questões discursivo-cognitivas, o Paralelismo revelou forte atuação tanto na versão tradicional (que analisa a sequência de formas ao longo de orações em série), quanto na versão que analisa a presença de outras reduções adjacentes ao pronome. No primeiro caso, verifica-se que: uma vez pronunciada uma variante (plena ou reduzida), a mesma tende a continuar sendo empregada nas orações seguintes. No segundo tipo de paralelismo, nota-se que reduções tendem a levar a outras reduções. Outra questão cognitiva importante é a frequência do verbo seguinte: verbos mais frequentes (acima de 3% no *corpus*) favorecem a redução do pronome, enquanto verbos menos frequentes desfavorecem essa variante.

Dentre as hipóteses de natureza linguística, a referência do pronome recebeu atenção especial no presente estudo. Embora a análise dessa variável independente tenha sido objeto de investigação em diversos estudos que tratam dos pronomes de segunda pessoa no PB, ela é analisada apenas de um ponto de vista binário: determinada *versus* indeterminada (ou específica *versus* genérica), sem que se levem em conta nuances que a referência genérica ou indeterminada pode assumir no discurso.

Foram recapitulados vários tratamentos dados à Referência na literatura que investiga pronomes pessoais. Uma oposição ternária é encontrada em Vitral & Ramos (2006 [1999]) – que, além da oposição “específico” *versus* “genérico”, admitem também um emprego expletivo do *você/cê*. Além destes, versões mais detalhadas para a abordagem da referência foram propostas por Corradello (1997) e Sousa (2008), que não chegaram a desenvolver análises variacionistas propriamente.

Neste estudo, foram estabelecidos nove tipos de empregos distintos do pronome *você/cê* no que diz respeito à sua referência, fundamentados em “pistas referenciais” presentes no contexto discursivo (Micheletti, 2000) ou em inferências pragmáticas que permitiram “delimitar” a referenciação do pronome quando genérico, de acordo com as pessoas envolvidas no processo de referenciação (Corradello, 1997). Refinou-se, assim, tanto a classificação proposta por Corradello (1997), quanto aquela defendida por Nascimento (2008a, b).

A partir das análises de Corradello (1997), Sousa (2008) e Nascimento (2008a, b), foi proposto que, além da referência específica ao interlocutor (P2), o pronome em estudo também podia ter uma referenciação: ao próprio locutor (P1); ao locutor mais um conjunto de pessoas dentre as quais o interlocutor não se inclui necessariamente (P4 ‘exclusivo’); a qualquer pessoa / às pessoas em geral (P4 ‘inclusivo’); a uma terceira pessoa (P3/P6); consideraram-se também “usos híbridos” (quando a referência do pronome é simultaneamente específica e genérica) e casos em que o pronome não apresenta uma referência, i.e., usos expletivos – como sugerem Vitral & Ramos (2006 [1999]) e Duarte (2003).

Os empregos que remetiam ao próprio locutor (P1) foram subdivididos entre aqueles em que havia um processo de “ocultamento do eu”, com vistas a preservar a face do locutor (P1 ‘oculto’) e aqueles em que isso não se verificava (P1 ‘expandido’). Os empregos do tipo P4 ‘inclusivo’ também foram subdivididos de acordo com o traço [$\pm$ definido].

Os resultados obtidos evidenciaram que não há correspondência entre “perda de substância fonológica” e “perda de conteúdo semântico” por que passariam os itens em processo de gramaticalização, uma vez que a variante reduzida tende a ser empregada com referência específica ao interlocutor e é fortemente desfavorecida nos empregos considerados “expletivos”. Esse resultado se mostra importante tanto para a teoria da gramaticalização, quanto para a discussão acerca da formação de itens expletivos no PB. Em primeiro lugar, a correlação entre *erosão* fonética e *perda* semântica não se aplica indistintamente a todos os casos de gramaticalização. Em segundo lugar, de um ponto de vista

qualitativo, embora se possam tecer críticas à denominação “expletivo” para o pronome *você/cê* que integra construções existenciais com o verbo “ter”, esses contextos comportaram-se de maneira distinta daqueles em que o pronome tem referência indiscutivelmente genérica.

Além disso, uma análise expandida da referência/referenciação desse pronome demonstra que, quando esta se encontra no âmbito da enunciação, remetendo ao próprio locutor ou ao interlocutor em específico (primeira e segunda pessoas = P1 e P2), há maior tendência de emprego da variante reduzida *cê*. Ao contrário, quando há referenciação no discurso a uma terceira pessoa (P3/P6) – uma *não-pessoa* (Benveniste, 1991) – o uso dessa variante é desfavorecido. Nos casos de P1 em que há um “ocultamento do eu”, o uso da variante reduzida do pronome é ainda mais favorecido, aproximando-se aos valores obtidos para a referência específica. Quando o pronome tem por referente discursivo “as pessoas em geral”, observa-se pequeno desfavorecimento ao emprego da variante reduzida (.46). A referenciação híbrida, por sua vez, mostrou-se próxima ao ponto neutro.

Foi depreendido, então, o seguinte percurso de implementação da variante reduzida *cê* no que diz respeito à sua referenciação:

P2 > P1 / P4 ‘exclusivo’ > Híbrido > P4 ‘inclusivo’ > P3 / P6 > Expletivo

Essa análise expandida da referência permitiu apreender a atuação de questões discursivas que se encontravam camufladas em uma análise mais concisa desse grupo de fatores (referência “específica” *versus* referência “genérica”), que não captava, por exemplo, a oposição “pessoa” (P2 e P1) e “não pessoa” (P3 / P6). Além disso, testes de qui-quadrado confirmaram que uma análise expandida da referência é preferível a análises concisas.

A análise em tempo real, por sua vez, não permitiu observar distinção entre todos os tipos de referência anteriormente apresentados. Neste caso, fez-se apenas uma oposição binária. Os resultados indicam que a referência específica favorece o emprego de *cê* e a genérica desfavorece essa variante tanto nos anos 1970, quanto nos anos 2000 entre os indivíduos mais escolarizados.

Conclusão

Apesar de a análise dos distintos graus de genericidade da referência não ter sido possível nos anos 1970, isso não significa necessariamente uma ampliação dos contextos de uso desse pronome nos anos 2000 – tal fato pode se dever exclusivamente a questões de amostragem ou ao pequeno número de dados obtidos naquela amostra.

Relativamente aos demais grupos de fatores, também se depreende uma atuação muito similar nas duas épocas – o que indica se tratar de uma variação estável nessa comunidade de fala. A exceção mais significativa encontrada entre as duas amostras é a atuação da variável coalescência do pronome ao verbo. Na década de 1970, a interpolação favorecia o uso de *cê*, enquanto nos anos 2000 esse contexto passa a desfavorecer o uso dessa variante. Esse resultado confirma a hipótese de que a variante pronominal reduzida estaria passando por um processo de cliticização (Vital, 2006b), de modo similar aos demais clíticos do português (Martins, 1996: apud. Vital, 2006b). Embora os contextos com interpolação não sejam categóricos ao uso de *cê* atualmente, eles tornaram-se desfavoráveis à redução do pronome – o que não se observa na década de 1970.

Em suma, os resultados desta pesquisa devem contribuir para um melhor entendimento do emprego dos pronomes de segunda pessoa na comunidade de fala paulistana, somando-se àqueles encontrados por outros pesquisadores em outras comunidades de fala. Além disso, esta dissertação apresenta desenvolvimentos tanto no campo da Sociolinguística Variacionista, quanto em áreas afins, como a Dialetoлогия e a Gramaticalização.

Referências Bibliográficas

- ABOUSALH, Elaine Silveira Ferreira. (1997). *Resolução de choques de acento no português brasileiro: elementos para uma reflexão sobre a interface sintaxe-fonologia*. Dissertação (Mestrado). Campinas/SP: Instituto de Estudos da Linguagem, IEL/UNICAMP.
- ALMEIDA, Napoleão Mendes de. (2005) *Gramática Metódica da Língua Portuguesa*. 45. ed. São Paulo: Saraiva.
- ALVES, Nilton Antônio (1998). *As Formas VOCÊ e CÊ e a Indeterminação do sujeito no Português Brasileiro*. 1998. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte.
- ANDRADE, Adriana Lília V. S. de. (2004) *A variação de você, cê, ocê no português brasileiro falado*. 2004. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade de Brasília, Brasília.
- AVELAR, Juanito. (2006) “The Status of (Supposed) Expletive in Brazilian Portuguese Existential Clauses”. In: 20th Symposium on Romance Linguistic. Vrije Universiteit Amsterdam.
- BAILEY, Guy. (2002) “Real and apparent time”. In: CHAMBERS, J. K., TRUDGILL, P. & SCHILLING-ESTES, N. (eds). *The Handbook of Language, Variation and Change*. Malden / Oxford: Blackweel.
- BENVENISTE, Emile. (1991) “A Natureza dos Pronomes”. In.: *Problemas de Linguística Geral I*. (Trad. de Novak, M. G. & Neri, M. L.). Campinas/SP: Pontes: Editora da UNICAMP.
- BOLÍVAR, Thiago M. V. (2008) *A Forma VOCÊ em interações comerciais em Porto Alegre, RS*. 2008. B638 f. Dissertação (Mestrado em Linguística) – Instituto de Estudos da Linguagem (IEL), UNICAMP, Campinas/SP.

- BRAGA, Maria Luiza (1991) “As sentenças clivadas no português falado do Rio de Janeiro”. In: *A variação no português do Brasil*. Revista do Instituto de Letras da Universidade Federal do Rio Grande do Sul. Organon 18, vol. 5, n. 18, p. 109-125.
- BROWN, Roger & GIMAN, Albert. (1972) “The pronouns of Power and Solidarity”. In: GIGLIOLI, P. P. (ed.) *Language and Social Context*. New York: Penguin Books, p. 252-282 [1960].
- BYBEE, Joan. (2002). “Cognitive process in Grammaticalization”. In. TOMASELLO, M. (ed.). *The new psychology of language: Cognitive and functional approaches to language structure* (vol. 2). New Jersey: Lawrence Erlbaum Associates.
- BYBEE, Joan. (2003) “Mechanisms of Change in Grammaticalization: The role of Frequency”. In.: JANDA, R. & BRION, J. (Eds.) *Handbook of Historical Linguistics*, Oxford: Blackwell.
- CALLOU, Dinah & AVELAR, Juanito. (2007) “From possessive to existencial: change and preservation in the history of Portuguese”. In: *NWAV 36*, University of Pennsylvania, Philadelphia.
- CÂMARA JR., Joaquim Mattoso. (1980) *Estrutura da língua portuguesa*. Rio de Janeiro: Petrópolis.
- CANÇADO, Márcia. (2005) *Manual de Semântica: noções básicas e exercícios*. Belo Horizonte, Editora da UFMG.
- CASTILHO, Ataliba T. de & PRETI, Dino. (orgs.) (1986-1990) “Linguagem falada culta na cidade de São Paulo” (vols. 1, 2, 3). São Paulo : T.A. Queiroz [NURC/SP – “Projeto de estudo da norma linguística urbana culta de São Paulo”]. <http://www.fflch.usp.br/dlc/nurc/historico.htm>
- CASTILHO, Ataliba Teixeira de. (2006) “Proposta Funcionalista de Mudança Linguística: os processos de lexicalização, semantização, discursivização e gramaticalização na constituição das línguas”. In.: LOBO, T., RIBEIRO, I., CARNEIRO, Z. & ALMEIDA, N. (orgs.). *Para a história do Português Brasileiro – vol. VI: Novos dados, novas análises* (Tomo I). Salvador – BA: EDUFBA.

Referências Bibliográficas

- CASTILHO, Ataliba Teixeira de. (1997) “A Gramaticalização”. In.: *Revista de Estudos Linguísticos e Literários*, nº. 19, p. 25-64.
- CASTILHO, Ataliba Teixeira de. (2009) “An approach to language as a complex system”. In.: *História do Português Paulista – Série Estudos*, vol. I. Campinas: UNICAMP/ Publicações IEL.
- CERVONI, Jean. (1989) “A dêixis”. In.: *A Enunciação*. São Paulo: Ática.
- CHAMBERS, J. K. (1995) *Sociolinguistic theory: linguistic variation and its social significance*. Oxford, UK; Cambridge, USA: Blackwell.
- CHAMBERS, J. K. “Patterns of variation including change”. (2002) In: CHAMBERS, J. K., TRUDGILL, P. & SCHILLING-ESTES, N. (eds). *The Handbook of Language, Variation and Change*. Malden / Oxford: Blackwell.
- CHESHIRE, J. (2002) “Sex and Gender in Variationist Research”. In.: CHAMBERS, J. K., TRUDGILL, P. & SCHILLING-ESTES, N. (eds). *The Handbook of Language, Variation and Change*. Malden / Oxford: Blackwell.
- CINTRA, Lindley. (1972) *Sobre as Formas de Tratamento na Língua Portuguesa*. Lisboa: Horizonte.
- CIRÍACO, Larissa S.; VITRAL, Lorenzo T. & REIS, César. (2004) “Intensidade e duração de formas reduzidas no português brasileiro”. In.: *Revista de Estudos da Linguagem*, v. 12, n.2, p. 143-147.
- COLLISCHONN, Gisela. (1996) “A Sílabas em Português”. In.: BISOL, L. (org.) *Introdução a Estudos de Fonologia do Português Brasileiro*. Porto Alegre: EDIPUCRS.
- CORRADELLO, Elaine de Fátima A. (1997) *Quem é você? – análise de um pronome pessoal*. 1997. Dissertação (Mestrado). Instituto de Estudos da Linguagem (IEL – UNICAMP), Campinas/SP.
- CUNHA, Celso & CINTRA, Lindley (2007) *Nova gramática do português contemporâneo* (4ª.ed.). Rio de Janeiro: Lexikon Editora Digital.
- DIAS, Edilene Patrícia (2007) *O uso do tu no português brasileiro falado*. Dissertação (Mestrado). Universidade de Brasília (UnB), Brasília.

Referências Bibliográficas

- DUARTE, Maria Eugênia Lamoglia (1996) “Do pronome nulo ao pronome pleno: a trajetória do sujeito no português do Brasil”. In.: ROBERTS, I. & KATO, M. A (orgs.) *Português Brasileiro – Uma viagem diacrônica*. Campinas, SP: Editora da UNICAM.
- DUARTE, Maria Eugênia Lamoglia (2003) “O Sujeito Expletivo e as Construções existenciais”. In.: RONCARATI, C. et.al. (orgs.) *Português Brasileiro: contato linguístico, heterogeneidade e história*. RJ: 7 Letras.
- DUARTE, Maria Eugênia L. & LOPES, Célia Regina dos S. (2003) “De *Vossa Mercê* a *você*: análise da pronominalização de nominais em peças brasileiras e portuguesas setecentistas e oitocentistas”. In.: BRANDÃO, S. F. & MOTA, M. A. (orgs.) *Análise Contrativa de Variedades do Português: primeiros estudos*. Rio de Janeiro.
- FARACO, Carlos Alberto (1996) “O tratamento *você* em português: uma abordagem histórica”. In: *Fragmenta*, Curitiba, UFPR, 13^a ed., p. 51-82.
- FERNÁNDEZ, Francisco Moreno. (1997a) “Sociolingüística, Sociología del language y Etnografía de la comunicación”. In: In.: *Principios de Sociolingüística y Sociología del lenguaje*. Barcelona: Editorial Ariel, S.A.
- FERNÁNDEZ, Francisco Moreno. (1997b) “Teoría y métodos de la sociolingüística variacionista”. In.: *Principios de Sociolingüística y Sociología del lenguaje*. Barcelona: Editorial Ariel, S.A.
- FIORIN, José Luiz. (2008) *Em busca do sentido: estudos discursivos*. São Paulo: Contexto.
- FREGE, Gottlob. (1978) *Lógica e filosofia da linguagem*. São Paulo: Cultrix.
- GALVES, Charlotte. C. (1996) “O enfraquecimento da concordância no português brasileiro”. In.: KATO, M. A. & ROBERTS, I. (orgs.) *Português Brasileiro – Uma viagem diacrônica*. Campinas, SP: Editora da UNICAM.
- GONÇALVES, Clézio Roberto. (2008) *Uma abordagem sociolinguística do uso das formas você, ocê e cê no português*. 2008. Tese (Doutorado). Departamento de Linguística, FFLCH/USP, São Paulo.

- GONÇALVES, Sebastião C. L., et.al. (2007) “Tratado Geral sobre Gramaticalização”. In.: GONÇALVES, S. C. L., et.al. (orgs.). *Introdução à Gramaticalização*. São Paulo: Parábola.
- GONÇALVES, Sebastião C. L. & CARVALHO, Cristina dos. S. (2007) “Critérios de Gramaticalização”. In.: GONÇALVES, S. C. L., et.al. (orgs.). *Introdução à Gramaticalização*. São Paulo: Parábola.
- GUY, Gregory R. (2007) “VARBRUL: Análise Avançada”. In.: GUY, G. R. & ZILLES, A. *Sociolinguística Quantitativa – Instrumental de Análise*. São Paulo: Parábola, [1988]. [Trad. de Ana Zilles].
- GUY, Gregory R. (2007) “Introdução à análise quantitativa da variação linguística”. In.: GUY, G. R. & ZILLES, A. *Sociolinguística Quantitativa – Instrumental de Análise*. São Paulo: Parábola, [1993]. [Trad. de Ana Zilles].
- GUY, Gregory R. (Ms., s/d). *Language variation and linguistic theory*. (Ms., inédito).
- GUY, Gregory R. & BOBERG, Charles. (1997) “Inherent variability and the obligatory counter principle”. In.: *Language Variation and Change*, v. 9, Cambridge University Press, p. 149-164.
- GUY, Gregory R. & ZILLES, Ana. (2007) *Sociolinguística Quantitativa: instrumental de análise*. São Paulo: Parábola.
- HEINE, B.; CLAUDI, U. & HÜNNEMEYER, F. (1991) *Grammaticalization: a conceptual framework*. Chicago and London: University of Chicago, Press.
- HERÊNIO, Kerlly K. P. (2006) “Tu” e “você” em uma perspectiva intra-linguística. 2006. Dissertação (Mestrado). Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia.
- HOPPER, Paul & BYBEE, Joan. (2001) “Introduction to frequency and the emergency of linguistic structure”. In: BYBEE, J. & HOPPER, P. *Frequency and the Emergence of Linguistic Structure*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- HOPPER, Paul & TRAUGOTT, E. C. (1993) *Grammaticalization*. Cambridge University Press, Cambridge.

- ILARI, R.; FRANCHI, C.; NEVES, M. H. de M. & POSSENTI, S. (1993) “Os pronomes do Português Falado: Roteiro para a Análise”. In: CASTILHO, A. T. de. & BASÍLIO, M. (orgs.). *Gramática do Português Falado (vol. IV) – Estudos Descritivos*. Campinas, SP: Editora da Unicamp, São Paulo: FAPESP, pp. 79-166.
- JURAFSKY, D.; BELL, A.; GREGORY, M. & RAYMOND, W. D. (2001) “Probabilistic relations between words: evidence from reduction in lexical production”. In: BYBEE, J. & HOPPER, P. *Frequency and the Emergence of Linguistic Structure*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- KAISER, Georg A. (2006) “Sobre a (alegada) perda do sujeito nulo no português brasileiro”. In.: LOBO, T. et. al. (orgs.). *Para a história do Português Brasileiro – vol. VI: Novos dados, novas análises* (Tomo I). Salvador – BA: EDUFBA.
- KOCH, Ingedore V. & MARCUSHI, Luiz A. (1998) “Processos de referenciação na produção discursiva”. In.: *DELTA*, v. 14, n. special, São Paulo.
- LABOV, William. (1972) *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- LABOV, William. (1978) “Where does the Sociolinguistic Variable Stop? A response to Beatriz Lavandera”. In: *Working Papers in Sociolinguistics*. Austin: Southwest Educational Development Laboratory.
- LABOV, William. (1994). *Principles of Linguistic Change: Internal Factors*. Oxford UK & Cambridge, EUA: Blackwell.
- LABOV, William. (2001) *Principles of Linguistic Change: Social Factors*. Oxford UK & Cambridge, EUA: Blackwell.
- LABOV, William. (2006) *The Social Stratification of English in New York City*. Cambridge University Press. [Edição revisada do original de 1966].
- LABOV, William. (2008) “Quantitative Reasoning in Linguistics”. In: *Linguistics* 563. Disponível em: <http://www.ling.upenn.edu/~wlabov/Papers/QRL.pdf>
- LABOV, William. (2010) *Principles of Linguistic Change: cognitive and cultural factors*, v. 3, Wiley-Blackwell.

- LAVANDERA, Beatriz. (1978) “Where does the Sociolinguistic Variable Stop?” In: *Language in Society* 7, p. 171-182.
- LEAL, Eneida de Goes. (2004) “Estudo sobre haplologia e redução no dialeto da cidade paulista de Capivari”. In: *VII ENAPOL - Encontro dos Alunos de Pós-graduação em Linguística*. São Paulo.
- LOREGIAN-PENKAL, Loremi. (2004) *Reanálise da referência de segunda pessoa na fala da região sul*. (Tese de doutorado). Curitiba: Universidade Federal do Paraná.
- LUCCA, Nívia Naves G. (2005) *A variação tu/você na fala brasiliense*. Dissertação (Mestrado em Linguística). Brasília: Universidade de Brasília.
- MAINGUENEAU, Dominique. (2004) *Análises de textos de comunicação*. [Trad. de Souza-e-Silva, C. P. & Rocha, D.] 3ª ed. São Paulo: Cortex. (caps. 9 e 10).
- MARTINS, Germano Ferreira. (2010) *A alternância Tu/Você/Senhor no município de Tefé – Estado do Amazonas*. Dissertação (Mestrado). Brasília: Instituto de Letras, Universidade de Brasília.
- MENDES, Ronald Beline. (1999) *A Gramaticalização de estar + gerúndio no Português Falado*. Dissertação (Mestrado). Campinas/SP: Instituto de Estudos da Linguagem, UNICAMP.
- MENDES, Ronald Beline. (2008) “Contraction, Grammaticalization and the Phonology/Syntax interface: Progressives in Brazilian Portuguese”. In.: *Proceedings of the 2008 annual conference of the Canadian Linguistic Association*.
- MENON, Odete P. da S. (2006) “A história de você”. In.: GUEDES, M. et.al.(orgs). *Teoria e Análise Linguísticas: novas trilhas*. Araraquara, SP: UNESP, Cultura Acadêmica Editora.
- MEYERHOFF, Miriam. (2002) “Communities of Practice”. In.: Chambers, J. K., Trudgill, P. & Schilling-Estes, N. *The Handbook of Language, Variation and Change*. Malden / Oxford: Blackweel.

Referências Bibliográficas

- MICHELETTI, Helena. (2000) *A Indeterminação do sujeito – um estudo do pronome de terceira pessoa do plural*. Dissertação (Mestrado). São Paulo: Departamento de Letras Clássicas e Vernáculas – FFLCH/USP.
- MICHELETTI, Helena. & FRANCHETTI, S. (1996) “A indeterminação do sujeito: um estudo sociolinguístico”. In: *Anais do XXV Seminário do GEL*. Taubaté, São Paulo, p. 629-635.
- MILNER, Jean-Claude. (2003) “Reflexões sobre a referência e a correferência” [1982]. In.: *Referenciação – clássicos da linguística*, v. 1, São Paulo: Contexto. (Trad./Revisão e Org.: CAVALCANTE, M. M et. al.)
- MILROY, Leslie. (2002) “Redes Sociais”. In.: Chambers, J. K., Trudgill, P. & Schilling-Estes, N. *The Handbook of Language, Variation and Change*. Malden / Oxford: Blackweel.
- MIOTO, Carlos et. al. (2004) *Novo Manual de Sintaxe*. Florianópolis: Insular.
- MODESTO, Artarxerxes T. T. (2006) *Formas de tratamento no português brasileiro: a alternância tu/você na cidade de Santos-SP*. Dissertação (Mestrado). São Paulo: Departamento de Letras Clássicas e Vernáculas – FFLCH/USP.
- MONDADA, Lorenza & DUBOIS, Danièle. (2003) “Construção dos objetos de discurso e categorização: uma abordagem dos processos de referenciação”. In.: *Referenciação – clássicos da linguística*, v. 1, São Paulo: Contexto. (Trad./Revisão e Org.: CAVALCANTE, M. M et. al.)
- MOTA, Maria Alice. (2008) *A variação dos pronomes ‘tu’ e ‘você’ no português oral de São João da Ponte (MG)*. Dissertação (Mestrado em Linguística). Belo Horizonte, Minas Gerais: Universidade Federal de Minas Gerais.
- MOURA NEVES, Maria Helena de. (2000) *Gramática de usos do português*. São Paulo: UNESP.
- MOURA NEVES, Maria Helena de. (2001) “O Processo de Gramaticalização”. In.: *A Gramática Funcional*. São Paulo: Martins Fontes.

Referências Bibliográficas

- NARO, Anthony Julius. (2003a) “O dinamismo das línguas”. I: MOLLICA, M. C. & BRAGA, M. L. (orgs.) *Introdução à Sociolinguística: o tratamento da variação*. São Paulo: Contexto.
- NARO, Anthony Julius. (2003b) “Modelos quantitativos e tratamento estatístico”. I: MOLLICA, M. C. & BRAGA, M. L. (orgs.) *Introdução à Sociolinguística: o tratamento da variação*. São Paulo: Contexto.
- NASCIMENTO, Ivanete Belém do. (2006) “O uso variável do ‘você’ no português”. In: *54º Seminário do GEL*, Araraquara / SP. Caderno de Resumos.
- NASCIMENTO, Ivanete Belém do. (2008a) “A variação "você" vs. "cê" na cidade de São Paulo”. In: *56º Seminário do GEL*, São José do Rio Preto / SP. Caderno de Resumos.
- NASCIMENTO, I. B. do. (2008b) “O uso variável de VOCÊ na cidade de São Paulo”. Relatório Final de Iniciação Científica (Bolsista CNPq).
- NASCIMENTO, I. B. do. (2009) “*Cê num tá se cliticizando?* – Processos de cliticização e mudança linguística no PB.” In.: VI Congresso Internacional da ABRALIN, João Pessoa. Anais - VI Congresso Internacional da ABRALIN. João Pessoa : Ideia, v. 1.
- OUSHIRO, Livia. (2010) *Uma análise variacionista para as interrogativas-Q*. [Relatório de Qualificação, inédito].
- PAIVA, Maria da Conceição de. (2003) “A variável sexo/gênero”. In.: MOLLICA, M. C. & BRAGA, M. L. (orgs.) *Introdução à Sociolinguística: o tratamento da variação*. São Paulo: Contexto.
- PAIVA, Maria da Conceição de. & DUARTE, Maria Eugênia Lamoglia (2003) *Mudança Linguística em tempo real*. Rio de Janeiro: Contra Capa.
- PAIVA, M. da Conceição de. & SCHERRE, M. Marta Pereira. (1999) “Retrospectiva Sociolinguística: Contribuições do PEUL”. In: *D.E.L.T.A.*, v. 15, n. especial, p. 201-232.

Referências Bibliográficas

- PAREDES SILVA, Vera Lúcia (1998) “O retorno do pronome tu à fala carioca”. In.: RONCARATI, C. & ABRAÇADO, J. *Português Brasileiro: contato, heterogeneidade e história*. Rio de Janeiro: 7 Letras.
- PATRICK, Petter L. (2002) “Speech Community”. In.: Chambers, J. K., Trudgill, P. & Schilling-Estes, N. *The Handbook of Language, Variation and Change*. Malden / Oxford: Blackweel.
- PERES, Edenize Ponzo. (2006) *O uso de você, ocê e cê em Belo Horizonte: Um estudo em tempo aparente e em tempo real*. Tese de Doutorado. Minas Gerais: UFMG.
- PETERSEN, Carol. (2008) “A tripartição pronominal e o estatuto das proformas *cê*, *ocê* e *você*”. In.: DELTA, vol. 24, n.2, p. 283-308.
- QUEIROZ-ANDRADE, Carolina. (2010) *Tu e mais quantos? – A segunda pessoa na fala brasiliense*. Dissertação (Mestrado). Brasília: Instituto de Letras, Universidade de Brasília.
- RAMOS, Jânia. (1997) “O Uso das Formas Você, Ocê e Cê no Dialeto Mineiro”. In.: HORA, D. da. *Diversidade Linguística no Brasil*. João Pessoa: Ideia, p. 43-60.
- RAMOS, Jânia. (1998) “Historia social do português brasileiro: perspectivas”. In.: CASTILHO, A. T. de. (org.) *Para a História do Português Brasileiro*, v. 1, São Paulo: Humanitas, p. 153-167.
- RAMOS, Jânia. (2000) “O surgimento de um novo clítico no português Brasileiro: análise quantitativa e qualitativa da forma *cê*”. In.: GÄRTNER, E. et al. (eds.) *Estudos de Sociolinguística Brasileira e portuguesa*, Frankfurt am MAIn: TFM, v. 15, p. 181-189.
- RAMOS, Jânia. (2006) “Gramaticalização em processo”. In.: VITRAL, L. & RAMOS, J. (orgs.) *Gramaticalização: uma abordagem Formal*. Rio de Janeiro: Tempo Brasileiro; Belo Horizonte, MG: Faculdade de Letras FALE / UFMG.
- RAMOS, Jânia. (2006) “O Processo *não > num* na Fala”. In.: VITRAL, L. & RAMOS, J. (orgs.) *Gramaticalização: uma abordagem Formal*. Rio de Janeiro: Tempo Brasileiro; Belo Horizonte, MG: Faculdade de Letras FALE / UFMG.

Referências Bibliográficas

- RAMOS, J. & CORRÊA, L. (2006) “Mais um pronome em processo de cliticização: o par eles / es”. In.: VITRAL, L. & RAMOS, J. (orgs.) *Gramaticalização: uma abordagem Formal*. Rio de Janeiro: Tempo Brasileiro; Belo Horizonte, MG: Faculdade de Letras FALE / UFMG.
- RAMOS, Jânia. & OLIVEIRA, Marilza. (2002) “O estatuto de *você* no preenchimento do sujeito”. In.: *XIII Congresso Internacional da ALFAL*, San José, Universidade de Costa Rica. Disponível em: http://www.lettras.ufmg.br/nucleos/nupevar/dados/arquivos/Oliveira%20e%20Ramos%20_2002_.pdf
- ROCHA LIMA, Carlos H. da (1985) *Gramática Normativa da Língua Portuguesa*. Rio de Janeiro: José Olímpio, 26ª ed.
- SACCONI, Luiz Antônio. (1987) *Gramática essencial da língua portuguesa: teoria e prática*. São Paulo: Atual, 2ª ed.
- SANKOFF, D. TAGLIAMONTE, S. A. & SMITH, E. (2005) *Goldvar X – A multivariate analysis application*. Toronto: Department of Linguistics; Ottawa: Department of Mathematics. <http://individual.utoronto.ca/tagliamonte/goldvarb.htm>
- SAUSSURE, Ferdinand de. (1969) *Curso de Linguística Geral*. São Paulo, Cultrix. [Edição brasileira; Trad.: Chelini, A.; Paes, J. P. & Blikstein, I.]
- SCHERRE, Maria Marta Pereira. (1998) “Paralelismo Linguístico”. In.: *Revista de Estudos da Linguagem* (vol. 7 – n. 2). Faculdade de Letras da UFMG: Belo Horizonte – MG, jul./dez.
- SCHERRE, M. M. P. (2007) “Aspectos sincrônicos e diacrônicos do imperativo gramatical no português brasileiro”. *Alfa*, São Paulo, 51 (1): 189-222.
- SCHERRE, M. M. P. (2010-2013) *Análise e mapeamento de três fenômenos variáveis no português brasileiro*. Projeto de pesquisa apresentado ao CNPq para o triênio 2010-2013, em parceria com o Prof. Dr. Anthony Julius Naro (inédito).
- SCHERRE, Maria Marta Pereira & NARO, Anthony Julius (2003) “Análise quantitativa e tópicos de interpretação do Varbrul”. I: MOLLICA, M. C. & BRAGA, M. L. (orgs.) *Introdução à Sociolinguística: o tratamento da variação*. São Paulo: Contexto.

Referências Bibliográficas

- SETTE, Neide Durães. (1980) *Formas de tratamento no Português coloquial*. Dissertação (Mestrado). Recife: Universidade Federal de Pernambuco.
- SILVA-CORVALÁN, Carmen. (1989) *Sociolingüística: Teoría y análisis*. Madrid: Alhambra.
- SOUSA, Valéria Viana. (2008) *Os (des)caminhos do VOCÊ: uma análise sobre a variação e mudança na forma, na função e na referência do pronome você*. Tese de doutorado. João Pessoa, Paraíba : Universidade Federal da Paraíba.
- SOUZA, Elizete Maria de. (2007) *O Uso do pronome 'eles' como recurso de indeterminação do sujeito*. Dissertação (Mestrado). Belo Horizonte: Faculdade de Letras da Universidade Federal de Minas Gerais (UFMG).
- SUZUKI, Keiichiro. (1998) *A Typological Investigation of Dissimilation*. Dissertation. Department of Linguistics, University of Arizona.
- TEIXEIRA, Eliana Sandra Pitombo. (2002) *Era uma vez você*. Tese de doutorado Salvador/Bahia: Instituto de Letras, Universidade Federal da Bahia.
- VIOTTI, Evani de Carvalho. (2003) “A Composicionalidade nas sentenças com o verbo *ter*”. In.: MÜLLER, A. L.; NEGRÃO, E. V. & FOLTRAN, M. J. (orgs.) *Semântica Formal*. São Paulo: Contexto.
- VITRAL, Lorenzo. (1996) “A forma *CÊ* e a noção de gramaticalização”. In.: *Revista de Estudos da Linguagem*. (ano 5, nº 4, v. 1). Minas Gerais: Faculdade de Letras da UFMG.
- VITRAL, Lorenzo. (2002) “Identificando Clíticos: evidências fonéticas”. In.: *Anais do 50º Seminário do GEL*. (Caderno de Resumos).
- VITRAL, Lorenzo. (2006a) “Identificando Clíticos: evidências fonéticas”. In.: VITRAL, L. & RAMOS, J. (orgs.) *Gramaticalização: uma abordagem Formal*. Rio de Janeiro: Tempo Brasileiro; Belo Horizonte, MG: Faculdade de Letras FALE / UFMG.

- VITRAL, Lorenzo. (2006b) “A Interpolação de SE e suas consequências para a teoria da Cliticização”. In.: VITRAL, L. & RAMOS, J. (orgs.) *Gramaticalização: uma abordagem Formal*. Rio de Janeiro: Tempo Brasileiro; Belo Horizonte, MG: Faculdade de Letras FALE / UFMG.
- VITRAL, Lorenzo & RAMOS, Jânia (2006) “Gramaticalização de “você”: um processo de perda de informação semântica?”. In.: VITRAL, L. & RAMOS, J. *Gramaticalização – uma abordagem formal*. Rio de Janeiro: Tempo Brasileiro; Belo Horizonte, MG: Faculdade de Letras FALE / UFMG. [1999]
- VOTRE, Sebastião Josué. (2003) “Relevância da variável escolaridade”. In: MOLLICA, M. C. & BRAGA, M. L. (orgs.) *Introdução à Sociolinguística: o tratamento da variação*. São Paulo: Contexto.
- WEINREICH, Uriel, LABOV, William & HERZOG, Marvin I. (1968) Empirical Foundations for a Theory of Linguistic Change. In: LEHMANN, W.P. & Y. MALKIEL (eds.) *Directions for Historical Linguistics: A Symposium*. Austin: University of Texas Press, p. 97-195. Disponível em: <http://www.utexas.edu/cola/centers/lrc/books/hist00.html>
- WEINREICH, U., LABOV, W.; & HERZOG, M. I. (2006) *Fundamentos empíricos para uma teoria da mudança linguística*. São Paulo: Parábola. (Tradução de Marcos Bagno).
- ZILLES, A. M. S. (2002) “Grammaticalization of 'a gente' in Brazilian Portuguese”. In.: *Working Papers in Linguistics*, Philadelphia: University of Pennsylvania, v. 8, n. 3.
- ZILLES, Ana M. S. (2004) “Grammaticalization of *a gente* as a cluster of changes: evidence from apparent and real time studies”. In.: Fórum Linguístico, Florianópolis, v. 4, nº 1, julho de 2004, pp. 13-46.
- ZILLES, Ana M. S. (2005) The development of a new pronoun: the linguistic and social embedding of *a gente* in Brazilian Portuguese. *Language Variation and Change*, v. 17, n. 1, p. 19-53.

ANEXO 1

- Amostra GESOL-SP-2000 (36 Informantes):

Número total de verbos nas estruturas “pronomes + V”: 2722.

Itens verbais até 1%
(Nº de ocorrências: 1 a 27; Total = 921 ocorrências; 217 itens) ¹
abrir (4); acabar (16); acertar (1); achar (23); acordar (1); acostumar (2); acreditar (4); adaptar(-se) (1); adequar(-se) (1); adquirir (1); afunilar (1); agilizar (1); agüentar (3); amar (1); amolar (2); analisar (1); andar (7); apaixonar (-se) (1); apoiar (1); apontar (1); apresentar (1); aprender (18); aprofundar (2); arrumar (4); assinar (1); assistir (8); assumir (1); atender (6); avançar (1); bater (1); brigar (1); brilhar (1); brincar (2); cair (2); catar (1); chegar (21); chorar (1); citar (2); classificar (1); colocar (8); começar (18); comer (1); cometer (2); complicar (1); comprar (16); conduzir (1); conhecer (22); consumir (4); construir (1); continuar (3); contribuir (3); conversar (13); correr (2); cortar (1); costuma (1); crescer (1); criar (1); criticar (2); cuidar (2); cumprir (1); dar (20); dançar (1); decidir (1); dedicar (1); definir (2); deixar (8); demonstrar (1); depender (4); descer (3); descobrir (2); desenvolver (1); desligar (1); deslocar (1); desvincular (1); detectar (1); deter (1); dever (12); dialogar (1); diferenciar (1); diminuir (1); dizer (14); dormiu (1); emocionar (1); emprenhar (1); encher (2); encontrar (17); encostou (1); enfrentar (3); enganar (1); enlouquecer (1); ensinar (2); entrar (24); entregar (1); entrevistar (1); envelhecer (1); era/ser (9); errar (1); escolher (3); escrever (2); escutar (4); esperar (2); esquecer (4); (es)tar (+ AP) (8); estimular (1); estourar (1); estragar (1); estudar (11); experimentar (1); explicitar (1); extravasar (1); faltar (1); fechar (2); for/ser (27); formar (2); forrar (1); frequentar (2); fugir (1); ganhar (8); garantir (1); gastar (7); gerar (1); gosta (7); há de convir (1); ia (3); ir + inf. (5); identificar(-se) (2); ignorar (1); imaginar (9); importar(-se) (1); ingerir (1); instruir (1); interessar(-se) (1); interromper (1); investir (1); ir(ia) (+ inf.) (4); jogar (3); julgar (2); ler (3); lembrar (11); levar (6); ligar (8); lutar (1); mandar (2); manter (2); matar (4); memorizar (1); merecer (1); misturar (1); montar (1); morar (17); morrer (2); movimentar (1); mudar (2); nascer (4); notar (4); olhar (14); orientar (1); ouvir (9); pagar (19); parar (6); parecer (2); passar (20); pedir (5); pensar (16); perceber (18); perder (7); perguntar (10); pôr (7); preferir (2); prejudicar (1); preparar (1); presenciar (1); prestar (+ SN) (3); prestar atenção (7); privar (se) (1); procurar (2); produzir (2); pronunciar (1); puxar (1); realizar (se) (1); receber (4); reciclar (1); reconhecer (1); reeleger (1); renovar (1); reparar (1); repetir (1); recusar (1); resolver (2); respeitar (1); sair (20); salvar (1); seguir (1); sentar (2); sentir (8); ser (+x) (8); simplificar (1); sintonizar (1); subir (1); suar (1); tentar (10); terminar (5); tirar (7); tocar (2); tomar (8); trabalhar (15); travar (1); trazer (3); tremer (1); usar (6); vir (19); verifica (1); viajar (1); virar (1); viver(9); voltar (6); votar (7).

Itens verbais de 1% a 2%
(Nº de ocorrências: 29 a 45; Total = 196 ocorrências; 5 itens)
conseguir (39); entender (45); ficar (41); pegar (42); precisar (29)
Itens verbais de 2% a 3%
(Nº de ocorrências: 63 a 75; Total = 277 ocorrências; 4 itens)
é/ser (69); fazer (75); querer (70); (es)tar (+ loc./PP) (63)

¹ O número entre parênteses corresponde ao número de ocorrências daquele item verbal.

Itens verbais de 3% a 4% (Nº de ocorrências: 85 a 97; Total = 274 ocorrências; 3 itens)
poder (92); saber (85); (es)tar + gerúndio (97)
Itens verbais de 4% a 6% (Nº de ocorrências: 115 a 152; 406 ocorrências; 3 itens)
falar (139); ter que (152); vai (+ loc.) (115)
Itens verbais acima de 6% (Nº de ocorrências: 190 a 203; Total = 592 ocorrências; 3 itens)
ter (+ X) (203); ir + inf. (199); ver (190)

- Amostra GESOL-SP-2000 (15 Documentadores paulistas/paulistanos):

Número total de verbos nas estruturas “pronomes + V”: 1840.

Itens verbais até 1% (Nº de ocorrências: 1 a 17; Total = 539 ocorrências; 161 itens) ²
Abrir (1); acabar (1); aceitar (1); acompanhar (3); acordar (1); acreditar (1); adquirir (3); ajudar (2); almoçar (2); andar (1); aposentar (1); apreciar (1); aprender (18); arrepender (1); assistir (8); associar (1); bater (2); beber (3); botar (1); brincar (1); buscar (1); cair (1); caracterizar (2); casar (5); checar (1); chegar (6); colocar (5); comer (2); começar (4); comentar (1); comparar (1); concluir (1); concordar (5); confiar (1); conseguir (6); considerar (1); contar (1); continuar (1); contar (2); conversar (15); conviver (2); corrigir (2); (a)costumar (10); crescer (2); cuidar (1); curtir (4); dar (5); debochar (1); decidir (1); dedicar (2); deixar (4); depender (1); descansou (1); descontar (1); desejar (1); desenvolver (1); desistir (1); detestar (3); dever (5); dirigir (1); dizer (18); divertir (2); dormir (1); encontrar (1); entender (1); entrar (2); enxerga (2); era (10); escolher (4); escutar (4); esperar (1); esquecer (1); esquentar (1); estar (12); ficar (16); flexionar (1); forçar (3); fumar (1); formar (2); fosse (3); frequentar (1); fugir (1); ganhar (1); gastar (2); gritar (1); ia + inf (8); identificar (1); imaginar (2); interessar (1); investir (1); ir/iria (6); jantar (1); jogar (1); largar (1); levar (4); levantar (1); manter (1); matar (1); mudar (1); nascer (7); notar (1); olhar (1); optar (2); orientar (1); ouvir (11); pagar (1); parar (6); participar (1); passar (8); pegar (7); perceber (15); perder (1); pesquisar (1); pintar (1); planejar (1); poder (17); pôr (1); poupar (1); praticar (5); precisar (5); preferir (15); preocupar (1); presenciar (1); prestar (2); pretender (2); proteger (1); pudessem (3); pular (1); puxar (1); receber (3); recomendar (1); reconhecer (2); reparar (6); repetir (1); reprovar (1); resolver (1); responder (3); rir (1); roubar (1); sair (7); sentar (1); sentir (16); ser (6); sofrer (3); subir (2); terminar (2); tocar (1); tomar (2); torcer (1); trabalhar (13); trocar (1); usar (14); utilizava (2); vai/ir (14); vir (7); viajar (4); virar (1); visitar (2); viver (8); voltar (4); votar (3).

² O número entre parênteses corresponde ao número de ocorrências daquele item verbal.

Itens verbais de 1% a 2% (Nº de ocorrências: 19 a 35; Total = 252 ocorrências; 10 itens)
conhecer (21); é/ser (29); estudar (35); lembrar(-se) (28); pensar (19); prestar atenção (19); querer (20); (es)tar + loc./PP (26); vai + inf. (28); ver (27)
Itens verbais de 2% a 3% (Nº de ocorrências: 54; Total = 54 ocorrências; 1 item)
morar (54)
Itens verbais de 3% a 4% (Nº de ocorrências: 57 a 67; Total = 187 ocorrências; 3 itens)
fazer (67); saber (57); (es)tar + ger. (63)
Itens verbais de 4% a 6% (Nº de ocorrências: 108; 108 ocorrências; 1 item)
gostar (108)
Itens verbais acima de 6% (Nº de ocorrências: 117 a 432; Total = 671 ocorrências; 3 itens)
achar (432); falar (122); ter (117)

- Amostra NURC-SP-1970 (18 Informantes):

Número total de verbos nas estruturas “pronomes + V”: 362.

Itens verbais até 1% (Nº de ocorrências: 1 a 3; Total = 105 ocorrências; 78 itens) ³
agradar (1); abandonar (1); abordar (1); achar (2); adaptar (1); agredir (1); ajudar (2); analisar (1); aprender (2); aprontar (2); arrumar (2); assentar (1); assinar (2); colocar (1); combinar (1); comprar (2); conhecer (3); conseguir (3); conservar (1); contar (1); controlar (2); cruzar (1); deixar (1); depositar (3); dever (1); diversificar (1); embarcar (1); enfrentar (1); entrar (2); escolher (1); estar (3); estudar (1); existe (2); for (2); formulasse (1); frequentar (1); fumar (1); ir (3); imaginar (2); interpretar (1); lembrar (1); levar (1); ligar (1); mandar (1); notar (1); olhar (1); pagar (2); passar (2); pedir (1); pendurar (1); pensar (3); perceber (1); perder (1); precisar (3); prejudicar (1); procurar (1); receber (1); referir (1); reparar (1); retirar (2); revisar (1); sair (2); segurar (2); sentar (1); sofrer (1); trabalhar (2); usar (1); ir + ger. (1); vestir (1); vir (2); voltar (2)

³ O número entre parênteses corresponde ao número de ocorrências daquele item verbal.

<p>Itens verbais de 1% a 2%</p> <p>(Nº de ocorrências: 4 a 7; Total = 81 ocorrências; 16 itens)</p> <p>apresentar (4); assistir (5); chegar (4); começar (4); dar (4); dizer (6); encontrar (4); entender (7); estar + ger. (5); falar (7); ficar (5); gostar (6); pegar (5); perguntar (6); (es)tar (4); ir (5)</p>
<p>Itens verbais de 2% a 3%</p> <p>(Nº de ocorrências: 8 a 9; Total = 34 ocorrências; 4 itens)</p> <p>ser/é (8); fazer (9); sentir (8); (es)tar + ger. (9)</p>
<p>Itens verbais de 3% a 4%</p> <p>(Nº de ocorrências: 13; Total = 13 ocorrências; 1 item)</p> <p>saber (13)</p>
<p>Itens verbais de 4% a 6%</p> <p>(Nº de ocorrências: 15 a 18; 33 ocorrências; 2 itens)</p> <p>poder (15); querer (18)</p>
<p>Itens verbais acima de 6%</p> <p>(Nº de ocorrências: 23 a 39; Total = 96 ocorrências; 3 itens)</p> <p>ter (39); ir + inf. (23); ver (34)</p>

ANEXO 2

Tabela de valores do *qui-quadrado*

Área à direita do valor crítico

Graus de liberdade	0,95	0,90	0,10	0,05	0,01
1	0,004	0,016	2,706	3,841	6,635
2	0,103	0,211	4,605	5,991	9,210
3	0,352	0,584	6,251	7,815	11,345
4	0,711	1,064	7,779	9,488	13,277
5	1,145	1,610	9,236	11,071	15,086
6	1,635	2,204	10,645	12,592	16,812
7	2,167	2,833	12,017	14,067	18,475
8	2,733	3,490	13,362	15,507	20,090
9	3,325	4,168	14,684	16,919	21,666
10	3,940	4,865	15,987	18,307	23,209

(Guy & Zilles, 2007: 234)