

Lúcia de Almeida Ferrari

**A EROSÃO LINGUÍSTICA DE ITALIANOS CULTOS EM CONTATO COM O
PORTUGUÊS BRASILEIRO: ASPECTOS DO SISTEMA PRONOMINAL**

Dissertação apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Minas Gerais como requisito parcial para obtenção do grau de Mestre em Linguística.

Área de Concentração: Linguística Teórica e Descritiva

Orientador: Prof. Dr. Tommaso Raso

Belo Horizonte

Faculdade de Letras da UFMG

2010

Ficha catalográfica elaborada pelos Bibliotecários da Biblioteca FALE/UFMG

F375e Ferrari, Lúcia de Almeida.
A erosão lingüística de italianos cultos em contato com o português brasileiro [manuscrito] : aspectos do sistema pronominal / Lúcia de Almeida Ferrari. – 2010.
147 f., enc. : il., p&b, color., grafs., tabs. + 1 CD-ROM.

Orientador: Tommaso Raso.

Área de concentração: Lingüística Teórica e Descritiva.

Linha de Pesquisa: Estudos Lingüísticos Baseados em Corpora.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Faculdade de Letras.

Inclui CD-ROM com as transcrições do *corpus* compilado para o trabalho em formato .txt.

Bibliografia: f. 126-143.

Anexos: f. 144-147.

1. Língua italiana - Pronomes – Teses. 2. Língua italiana - Variação – Brasil – Teses. 3. Corpora (Lingüística) – Teses. 4. Aquisição da segunda linguagem – Teses. 5. Italianos – Linguagem – Brasil – Teses. 6. Língua portuguesa – Influências sobre o italiano – Teses. 7. Bilingüismo – Teses. 8. Sociolingüística – Teses. I. Raso, Tommaso. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III. Título.

CDD: 457.9981

A Lucílio, pela falta que sua alegria me faz.

Agradecimentos

Desejo expressar minha gratidão a todos aqueles que direta ou indiretamente tornaram possível a realização de mais esta etapa da minha vida.

Agradeço aos meus pais que sempre apoiaram meus esforços e meus estudos.

Um agradecimento especial a Marco, pela paciência constante.

Aos amigos e colegas que estiveram perto nas horas de desânimo e me deram forças para continuar.

Aos professores que me incentivaram.

Aos meus informantes pela ajuda valiosa e gentileza quando da gravação das falas.

Ao meu orientador, por ter pacientemente me ensinado tudo o pouco que sei sobre linguística.

RESUMO

Este trabalho explora a erosão linguística de italianos cultos em contato prolongado com o português do Brasil. A erosão linguística segue a definição de Kopke e Schmid (2004) e nesta pesquisa busca-se verificar a Hipótese da Regressão de Jakobson (1941). Optou-se por utilizar informantes cultos para evitar que uma aquisição incompleta da L1, o italiano, fosse confundida com a própria erosão. A opção pelo contato prolongado, ou seja uma estadia no Brasil por parte dos informantes de pelo menos oito a dez anos, garantiria uma aquisição quase nativa da L2, o português brasileiro. A investigação é baseada em um *corpus* de 21298 palavras de bilíngues, italianos que tenham crescido e estudado na Itália até a idade adulta e possuam título superior, *corpus* por nós coletado e transcrito nos anos de 2008 e 2009. Este *corpus* foi comparado com uma amostra de 21224 palavras extraída do *C-ORAL-ROM italiano* (Cresti-Moneglia, 2005). Ambos os *corpora* possuem, além das transcrições (em formato CHAT), as respectivas gravações das interações. Objetivo deste trabalho era investigar e verificar os resultados dos trabalhos de Raso e Vale (2009), do qual este estudo quer ser a continuação. As pesquisas guiadas por Raso, que definiu os critérios que foram aqui seguidos, foram feitas sobre um diferente *corpus* de italianos sujeitos a erosão, e comparadas com um outro *corpus* de referência. Naquele estudo, como neste, foram investigados aspectos do sistema pronominal dos clíticos, a dizer: o pronome *ci* em seus valores atualizante, lexicalizante e locativo, os pronomes *ne* em suas funções partitiva, argumental e locativa e os pronomes acusativos de terceira pessoa, ou seja, *lo, la li, le, l'*. A metodologia adotada foi aquela da análise de *corpora*, apesar de não ter sido possível utilizar ferramentas computacionais pela dificuldade de encontrar um *software* adequado. Procedeu-se a uma varredura manual dos *corpora* comparados em busca dos clíticos objeto de estudo e de seus contextos significativos. Em seguida foram comparados os resultados obtidos com aqueles de Raso e Vale (2009). Os resultados confirmaram a presença de erosão linguística também no *corpus* por nós coletado, com algumas diferenças substanciais devidas a uma diferente arquitetura deste último em relação ao anterior, e à escolha por compará-lo com um *corpus* diferente, o que proporcionou a eliciação de um tipo de fala mais interativo e delimitado situacionalmente, e portanto variações no nível de erosão dos vários pronomes entre um estudo e outro. Em nosso *corpus*, a erosão não se mostra tão marcante como aquela detectada no *corpus* anterior. Supõe-se que o fenômeno não se estabilize após a primeira década, como afirmam estudos anteriores, mas continue com o passar do tempo. Abre-se portanto a possibilidade de se dar continuidade aos trabalhos iniciados para confirmar as hipóteses aqui apresentadas e verificar de forma mais segura a Hipótese da Regressão.

Palavras-chave: erosão; *corpus*; pronomes; clíticos; italiano

RIASSUNTO

Questa ricerca indaga l'erosione linguistica di italiani colti in contatto prolungato con il portoghese brasiliano. La definizione di erosione linguistica segue quella indicata da Kopke e Schmid (2004) ed in questo studio si cercherà di verificare l'ipotesi della Regressione di Jakobson (1941). Si è scelto di utilizzare informatori colti per evitare che un'acquisizione incompleta della L1, l'italiano, potesse essere confusa come un aspetto dell'erosione stessa. La preferenza per il contatto prolungato, ossia una permanenza in Brasile da parte degli informatori di perlomeno otto a dieci anni servirebbe a garantire una loro acquisizione quasi nativa della L2, il portoghese brasiliano. La ricerca si è basata su un *corpus* di 21298 parole di bilingui, italiani che siano cresciuti e abbiano concluso il loro ciclo di studi in Italia fino alla maggiore età e siano in possesso di laurea, che abbiamo raccolto e trascritto negli anni 2008 e 2009. Questo *corpus* è stato confrontato con stralci di 21224 parole tratte dal C-ORAL ROM italiano (Cresti-Moneglia, 2005). Entrambi i *corpora* sono corredati, oltre che dalle trascrizioni (in formato CHAT), anche dalle corrispettive registrazioni delle interazioni. L'obiettivo di questa ricerca era verificare ed approfondire i risultati delle indagini di Raso e Vale (2009), di cui questo studio vuole essere la continuazione. Le ricerche dirette da Raso (2009), che ha definito i criteri che sono stati qui seguiti, sono state condotte su un diverso *corpus* di italiani sottoposti a erosione ed i risultati sono stati confrontati con un altro *corpus* di riferimento. In quella ricerca, così come in questa, sono stati esaminati aspetti del sistema pronominale dei clitici, in particolare: il pronome ci nei valori attualizzante, lessicalizzante e locativo, i pronomi ne nelle funzioni partitiva, argomentale e locativa, ed i pronomi accusativi di terza persona, ossia: lo, la, li, le, l'. La metodologia utilizzata è stata quella dell'analisi di *corpora*, nonostante non sia stato possibile utilizzare alcun tipo di ferramenta informatica a causa della difficoltà nell'incontrare un *software* adeguato alle nostre esigenze. Si è proceduto ad una analisi manuale dei *corpora* confrontati alla ricerca dei clitici oggetto di studio e dei loro corrispettivi contesti significativi. In seguito sono stati confrontati i risultati da noi ottenuti con quelli di Raso e Vale (2009). I risultati hanno confermato la presenza di erosione linguistica anche nel *corpus* da noi raccolto, pur con alcune differenze significative dovute ad una diversa architettura di quest'ultimo in rapporto a quello precedente e alla scelta di confrontarlo con un *corpus* ancora diverso, il che ha favorito la manifestazione di un tipo di parlato più interattivo e delimitato situazionalmente, e perciò variazioni nel grado di erosione dei vari pronomi tra uno studio e l'altro. Nel nostro *corpus* tuttavia l'erosione non è tanto espressiva quanto attestato dagli studi precedenti. Appare di conseguenza plausibile, e meritoria di ulteriori ricerche, l'ipotesi che l'erosione linguistica non si stabilizzi dopo il primo decennio, come affermano vari studi, ma continui con il passare degli anni. Si apre dunque la possibilità di poter dare continuità agli studi intrapresi per confermare le ipotesi qui presentate e verificare in maniera più sicura l'ipotesi della Regressione.

Parole chiave: erosione, *corpus*, pronomi, clitici; italiano

LISTA DE TABELAS

Tabela 1-	Ocorrência dos clíticos nos <i>corpora</i> de italianos bilíngues e monolíngues.....	84
Tabela 2-	Ocorrência dos clíticos ne em suas várias funções nos <i>corpora</i> de italianos bilíngues e monolíngues.....	85
Tabela 3-	Ocorrência do clítico <i>ci</i> com valor lexicalizante nos <i>corpora</i> de italianos bilíngues e monolíngues.....	86
Tabela 4-	Ocorrências absolutas dos <i>types</i> lexicais e <i>tokens</i> do clítico <i>ci</i> com verbo.....	87
Tabela 5-	Ocorrência dos clíticos acusativos de terceira pessoa nos <i>corpora</i> de italianos bilíngues e monolíngues.....	89
Tabela 6-	Ocorrência dos clíticos acusativos de terceira pessoa nos <i>corpora</i> de italianos bilíngues e monolíngues total e em diferentes posições.....	89
Tabela 7-	Ocorrência (normalizada para 11.000 palavras) dos diferentes tipos de tematização nos <i>corpora</i> de monolíngues italianos, monolíngues brasileiros e italianos bilíngues.....	93
Tabela 8-	Ocorrência dos clíticos nos <i>corpora</i> de italianos bilíngues e monolíngues.....	97
Tabela 9-	Ocorrência do clítico <i>ne</i> em suas várias funções nos <i>corpora</i> de italianos bilíngues e monolíngues.....	100
Tabela 10-	Ocorrência do clítico atualizante <i>ci</i> +verbo (<i>esserci</i>) em suas funções existencial e apresentacional nos <i>corpora</i> de italianos bilíngues e monolíngues.....	103
Tabela 11-	Ocorrência do clítico <i>ci</i> atualizante com o verbo <i>avere</i> nos <i>corpora</i> de italianos bilíngues e monolíngues.....	104
Tabela 12-	Ocorrência de <i>esserci</i> existencial e apresentacional nos <i>corpora</i> de italianos bilíngues e monolíngues.....	105
Tabela 13-	Ocorrência normalizada da forma <i>esserci</i> nas funções existencial e apresentacional nos diferentes <i>corpora</i> de italianos bilíngues e monolíngues.....	106
Tabela 14-	Ocorrência normalizada do clítico <i>ci</i> atualizante com o verbo <i>avere</i> nos diferentes <i>corpora</i> de italianos bilíngues e monolíngues.....	106
Tabela 15-	Comparação entre a ocorrência normalizada de <i>esserci</i> existencial e apresentacional do <i>corpus</i> Raso-Vale em relação ao <i>C-ORAL-ROM Italiano</i> e ao <i>BADIP</i>	109
Tabela 16-	Comparação entre a ocorrência normalizada de <i>esserci</i> existencial e apresentacional do <i>corpus</i> Raso-Ferrari em relação ao <i>BADIP</i> e ao <i>C-ORAL-ROM italiano</i>	109
Tabela 17-	Comparação entre a ocorrência normalizada de <i>averci</i> dos <i>corpus</i> Raso-Vale em relação ao <i>C-ORAL-ROM italiano</i> e ao <i>corpus BADIP</i>	111
Tabela 18-	Comparação entre a ocorrência normalizada de <i>averci</i> do <i>corpus</i> Raso-Ferrari em relação ao <i>BADIP</i> e ao <i>C-ORAL-ROM Italiano</i>	111
Tabela 19-	<i>Types</i> lexicalizantes do clítico <i>ci</i> e ocorrência de <i>tokens</i> nos <i>corpora</i> Raso-Ferrari e <i>C-ORAL-ROM Italiano</i>	112-113
Tabela 20-	Ocorrência dos clíticos nos <i>corpora</i> de italianos bilíngues e monolíngues.....	114
Tabela 21-	Comparação entre as ocorrências normalizadas do <i>corpus</i> Raso-Vale em relação ao <i>C-ORAL-ROM Italiano</i> e do <i>corpus</i> Raso-Ferrari em relação ao <i>BADIP</i>	116

LISTA DE FIGURAS

Quadro 1-	Os pronomes pessoais clíticos divididos por funções	26
Quadro 2-	Informações sobre o <i>corpus</i> e cada um dos textos.....	69-70
Quadro 3-	Dados sobre os informantes investigados na pesquisa.....	73
Gráfico 1-	Variações percentuais dos clíticos sujeitos a erosão em comparação com os dois <i>corpora</i> de referência nos dois estudos apresentados.....	98
Gráfico 2-	Ocorrências normalizadas do clítico <i>ne</i> em suas várias funções nos dois <i>corpora</i> de informantes bilíngues.....	100
Gráfico 3-	Variações percentuais entre os dois <i>corpora</i> de bilíngues e seus respectivos <i>corpora</i> de comparação de monolíngues.....	101
Gráfico 4-	Variação percentual entre o <i>corpus</i> Raso-Ferrari e o <i>C-ORAL-ROM italiano</i>	104
Gráfico 5-	Dados normalizados dos dois <i>corpora</i> de comparação.....	107
Gráfico 6-	Variação percentual entre os dados normalizados do <i>BADIP</i> e do <i>C-ORAL-ROM italiano</i>	108
Gráfico 7-	Comparação cruzada entre os <i>corpora</i> de monolíngues e bilíngues e as respectivas comparações originais.....	110
Gráfico 8-	Variação percentual entre o <i>corpus</i> Raso-Ferrari e o <i>C-ORAL-ROM italiano</i>	115
Gráfico 9-	Ocorrências normalizadas dos acusativos de terceira pessoa nos dois <i>corpora</i> de bilíngues.....	116
Gráfico 10-	Variações percentuais entre os <i>corpora</i> comparados nos dois estudos.....	117
Gráfico 11-	Ocorrências normalizadas dos acusativos de terceira pessoa nos dois <i>corpora</i> de monolíngues.....	118
Gráfico 12-	Comparação cruzada entre os <i>corpora</i> de bilíngues e monolíngues e as respectivas comparações cruzadas.....	119

SUMÁRIO

1 INTRODUÇÃO.....	11
1.2 Justificativa	12
1.3 Organização da Dissertação.....	12
2 REVISÃO DA LITERATURA.....	14
2.1 A erosão linguística.....	14
2.1.2 A erosão linguística da L1.....	17
2.2 Bilinguismo: contato e aquisição.....	19
2.2.1 Ordem e sequência aquisicional de L2.....	22
2.3 Sequência de aquisição dos clíticos no italiano L2.....	25
2.3.1 Os clíticos em italiano.....	26
2.3.1.1 Os acusativos de terceira pessoa.....	29
2.3.1.2 O ci locativo.....	30
2.3.1.3 O ne e suas funções.....	31
2.3.1.4 O ci atualizante e lexicalizante.....	32
2.3.2 A história linguística italiana e o sistema pronominal.....	36
2.4 Aspectos psicolinguísticos e sociais da erosão linguística.....	39
2.4.1 Fatores de indução interna.....	39
2.4.2 Fatores de indução externa.....	41
2.4.3 Aspectos psicolinguísticos.....	45
2.4.4 Sequência de erosão linguística de L1.....	47
2.5 A Hipótese da Regressão de Jakobson.....	48
2.5.1 O teste de Keijzer sobre a Hipótese da Regressão.....	50
2.6 O debate metodológico nas pesquisas sobre erosão linguística.....	53
2.7 Estudos sobre a erosão linguística do italiano.....	55
2.8 Estudos sobre a erosão linguística do Português do Brasil.....	59
2.9 Os estudos coordenados por Raso.....	60
3 METODOLOGIA.....	62
3.1 A importância dos estudos de corpora.....	62
3.2 O corpus analisado por Raso e Vale.....	65
3.2.1 Características do corpus de São Paulo.....	65
3.3 O novo corpus Raso-Ferrari.....	67
3.3.1 Coleta de dados e transcrições.....	68
3.3.2 Perfil sociolinguístico dos participantes do novo corpus.....	71
3.3.3 Método de análise dos dados: erros e acertos.....	73
3.3.4 Limitações do corpus Raso-Ferrari.....	74
3.4 Os corpora de comparação.....	76
3.4.1 O BADIP.....	77
3.4.1.1 Textos utilizados na pesquisa de Raso-Vale.....	78
3.4.2 O NURC.....	79
3.4.3 O C-ORAL-ROM italiano.....	80
3.4.3.1 Textos utilizados na presente pesquisa.....	81
4 ANÁLISE DOS RESULTADOS.....	83
4.1 Os resultados obtidos nos estudos de Raso e Vale (2009).....	83
4.1.1 O pronome ci com valor locativo.....	85
4.1.2 O pronome ne em suas várias funções.....	85
4.1.3 O clítico ci em suas funções atualizante e lexicalizante.....	86
4.1.3.1 Análise dos resultados.....	87
4.1.4 Os acusativos de terceira pessoa.....	89
4.1.4.1 Análise dos resultados.....	90

4.1.4.1.2 As topicalizações temáticas nos três grupos de falantes.....	92
4.1.4.1.3 Os deslocamentos à direita.....	94
4.1.5 Conclusões de Raso e Vale (2009).....	96
4.2 Os resultados obtidos no presente estudo.....	97
4.2.1 O pronome ci com valor locativo.....	99
4.2.2 O pronome ne em suas várias funções.....	99
4.2.2.1 Análise dos resultados do clítico ne.....	100
4.2.3 O pronome ci com valor atualizante e lexicalizante.....	102
4.2.3.1 Análise contrastiva dos resultados.....	104
4.2.4 Os acusativos de terceira pessoa.....	113
4.2.4.1 Análise contrastiva dos resultados.....	115
5 CONSIDERAÇÕES FINAIS.....	122

1 INTRODUÇÃO

1.1 Delimitação do objeto e objetivos

Esta dissertação de Mestrado propõe-se abordar o tema da erosão linguística de italianos cultos em contato prolongado com o português do Brasil (doravante PB). Será feita uma análise comparativa com estudos anteriores de Raso e Vale (2007, 2009) e será analisado em específico um grupo de clíticos: o pronome *ci* em seus valores atualizante, lexicalizante e locativo, os pronomes *ne* em suas funções partitiva, argumental e locativa e os pronomes acusativos de terceira pessoa, a dizer: *lo, la li, le, l'*.

A erosão linguística como entendida neste estudo é a perda não patológica na proficiência da língua materna, a L1, por parte de um indivíduo, isto é, uma perda intrageneracional, como definida por Kopke e Schmid (2004). No estudo aqui apresentado a L1 analisada será o italiano, doravante It. Pretende-se verificar até que ponto esta perda é devida ao influxo da língua de contato L2, aqui o PB, ou à reestruturação natural do enunciado por falta de insumo da L1 ou pelo insumo de uma L1 já por si sujeita a erosão¹.

Objetivos desta dissertação são verificar as conclusões de Raso e Vale e tentar esclarecer o que desencadeia a erosão linguística; qual o peso dos fatores internos e qual o daqueles externos; se há uma gradualidade na perda, ou seja, se a perda continua ao longo do tempo, ou se após um certo período há uma estagnação no processo de erosão. As escolhas metodológicas que serão citadas ao longo do trabalho servirão de guia para responder a estas perguntas.

O trabalho se baseia na análise de um *corpus* por nós coletado e transcrito entre 2008 e 2009 composto por 21.298 palavras de interações espontâneas de informantes italianos cultos, ou seja, que tenham concluído até o ensino médio, que é de cinco anos, na Itália, e possuam título superior². As transcrições estão gravadas no CD que é anexado à cópia desta Dissertação. Este *corpus* foi comparado com trechos extraídos do *C-ORAL-ROM italiano*³, para um total de 21.224 palavras.

Esta pesquisa se insere no projeto sobre erosão linguística coordenado por Raso junto à UFMG.

1 Veja-se a *Hipótese da Ilha Deserta* citada adiante.

2 Os motivos destas escolhas e daquela do *corpus* de comparação serão explicitados adiante.

3 Cresti; Moneglia (2005)

1.2 Justificativa

A presente pesquisa mostra-se relevante por dois aspectos. Em primeiro lugar os estudos sobre a erosão linguística de línguas em contato com o PB são escassos, portanto pode-se revelar como um âmbito de estudo promissor. Além disso o estudo da erosão linguística poderia permitir uma correlação interessante com os estudos sobre aquisição, e a análise de uma língua específica pode auxiliar nas investigações de outras línguas, pelo menos as mais próximas ao It.

Em segundo lugar, existe a possibilidade de que este estudo desencadeie aplicações práticas no ensino linguístico. Os docentes não nativos de It muitas vezes não chegam a adquirir totalmente, por sua dificuldade, o inteiro sistema pronominal italiano nas partes mais marcadas, assim como todos os aspectos marcados do italiano, atingindo uma formação incompleta e assim repassando a mesma a seus alunos. Outrossim os próprios professores nativos, por falta de *input* ou pelo contato constante com um It já por si sujeitos à erosão, acabam por serem ainda mais sujeitos à erosão e perderem assim as formas mais marcadas.

Conseguir detectar quais traços do sistema estão mais sujeitos à erosão e buscar uma explicação para isso possibilitaria a criação de material didático e exercícios de suporte ou situações específicas que possam servir de "antídoto" à erosão.

1.3 Organização da Dissertação

Além desta breve seção introdutória a Dissertação será dividida em outras quatro partes. No segundo capítulo, será feita uma revisão da literatura, tanto a respeito da erosão linguística que da aquisição, tanto sobre o comportamento das línguas em contato. Abordaremos também a hipótese que se quer verificar nesta pesquisa, ou seja, a Hipótese da Regressão de Jakobson⁴, e discutiremos sobre os pontos de interesse do sistema dos clíticos em italiano. O terceiro capítulo será dedicado às escolhas metodológicas, como aquela por uma pesquisa de *corpus*, e quais problemas foram enfrentados e como. No quarto capítulo, analisaremos os resultados obtidos: na primeira parte, serão fornecidas as conclusões das pesquisas anteriores, em seguida serão evidenciados nossos dados que serão comparados com aqueles de Raso e Vale. O último capítulo é dedicado a algumas considerações finais sobre os rumos teóricos da pesquisa e sobre a possibilidade de realizar estudos futuros mais aprofundados, com a ampliação do *corpus* e a análise de outros traços do sistema que não

4 Jakobson (1941).

foram abordados até o presente momento.

2 REVISÃO DA LITERATURA

Esta seção busca fornecer ao leitor um panorama do que se entende por erosão linguística em suas várias acepções, para em seguida focalizar o discurso sobre a erosão de L1 em falantes cultos. O fenômeno é então contextualizado no âmbito do bilinguismo, haja vista que na pesquisa é observada a erosão da L1 It no contexto da L2 PB. Em seguida, enfoca-se a relação entre erosão e aquisição, seguindo a teoria subjacente aqui adotada, a Hipótese da Regressão de Jakobson. Juntamente à revisão da literatura são discutidas questões teóricas e metodológicas sobre os pronomes em italiano, em específico os clíticos, objeto de nosso estudo; sobre os problemas que estes desencadeiam e sobre como decidimos enfrentá-los, isto para facilitar a compreensão dos capítulos seguintes e das escolhas tomadas nesta dissertação. Finalmente, após um debate metodológico, faz-se uma revisão sobre os estudos de erosão do PB e do It.

2.1 A erosão linguística

A erosão linguística é definida por Köpke e Schmid (2004, p. 5, tradução nossa) como "a redução não patológica no desempenho em uma língua que fora adquirida anteriormente por um indivíduo, ou seja, uma perda intrageracional"⁵. Esta definição tornou-se um consenso entre os linguistas somente em tempos recentes, pois o termo pode indicar fenômenos diferentes que serão abordados nas próximas páginas. Antes de mais nada, é importante distinguir a perda, total ou parcial, de uma língua devido a patologias diferentes, como lesões cerebrais. Há vários estudos⁶, seja de tipo médico que linguístico, que se interessam por este tipo de dano, possibilitando um melhor entendimento dos mecanismos cognitivos, neuronais e cerebrais que servem de fundamento para as estruturas da linguagem.

Contudo, como na realidade na afasia os danos cerebrais atingem alguns aspectos específicos do sistema linguístico, e não o sistema como um todo, neste estudo a erosão devida a problemas patológicos não será contemplada. Além disso, geralmente estes tipos de danos produzem um efeito imediato, e não gradual, portanto não é possível fazer uma comparação inversa com o processo de aquisição linguística.

Sendo a erosão um campo de pesquisa bastante recente, as definições ainda são

5 *"the non pathological decrease in proficiency in a language that had previously been acquired by an individual, i.e. intragenerational loss"*. Para uma discussão mais aprofundada sobre o assunto veja-se também: Seliger e Vago (1991); Schmid *et al.* (2004).

6 Veja-se, por exemplo, no âmbito da discussão linguística, os seguintes estudos: Obler e Mahecha (1991); Dressler (1991).

parciais e não compartilhadas por todos. Nas próximas páginas será feita uma revisão dos pontos principais que levaram a erosão linguística de mero campo de interesse de alguns poucos estudiosos a se tornar uma das novas áreas entre as mais promissoras nas pesquisas linguísticas.

2.1.1 Breve revisão histórica

O início do interesse pela erosão linguística deve-se ao estudioso Richard Lambert, organizador da histórica conferência na Universidade da Pensilvânia, intitulada "A perda das habilidades linguísticas", e o lançamento posterior de suas atas (Lambert & Freed, 1982). O objetivo deste encontro era discutir os principais itens em que ocorria a erosão na L2 e debater sobre futuras áreas de pesquisa, sempre no campo da L2. Na época, as observações sobre a erosão faziam-se prevalentemente através da introspecção, mas o livro e a conferência assumiram tal importância que constituíram-se como uma pedra basilar, ainda hoje relevante em suas contribuições. De fato, já se discutira sobre transferência e morte linguística, perda linguística patológica, fatores políticos e sociais e princípios metodológicos sobre como seriam testadas as hipóteses na base de disciplinas adjacentes. O mérito das conclusões a que os estudiosos chegaram foi justamente perceber que a erosão cobre diversos tipos de perda linguística, e as razões para esta perda podem ser várias.

A partir daquele momento, foram criadas instituições para o estudo da erosão linguística: nos Estados Unidos fundou-se o *Language Skill Attrition Project* e na Europa, especificamente na Holanda, Richard Lambert e seu colega Theo van Els organizaram um importante *workshop* de dois dias na Nijmegen University em outubro de 1982. Pela primeira vez surgiu a hipótese de que a erosão não poderia ser estudada meramente do ponto de vista linguístico, mas deveria levar em conta fatores cognitivos e ser entendida como componente integrante do processo de aquisição de outra língua. Um novo simpósio, em 1986, em Kerkrade, na Holanda, o *Language Loss Symposium*, foi pouco produtivo para a discussão teórica sobre a erosão da L1, já que os trabalhos apresentados concentravam-se sobre estudos pilotos de transferência linguística, morte de dialetos e erosão de L2.

Inicialmente, os estudos sobre erosão convergiam sobre a perda da L2, e os pesquisadores provinham em, sua maioria, da área de ensino de línguas estrangeiras. De fato, outra distinção importante que foi sendo feita ao longo dos anos, foi aquela entre erosão de L2 e erosão de L1. A primeira é aquela que muitas pessoas podem experimentar quando aprendem

uma língua estrangeira, na escola, por exemplo, ou em uma estadia no exterior, e após algum tempo começam a perder suas habilidades naquela língua quando não mais têm a chance de praticá-la, até o ponto em que "esquecem", às vezes totalmente, aquela L2⁷. É um campo ainda hoje muito explorado, mas que, com o passar dos anos, ficou sempre mais restrito aos estudos sobre aquisição de língua estrangeira e sua perda. A erosão de L1, como será discutido adiante, é a perda gradual da língua materna, que pode ter como causa influxos externos, o contato linguístico, por exemplo, ou internos, a re-estruturação da língua por ausência de insumo.

Voltando à história dos estudos sobre erosão, que levou ao panorama atual, o final dos anos oitenta, e também o período dos anos noventa, foi caracterizado por pesquisas isoladas sobre erosão, e muitos pesquisadores, após terem concluído seus doutorados, abandonaram este campo. Somente em 1989, em dois diferentes periódicos: *Studies on Second Language Acquisition* 11:2 e *Review of Applied Linguistics*, apareceram alguns artigos que ilustravam experiências de erosão linguística de L1 em um ambiente de L2.

A grande contribuição para a afirmação da erosão de L1 como campo de pesquisa próprio deu-se graças ao empenho de dois estudiosos, Seliger & Vago (1991), e seu volume *First Language Attrition*, que quis levar no próprio título o empenho para que este objeto de estudo ganhasse *status* específico. Pela primeira vez foi dada uma definição de forma unívoca da erosão linguística, e estudiosos foram incentivados a publicarem suas pesquisas. A partir deste momento e nos dez anos seguintes, os estudos recomeçaram, e foi discutida e concebida a distinção entre o *shift* e a erosão linguística propriamente dita. Como foi sendo percebido, outra subdivisão haveria de ser feita: aquela entre a erosão de L1 que afeta uma comunidade linguística inteira, como no caso de grupos de migrantes que se transferem para outro país, no qual constituem um grupo linguístico minoritário, e aquela de um único indivíduo. Pode acontecer que, de uma geração para outra, a L2 vá prevalecendo no uso sobre a L1 da comunidade de imigrantes: este é o *shift* linguístico, ou seja, quando a L1 é gradualmente substituída pela L2. O processo pode passar por várias etapas e pode até levar ao que é chamado de morte da língua⁸ dos migrantes, mas acontece sempre entre gerações, podendo ser da primeira para a segunda, ou até da segunda para a terceira. Já a erosão linguística é a perda gradual da língua materna, mas sempre no espaço de uma geração, e pode acontecer em nível de comunidade ou de um único indivíduo.

7 Veja-se por exemplo Tomiyama (1999, pp. 59-79).

8 *Language death*.

Os anos noventa, se por um lado caracterizaram-se por pesquisas solitárias, por outro foram aqueles em que o grande público teve acesso ao tema, com a ampla divulgação de vários encontros que ocorreram no começo daquela década. O interesse, contudo, diminuiu com o passar do tempo e somente na virada do milênio o tema retornou a ser objeto de debates.

Na última década, os estudos de erosão linguística concentraram-se sobre a erosão de L1⁹, que é o objeto deste trabalho. As discussões focalizaram dois grandes temas: um de tipo teórico, o outro metodológico. As questões que norteiam as várias pesquisas são:

O que desencadeia a erosão?

A redução é devida somente ao contato ou o processo de reestruturação interno à língua devido à ausência de insumo tem uma força maior?

Há uma gradualidade na perda ou após um certo período há uma estagnação do processo de erosão?

Qual o melhor método, ou os melhores métodos, de coleta de dados?

Qual método permite identificar os sinais de erosão e não confundi-los com outros fatores?

Estas são também as perguntas às quais esta pesquisa se propôs, pelo menos em parte, responder.

2.1.2 A erosão linguística da L1

Como foi visto acima, o tema da erosão linguística de L1 levou vários anos para ser focalizado e para que ganhasse status de ramo da linguística. Os estudos nestes últimos anos estão-se proliferando e cada vez mais é necessário que os conceitos sejam claros para os pesquisadores: por isso das discussões, não somente metodológicas, mas mesmo teóricas, serem sempre presentes na produção bibliográfica e nos encontros que são promovidos¹⁰.

Assim sendo, é importante um parêntese terminológico, para esclarecer a escolha do termo *erosão* em relação àquele de *atrimento*. O termo usado em inglês é *attrition*, o que sugere um desgaste, neste caso da língua, devido a uma espécie de fricção entre duas partes, ou seja

9 Os volumes de referência para a erosão da L1 que discutem sobre as questões teóricas acima postas e são acompanhados por vários estudos sobre o tema são: Seliger e Vago (1991, reimpressão 1999); Köpcke e Schmid (2004), além de vários artigos publicados, especialmente nas revistas *International Journal of Bilingualism* (Vol. 8:3, 2004) e em vários números da *Bilingualism: Language and Cognition*.

10 A título de exemplo, sobre estas discussões e propostas, citamos Schmid (2004, pp. 349-362).

de uma língua por outra, neste caso, no contato entre as línguas. Mas a tradução em línguas como o português e o italiano deixa espaço para a escolha entre *atrído/attrído* e *erosão/erosione*, e até mesmo em inglês é às vezes encontrado o termo *erosion*. Segundo Mello e Raso (no prelo, p. 141), o termo atrído tem a vantagem de ser politicamente correto, não deixando margem para uma interpretação depreciativa do fenômeno. Contudo, sua desvantagem consiste em prefigurar uma explicação do mesmo, restringindo-a somente para aquela do contato linguístico e da influência de uma língua sobre a outra. O termo erosão, por outro lado, simplesmente indica o desgaste de uma língua, enquanto usado como termo técnico é, portanto, não conotado, e permite mais de uma explicação do fenômeno, como a perda de insumo, por exemplo. Por concordar com esta argumentação, o termo usado nesta dissertação será aquele, talvez menos difundido, mas a nosso ver mais apropriado por não prefigurar uma explicação do fenômeno, de *erosão* linguística.

2.1.2.1 Hipóteses teóricas sobre a erosão linguística de L1

Como foi visto, a erosão linguística de L1 pode dar-se em dois contextos distintos:

O primeiro seria aquele de um grupo de imigrantes, ou até mesmo de um indivíduo que, vivendo em um ambiente de L2, devido ao contato com outra língua, acaba por perder em alguma medida aspectos da sua L1. Seria portanto o contato linguístico uma das causas da erosão; e em que medida isto acontece será visto adiante. Deve ser desde já esclarecido, contudo, que por contato linguístico não necessariamente está-se falando de contato entre grupos. Pode-se entender com isto também o que acontece no cérebro do falante quando há um encontro de códigos diferentes, e os efeitos que se dão neste encontro, seja a interferência sobre a L1 de uma L2 que venha a ser adquirida, ou em cujo ambiente se viva, seja o efeito contrário, da L1 sobre a L2. Os contextos de bilinguismo são aqueles com que a maioria dos estudiosos de erosão têm se defrontado, justamente pela possibilidade de sistematizar as pesquisas e verificar hipóteses.

Outra hipótese, de difícil investigação mas muito instigante do ponto de vista teórico, é a Hipótese da Ilha Deserta¹¹: um falante poderia perder sua competência linguística por ausência de insumo desta, ou de qualquer outro insumo linguístico, como no caso de alguém que se encontrasse abandonado em uma ilha deserta. Esta hipótese, na prática

11 Isurin (2000); Sharwood Smith & van Buren (1991, p. 22).

impossível de ser testada, deve sua importância ao fato de que em uma situação como aquela descrita, poderiam-se isolar os efeitos da falta de insumo daqueles de uma língua de contato, e desta forma verificar qual o efetivo peso do contato na erosão linguística e qual aquele de uma re-estruturação linguística ditada por características universais da linguagem.

A grande dificuldade dos estudiosos reside em separar estes dois aspectos para poder entender qual fator está pesando no desencadear da erosão linguística e qual a força deste peso. Entender até que ponto o contato linguístico tem influência na L1 e até que ponto a língua é reestruturada por ausência de insumo, ou por um insumo por si já sujeito à erosão, é uma das grandes questões que os pesquisadores têm que enfrentar.

Uma maneira de minimizar os efeitos do contato e investigar outros fatores foi aquela utilizada por Raso e seu grupo de pesquisa, do qual esta dissertação faz parte, que tenta utilizar como informantes somente pessoas cuja aquisição da língua materna seja, com certeza, completa¹², de forma a evitar que outros elementos, como uma aquisição incompleta, possam influir sobre os resultados encontrados.

2.2 Bilinguismo: contato e aquisição

Entende-se por bilinguismo, ou multilinguismo, o uso de duas ou mais línguas por parte de um indivíduo ou de uma comunidade (Appel e Muysken, 1987). A primeira distinção a ser feita é entre bilinguismo sociatal, ou seja, quando uma comunidade inteira, ou uma parte significativa dela, é bilíngue, e o bilinguismo individual, entendendo-se com isto o contato de duas ou mais línguas no cérebro de um indivíduo (Appel e Muysken, 1987).

Serão aqui abordados os conceitos basilares de como acontece o bilinguismo, ou seja, através do contato e da aquisição, e de quais são os aspectos cognitivos fundamentais que ocorrem na mente de um indivíduo bilíngue. Isto porque, na presente pesquisa, será analisada a fala de sujeitos bilíngues, e um dos objetivos é tentar discriminar o que, na erosão linguística destes indivíduos, é devido ao contato e à aquisição e o que, a outros fatores.

Segundo Appel e Muysken (1987), o contato linguístico leva sem dúvidas ao bilinguismo. Este contato pode acontecer de várias formas, como por via de migrações de uma pessoa ou grupo: pode-se citar, a título de exemplo, o caso de um italiano que se muda

12 Adiante serão explicitados os detalhes dos requisitos dos informantes com a explicação de cada escolha.

para o Brasil e se encontra em uma situação em que a L1 por ele falada estará circunscrita a poucas situações, familiares ou talvez de trabalho, enquanto a L2 terá um peso sempre maior em sua vida, condicionando fortemente sua produção linguística. A outra situação de contato, aquela que acontece no cérebro de cada um, é até mais comum do que se pensa: advém toda vez que aprendemos uma língua estrangeira, por exemplo, em um curso de línguas.

A questão do contato linguístico não pode ser separada daquela da aquisição de L2. Aqui a distinção a ser feita é aquela entre a aquisição simultânea, ou aquisição bilíngue precoce, e a aquisição de L2 em idade adulta. A primeira é aquela que ocorre quando as duas ou mais línguas são aprendidas mais ou menos contemporaneamente, na primeira infância, como no caso de filhos de pais imigrantes que aprendem a L1 em casa e a L2 fora do meio doméstico. Quando há aquisição em idade adulta, ou aquisição sucessiva, geralmente supõe-se que o indivíduo possua uma L1 organizada em um sistema já estável e em seguida venha a aprender a L2.

Alguns autores¹³ fazem uma distinção ulterior entre aquisição e aprendizagem de L2. A aquisição daria-se de forma espontânea, em um contexto de L2 em que o insumo viria naturalmente do convívio diário do aprendiz, como acontece nas imersões. O exemplo é o de um falante que se encontra em um país estrangeiro e acaba por adquirir uma nova língua através da exposição à L1 local. A aprendizagem de L2, ao contrário, acontece em situações de ensino formal, de forma consciente, como nos cursos de línguas em que o aprendiz escolhe e aprende gradualmente uma nova língua.

Neste trabalho não distinguiremos nem nos delongaremos sobre aquisição simultânea, pois todos os informantes mudaram-se para o Brasil já em idade adulta, e justamente a idade foi um dos critérios da escolha destes participantes. Da mesma forma, não é pertinente para esta pesquisa discriminar aquisição e aprendizagem, pois não foi feita esta diferenciação no perfil sociolinguístico dos informantes.

Serão abordados a seguir alguns conceitos básicos sobre a aquisição de L2 que serão úteis no decorrer da discussão.

Weinreich (1953) introduz o conceito de *interferência*. Com este termo o estudioso entende a reorganização dos padrões que resultam dos desvios derivantes de elementos da L1 que se inserem nas partes mais estruturadas dos domínios da linguagem, como o sistema

13 Veja-se por exemplo Krashen (1981); Klein (1986); Ellis (1992); Ellis (1992); Chini (1995), Cook (1996); Ellis (1998); Mitchell, Myles (1998); Lightbounwn, Spada (1999); Bettoni (2001); Doughty, Long (eds, 2003); Hendriks (2005), que abordam todos em detalhes as questões referentes a aquisição de segunda língua.

fonológico, morfológico, sintático ou até mesmo algumas áreas do vocabulário¹⁴. O que é sugerido é que a L1 influencia na aquisição da L2.

É necessário agora introduzir um outro conceito fundamental: a *interlíngua*, criado por Selinker (1972), que designa a versão ou variedade da língua-alvo que faz parte do conhecimento implícito, ou competência do aprendiz de L2. Durante todo o processo de aprendizagem de uma língua estrangeira, o falante passa por uma série de interlínguas até atingir um estágio de proficiência avançado. A interlíngua deve ser entendida como um sistema linguístico independente, resultante das tentativas de produção por parte do aprendiz de uma regra da língua-alvo. Sendo caracterizada como sistema, Selinker postula a interlíngua como a competência em seus vários estágios, que leva o aprendiz a buscar as regras e os princípios que controlam a língua que ele está aprendendo e que possivelmente podem explicar seus desvios. Por outro lado, a interlíngua é entendida pelo estudioso também em suas características dinâmicas de um processo *in fieri*, e nos quais podem ser identificados cinco aspectos cognitivos fundamentais:

A *transferência* de regras ou termos da L1, isto é, a influência direta da língua materna sobre o sistema que está se formando.

A transferência de treinamento, devido ao tipo de ensino ao qual o falante é submetido, e que pode influenciar na produção, ou superprodução, de certos construtos por parte do aprendiz.

As estratégias de aprendizagem da L2, que são as maneiras em que o aprendiz se põe diante da língua a ser adquirida, por exemplo através de processos de simplificação da morfologia.

As estratégias de comunicação na L2, ou seja, como o aprendiz tenta se comunicar com os nativos utilizando os meios de que dispõe.

A hipergeneralização de regras da L2: é um fenômeno conhecido também como hipercorreção, em que o falante, na tentativa de ser correto, acaba por utilizar uma forma que naquele contexto é incorreta, mas que ele interpreta como correta.

Outro conceito importante para Selinker (ano 1972) é aquele de *fossilização*, que diferencia o aprendizado de L1 daquele de L2. A fossilização é, segundo o autor, um mecanismo que existe na estrutura psicológica latente do aprendiz de L2, incluindo estruturas,

14 Weinreich (1953, p. 1).

regras e subsistemas que o sujeito retêm em sua interlíngua, ou seja, todos aqueles aspectos da interlíngua que se tornam entrelaçados e permanentes, independente da quantidade de instruções formais que o aprendiz receba na língua-alvo. "O mecanismo de fossilização é responsável pelo fenômeno do regular reaparecimento no desempenho produtivo da interlíngua de material linguístico que se pensava ter sido erradicado"¹⁵ Selinker (1972, p. 49, tradução nossa).

2.2.1 Ordem e sequência aquisicional de L2

No quadro geral da interlíngua, há um consenso entre os pesquisadores¹⁶ de que haja uma ordem aquisicional e uma sequência em que isso ocorre. As primeiras produções dos aprendizes de L2 geralmente são precedidas por um período de silêncio, em que o sujeito analisa o insumo a que é exposto, buscando as partes mais salientes e memorizando, muitas vezes através da fala consigo mesmo, em uma espécie de forma de treinamento, as estruturas que lhe são mais úteis e frequentes no uso cotidiano. O segundo momento é caracterizado pelo uso de rotinas formulaicas, que se tornam com o tempo padrões de fala simplificados. A partir deste momento, o desenvolvimento na aquisição dependeria de cada falante, mas em geral, quando se inicia a produção por parte do aprendiz na L2, manifestam-se as seguintes fases¹⁷:

Fase pré-básica: a organização do enunciado é nominal. As frases são breves e muito elementares, pois o aprendiz não possui todos os meios necessários para expressar-se e recorre à ajuda do nativo ou a gestos. A estrutura dos enunciados é caracterizada pela ausência do verbo, às vezes há somente a parte do *comentário*¹⁸, comunicativamente nova, ou no máximo pode-se observar a ordem tópico-comentário com contiguidade dos dois elementos sem ligações de tipo sintático. Como o aprendiz possui um léxico extremamente reduzido, utiliza o recurso das rotinas formulaicas.

Fase básica: a organização do enunciado é ainda de base pragmática, o verbo é o núcleo da frase, mas geralmente não é conjugado e muitos traços morfológicos não são expressos. Há uma variedade sintática em que as palavras funcionais, como artigos e

15 "The fossilization mechanism accounts for the phenomenon of the regular reappearance in IL productive performance of linguistic material which was thought to be eradicated"(Selinker 1972, p.49).

16 Remetemos novamente os detalhes aos trabalhos de: Krashen (1981); Ellis (1992); Klein (1986), Chini (1995), Bettoni (2001), Cook (1996), Doughty, Long (eds, 2003), Ellis (1998); Hendriks (2005), Lightbown, Spada (1999), Mitchell e Myles (1998).

17 Veja-se Cook (1993 e Ellis (1994, cap. 3).

18 O *comentário* é um termo usado para descrever a estrutura informativa e é o que é dito a respeito do *tópico*, que se refere a algo novo que é introduzido na ilocução.

preposições, são praticamente ausentes, a morfologia é usada nas formas não marcadas e a expressão acontece através de meios lexicais. Geralmente o foco, a informação nova no enunciado, ou o tópico, são postos no final do enunciado, segundo a ordem comentário-tópico-foco, pondo em evidência desta forma o que é mais saliente. Os enunciados nesta fase são geralmente justapostos, predominando a parataxe. O léxico vai sendo ampliado e o aprendiz começa a distinguir classes de palavras, apesar de ser incerto e pouco presente o uso de elementos funcionais.

Fase pós-básica: é caracterizada por um progressivo uso da morfossintaxe, pelo uso de verbos conjugados e elementos funcionais cujo emprego é ditado pelas normas específicas da L2. Os enunciados possuem uma estrutura mais próxima daquela da língua-alvo. A morfologia é mais rica e complexa e o aprendiz reconhece o valor morfológico e funcional dos morfemas, apesar de muitas vezes ocorrer uma hipergeneralização de alguns morfemas, entres os mais regulares (em inglês, por exemplo, um caso clássico é aquele do *past tense* do verbo *to go*, que os aprendizes de L2 muitas vezes hipercorrigem em *he go-ed* ao invés de *he went*). A variedade pós-básica pode ir desde estágios intermediários, até aqueles avançados e quase nativos.

A ordem de aquisição de uma L2 parece seguir um padrão similar em várias línguas estudadas: seja em estudos longitudinais ou transversais¹⁹, emergiu que há uma ordem de aquisição para as estruturas. Não é a forma de um morfema em si, ou a estrutura sintática a ser adquirida, mas sua função e seu grau de marcação²⁰, geral ou em relação a uma língua específica, que vão ditar quando esta forma vai ser adquirida. Em inglês, por exemplo, a função de marcar o plural possivelmente é mais relevante comunicativamente, mais simples do ponto de vista cognitivo ou menos marcada linguisticamente em relação à função de assinalar o possessor ou indicar o acordo com o sujeito de terceira pessoa singular. Por exemplo, Krashen (1977) pressupõe que haja em inglês a seguinte ordem natural de aquisição dos morfemas:

- (1) *-ing* / *-s* plural / copula → auxiliar *be* / artigo *the/a* → passado irregular (ex. *went*) →

19 Os primeiros estudos sobre a ordem de aquisição da L1, de tipo longitudinal, foram aqueles de Brown (1979), que identificou um padrão de aquisição para morfemas do inglês. Em seguida grupos de estudiosos se debruçaram sobre a ordem de aquisição de várias L2, especialmente o inglês, confirmando que a ordem é parecida com aquela de aquisição de L1: vejam-se a este respeito os estudos de Krashen (1977) e Dulay, Burt, Krashen (1985), Chini (1995), Bettoni (2001), Cook (1996), Doughty, Long (eds, 2003), Ellis (1998), Ellis (1992), Hendriks (2005), Lightboun, Spada (1999), Mitchell, Myles (1998).

20 Sobre o conceito de marcação falaremos adiante.

passado regular (-ed), 3ª pessoa sing -s, possessivo 's

Como se pode observar, a desinência -s do plural antecede a desinência -s da terceira pessoa o 's possessivo é o último a ser adquirido.

Graças aos estudos sobre ordem de aquisição foi possível elaborar hierarquias de aquisição. Estas ordenam as estruturas aprendidas de forma sequencial, ou seja, para que as estruturas de um grupo sejam aprendidas, é necessário antes ter aprendido aquelas do grupo anterior. Isto, de uma certa forma, subentende que há mecanismos universais que são responsáveis pela ordem de aquisição.

Estudos relativos especialmente ao inglês²¹ estabeleceram uma série de sequências de aquisição, ou evolucionais; entre as mais estudadas por sua complexidade no subsistema, citamos por exemplo:

A aquisição das estruturas interrogativas: os principais tipos de perguntas são 1. as parciais (*wh-questions*) que vertem sobre um constituinte; 2. as totais ou polares (*yes/no questions*) que vertem sobre o inteiro conteúdo do enunciado. Geralmente na gramática inglesa estas perguntas requerem a inversão verbo-sujeito e a utilização do auxiliar *do* flexionado, à exceção de que se trate de verbos com cópula ou modal, para os quais há somente a inversão. Os trabalhos sobre a aquisição da estrutura da interrogativa no inglês L2 mostraram que há uma grande dificuldade por parte dos aprendizes na aquisição da inversão, mesmo em sujeitos cujas línguas L1, como o francês, possuem o mesmo tipo de estrutura. A inversão parece ser um tipo de estrutura interrogativa marcada nas línguas do mundo, que se opõe à ordem de base das línguas SVO e SOV. Quando esta estrutura é aprendida pode acontecer que seja superextendida a formas que não a requerem, como as interrogativas encaixadas.

A frase negativa apresenta muitos problemas para os aprendizes de inglês L2: inicialmente a tendência é uma negação externa ao enunciado com um *no/not* colocado antes ou depois do elemento negado; em seguida os aprendizes utilizam um tipo de negação interno à frase, preverbal e sem o uso do auxiliar; após este estágio há uma negação posposta ao auxiliar, que pode ser *be*, *can* ou *do*; finalmente, chega-se à negação posposta ao auxiliar *do* analisado e flexionado em pessoa e tempo verbal.

Como pôde ser visto, o auxiliar *do* comparece tardiamente nas várias estruturas, em

21 Vejam-se novamente os resultados de Krashen (1977) e Dulay, Burt, Krashen (1985) e também de Cook (1993).

fases pós-básicas. Também em outras línguas²², como o sueco e o alemão, a posição da negação torna-se mais correta somente em um momento posterior. É importante salientar que a sequência de aquisição de uma L2 parece não depender da L1 do aprendiz. Por exemplo, os alemães, que possuem a negação pós-verbal, durante a aprendizagem do inglês, passam por uma fase em que utilizam a negação pré-verbal.

Será analisada em seguida uma das sequências de aquisição do italiano L2 e serão explicitadas, de forma mais clara, quais as implicações que este discurso tem para a pesquisa aqui apresentada.

2.3 Sequência de aquisição dos clíticos no italiano L2

Muitas das pesquisas sobre as sequências de aquisição do italiano L2, foram coordenadas pelo grupo de estudo de Anna Giacalone Ramat²³, no Projeto de Pavia. Foram teorizadas sequências de aquisição para o grupo dos clíticos, para tempos e modos verbais, gênero, negação, subordinação, anáfora e ordem sintática marcada e foi feita uma comparação entre a aquisição do italiano L1 e aquela do italiano L2²⁴. Remetemos à bibliografia citada o conjunto de resultados obtidos, e nesta seção aprofundaremos somente a sequência de aprendizagem dos clíticos, que são objeto do trabalho aqui apresentado.

O discurso sobre como um falante de L1 consegue obter uma determinada proficiência em uma L2 é muito importante no contexto da erosão linguística de L1. Como já foi introduzido anteriormente, a investigação deste fenômeno geralmente é feita analisando o contato linguístico com a L2. Conhecendo a sequência de aquisição da L1, e daquela da mesma língua operada por estrangeiros, e comparando o que é perdido ou reestruturado na erosão, é possível verificar se estas sequências são especulares. Isto implica poder verificar se a Hipótese da Regressão de Jakobson (1941)²⁵, que, em substância, afirma que os itens linguísticos adquiridos antes são perdidos por último, se sustenta, ou se entram em jogo outros fatores, e em qual medida. Segundo Montrul (2004, p. 260, tradução nossa) "a gramática sujeita a erosão dos bilíngues, em um certo estágio, ou até mesmo no estágio final fossilizado, assemelha-se à gramática incompleta típica dos estágios intermediários e avançados da aquisição de L2"²⁶. Um estudo comparativo entre as fases de aquisição da L1 que será

22 Remetemos aos estudos de Meisel e Clahsen, em Giacalone-Ramat (1986).

23 Os textos de referência são Giacalone Ramat (1993 e 2003).

24 Giacalone Ramat (2003, pp. 220-252).

25 A Hipótese de Jakobson (1941) será discutida mais aprofundadamente adiante, citando também o estudo de Keijzer (2007) que a testou de forma detalhada.

26 *the eroded L1 grammars of bilinguals at a given state or even the fossilised endstate resemble the incomplete*

estudada como sujeita à erosão, e da mesma língua adquirida por estrangeiros, portanto como L2, poderá mostrar se realmente há especularidade na erosão linguística, e se isto acontece em relação à aquisição dos nativos, como afirma Jakobson, ou se é em relação aos não nativos.

2.3.1 Os clíticos em italiano

O grupo dos pronomes clíticos, ou átonos, é bastante complexo em italiano. Todos possuem baixa saliência fonológica e morfologia bastante complexa, mas alguns deles são logo adquiridos ainda na primeira infância, por seu uso frequente e sua facilidade de assimilação. Outros, ao contrário, são entre os últimos itens morfológicos a serem adquiridos em italiano. Mesmo os nativos tendem a aprendê-los muito tarde e, às vezes, não conseguem fazê-lo por completo, apresentando frequentemente omissões ou desvios. Obviamente, nem todo o grupo apresenta estes problemas, mas nos referimos aos pronomes mais complexos ou que somam mais de uma função, como nos casos dos nexos, das várias funções somadas pelo clítico *ne*, pelo *ci* e pelo *vi*, como será visto adiante com vários exemplos.²⁷

O quadro abaixo mostra o grupo dos clíticos, divididos por funções e pessoas²⁸. Como é possível notar, há muitas homônimas entre as formas que, possuindo a mesma grafia, acumulam usos diferentes.

	SINGULAR				PLURAL			
	1 ^a	2 ^a	3 ^a	Reflex.	1 ^a	2 ^a	3 ^a	Reflex.
Acusativo	mi	ti	lo, la	si	ci	vi	li, le	Si
Dativo	mi	ti	gli, le	si	ci	vi	loro, gli	Si
Genitivo partitivo			ne				ne	
Locativo			ne				ne	
Argumental			ne				ne	
Locativo			ci, vi				ci, vi	
Atualizante			ci				ci	
Pronomes lexicalizantes ²⁹			ci, si e nexos				ci, si e nexos	

grammars typical of the intermediate and advanced stages in L2 acquisition.

27 Sobre os clíticos em italiano e suas dificuldades de aquisição veja-se Sabatini (1985), Berretta (1986), Cipriani *et al.* (1993).

28 Quadro de nossa autoria, baseado em várias gramáticas de referência do italiano.

29 Os pronomes lexicalizantes são clíticos que, ligando-se de forma enclítica a alguns verbos, os modificam semanticamente. Como exemplos podemos citar: *avercela*(estar com raiva de alguém) e *cavarsela*(se virar), que possuem significado muito diferente das formas *avere*(ter) e *cavare*(tirar), constituindo-se como lexemas

Quadro 1. Os pronomes pessoais clíticos divididos por funções.

O subsistema dos clíticos é bastante complexo, marcado, e sua colocação sintática pode ser pré ou pós-verbal, dependendo da forma verbal. Hoje em dia, seu estatuto está variando de pronome para marca verbal, como acontece com os pronomes lexicalizantes. A escala de aquisição do italiano L1, extraída de Berretta (1986), é a seguinte:

- (1) *mi* dativo > *ti* dativo/acusativo > *ci* atualizante > *si* impessoal > *mi/ti/gli* (masculino, singular) > *si* (reflexivo) > *lo* flexionado (*lo* > *la* > *li* > *le*) > *ci* locativo > nexos *me lo/te lo* > dativos de 3ª *gli/le//loro* > nexos *glielo/ce ne* > *ne* partitivo > *ne* argumental > *ne* locativo > *ci/vi* acusativos e dativos de 1ª e 2ª plural

Para facilitar a compreensão do leitor, abaixo citaremos um exemplo para cada forma pronominal acima exposta:

- (1) ***Mi*** (DAT) *presti il libro?*
Me empresta o livro?
- (2) ***Ti*** (DAT) *offro un caffè?*
Te ofereço um café?
- (3) ***Ti*** (ACC) *porto al mare.*
Te levo pra praia.
- (4) ***Ci sono*** (Atualizante) *molte zanzare.*
Há muitos pernilongos.
- (5) ***Si*** (Impessoal) *è fatto tardi.*
Se fez tarde. (Está tarde)
- (6) ***Mi*** (DAT masc. sing) *dici cosa ho fatto?*
Me diz o que (eu) fiz?
- (7) ***Ti*** (DAT masc. sing.) *porto un cappuccino?*
Te trago um cappuccino?
- (8) ***Gli*** (DAT masc. sing.) *ho prestato la giacca.*
A ele emprestei o casaco.
- (9) ***Si*** (reflexivo) *sta lavando le mani.*
Se está lavando as mãos.
- (10) *Puoi accompagnarlo* (acusativo de terceira pessoa) *a casa?*
(Você) pode acompanhá-lo para casa?
- (11) *Vai dal dentista? Ci* (locativo) *vado domani!*

- (Você) vai ao dentista? Vou aí (no dentista) amanhã!
- (12) *Mi presti il libro? Si, te* (DAT) *lo* (ACC) *presto subito*
Me empresta o livro? Sim, a ti o empresto logo.
- (13) *Gli* (DAT 3ª sing. masc.) *hai parlato?*
Com ele (você) falou? (Você falou com ele?)
- (14) *Le* (DAT 3ª sing. fem.) *hai scritto?*
A ela (você) escreveu? (Você escreveu para ela?)
- (15) *Hai portato loro* (DAT 3ª plur. masc.) *la torta?*
(Você) levou a eles o bolo?
- (16) *Glielo* (Nexo dativo e acusativo) *diciamo?*
A ele o dizemos? (Vamos contar para ele aquilo?)
- (17) *Vuoi la torta? Ne* (partitivo) *voglio una fetta.*
(Você) quer o bolo? Dele (eu) quero uma fatia.
- (18) *Parliamone* (argumental) *più tardi.*
Vamos falar disso mais tarde.
- (19) *Sei stato da Giulia? Ne* (locativo) *vengo proprio ora.*
(Você) esteve na casa de Giulia? Da casa dela chego agora mesmo.
- (20) *Ci* (ACC 1ª pl.) *porti con te?*
Nos leva com você?
- (21) *Vi* (ACC 2ª pl.) *accompagno a casa.*
Vos acompanho em casa.
- (22) *Ci* (DAT 1ª pl.) *porti un dolce?*
(Você) nos traz um doce?
- (23) *Vi* (DAT 2ª pl.) *ha ascoltato?*
(Ele) vos ouviu?

É proposta abaixo a escala de aquisição do italiano L2, sempre extraída de Berretta (1986), a título de comparação.

- (2) *ci* atualizante > *mi* dativo/reflexivo > *si* impessoal/apassivador/reflexivo > *ti* dativo/acusativo/reflexivo > *lo* flexionado (*lo* > *la* > *li* > *le*) > nexos como *me lo / te lo* > *ci* locativo > dativos de 3ª > *ci/vi* acusativos e dativos de 1ª e 2ª plural > *ne* em nexos > *ne* partitivo > *ne* argumental > *ne* locativo

Pode-se observar que é bastante similar àquela do italiano L1, mas apresenta também diferenças significativas: o *ci* atualizante é o primeiro a ser adquirido pelos aprendizes de L2,

enquanto nos nativos este somente é aprendido depois que o falante adquiriu o *mi* dativo e o *ti* dativo/acusativo; já as várias funções de *ne* são aprendidas tardiamente, tanto pelos nativos quanto pelos aprendizes de L2.

No ensaio de Berretta (1986), alega-se que os pronomes tônicos são adquiridos antes da morfologia verbal (*io/me, tu/te, lui, lei, noi, voi, loro e questo*, entendido como um "isso" neutro de terceira pessoa) e usados antes dos clíticos. Supõe-se que os clíticos sejam aprendidos juntamente com a morfologia verbal ou como parte dela. Supostamente os adultos, portanto aprendizes de L2, adquirem todos os pronomes tônicos e em seguida todos os clíticos, em uma estratégia inicial de simplificação redutora da morfologia verbal. As crianças, portanto os nativos, ao contrário, utilizam um comportamento de aprendizagem de: tônico x > clítico x > tônico y > clítico y , de acordo com a saliência pragmática das pessoas e com amplo uso no começo de dêixis não verbal e/ou lexicalização, substituindo a flexão verbal.

Nos próximos parágrafos nos dedicaremos a aprofundar alguns dos clíticos, em específico aqueles que serão objeto de análise nesta dissertação: os acusativos de terceira pessoa (*lo, la, li, le*), o *ci* atualizante, lexicalizante e locativo, e o *ne* em suas funções argumental, partitivo e locativo de origem. A discussão explanará a situação destes clíticos no italiano contemporâneo e mostrará a situação específica da língua falada, que é a que nos interessa nesta pesquisa.

2.3.1.1 Os acusativos de terceira pessoa

Segundo Berretta (1985), no italiano atual há um uso muito grande dos clíticos devido à alta frequência de estruturas sintáticas não canônicas, como deslocamentos à esquerda e à direita, ou de clivadas em geral, que levam a um aumento do uso dos pronomes átonos, especialmente os acusativos de terceira pessoa. Vejamos os exemplos abaixo:

- (24) *Hai mangiato tutta la minestra?*
(Você) comeu toda a sopa?
- (25) *La minestra, l'hai mangiata tutta?*
A sopa, (você) **a** comeu toda?
- (26) *L'hai mangiata tutta la minestra?*
(Você) **a** comeu toda a sopa?

O enunciado (24) apresenta a frase canônica, na ordem SVO. No enunciado (25) o

tópico é deslocado para a esquerda, a frase é comunicativamente marcada sobre um ponto de atenção, *la minestra* (a sopa). Em italiano, é obrigatória a retomada pronominal, portanto se faz necessário o uso do clítico *la*. Já o exemplo (26) mostra outro tipo de deslocamento, aquele à direita, e a catáfora pronominal pode ser facultativa³⁰, mas muito usada. Os acusativos de terceira pessoa, *lo, la, li, le*, são, portanto, em conformidade com Berreta (1985), mais frequentes na fala que na escrita pois, por motivos pragmáticos, acaba-se por utilizar muito os clíticos, o que em PB é raro e marcado em termos de registro.

Outro motivo que contribui para aumentar as ocorrências dos clíticos, segundo Berretta (1985), é a frequência dos verbos assim ditos pronominais, como *andarsene* (ir embora), em que o verbo *andare* (ir) é conjugado a um *si* pseudo-reflexivo e a um *ne* locativo de origem, dando vida a um verbo novo; ou também *cavarsela* (conseguir dar conta, conseguir atingir o objetivo), em que o verbo original *cavare* (escavar, extrair) conjuga-se a um *si* pseudo-reflexivo e a um *la* invariável, lexicalizando totalmente o verbo. Nestes casos, "os átonos são muito usados não tanto, ou não somente, porque escolhidos por si, mas porque sua presença é efeito de escolhas de outro nível, sintático e lexical" (BERRETTA, 1993, p. 226, tradução nossa)³¹.

Segundo Berretta (1993), o caso dos clíticos é bastante delicado pois, sendo formas já por si fracas, o fenômeno de desgaste é mais avançado que em relação aos tônicos. Com isto, entende-se que, quando usados com os verbos pronominais, apesar de manterem o comportamento sintático de morfemas semilivres, acabam por se tornarem funcionalmente, simples partes de palavras. Estes mesmos pronomes, quando utilizados na retomada pronominal em casos de deslocamentos, "parecem funcionar como marca de acordo entre o próprio verbo e os nominais, e somente em segundo lugar possuem o papel de marca de caso dos nominais [...] parecem adquirir o estatuto de morfemas verbais, perdendo o já enfraquecido estatuto pronominal"³² (BERRETTA, 1986, p. 227, tradução nossa).

2.3.1.2 O *ci* locativo

Citando alguns exemplos de formas que estão se perdendo, há o caso do *vi* locativo,

30 A discussão sobre os deslocamentos de constituintes à direita é muito complexa e remetemos ao parágrafo 4.1.4.1.3 onde este assunto será abordado mais amplamente.

31 "*gli atoni sono molto usati non tanto, o non solo, perché siano scelti di per sé, quanto perché la loro comparsa é effetto di scelte di altro livello, sintattico e lessicale*".

32 "*sembrano fungere da marca di accordo fra il verbo stesso e i nominali, e secondariamente da marca di caso di questi ultimi [...] sembrano acquisire lo statuto di morfemi verbali, perdendo il già debole statuto pronominale*".

quase sempre substituído pelo *ci* locativo. Russi (2008) esclarece que, inicialmente, havia uma diferença entre o *ci* com significado de "aqui" contraposto ao *vi* significando "ali", mas no italiano moderno eles se tornaram sinônimos e um acabou prevalecendo no uso sobre o outro.

Por outro lado, esclarecemos que ainda permanece uma certa diferença de registro entre os dois pronomes: o *ci* é mais utilizado em uma diafasia e diamesia informal e falada, enquanto o *vi* sobrevive em algumas diamesias escritas, raramente faladas, muito altas.

2.3.1.3 O *ne* e suas funções

Também o *ne* locativo comparece quase que somente ligado a verbos, como *andarsene* (ir-se embora) e portanto, segundo Berretta (1986), pode-se dizer que tenha desaparecido como forma pronominal. O mesmo pode-se dizer do *ne* no uso genitivo partitivo³³. Russi (2008) concorda com esta observação e esclarece as várias funções do *ne*, que pode:

Pronominalizar expressões nominais quantificadas como em (27):

- (27) *Ho portato la torta. Ne vuoi una fetta?*
 Trouxe o bolo. Você quer dele uma fatia?

Pronominalizar sintagmas preposicionais introduzidos por *da* (de) como os complementos causais em um número restrito de construções passivas como *essere* (ser), *rimanere* e *restare* (ficar), *venire* (vir), etc. e o particípio passado de verbos como *affascinare* (encantar), *colpire* (abalar) etc. como no exemplo (28):

- (28) *Ho visto il suo ritratto. Ne sono rimasto affascinato.*
 Vi seu retrato. Fiquei encantado com ele.

Como locativo: este uso quase desapareceu e o pronome comparece somente em combinação com o clítico *si* em verbos como *andarsene* (ir-se embora) e *venirsene* (vir-se embora)³⁴.

Pode-se porém notar que o uso do *ne* locativo de origem ainda não desapareceu completamente: ainda é usado em expressões sim ligadas a um verbo, mas mantendo intacto seu significado original, como em:

- (29) *Da quell'incidente ne sono uscito senza un graffio.*
 Daquele acidente dele (eu) saí sem um arranhão.

33 Este mesmo tipo de argumentação pode ser encontrado em Berretta (1994, pp. 239-270).

34 Em seu livro Russi (2008) dedica o inteiro capítulo 5 à lexicalização de verbos com o pronome *ne*. Aqui não vamos nos delongar neste discurso porque, apesar de ser muito interessante, não será abordado nesta pesquisa, portanto faça-se referência à bibliografia para este assunto.

2.3.1.4 O *ci* atualizante e lexicalizante

Vejam agora as funções do *ci* atualizante e lexicalizante e quais as diferenças que os distinguem.

Em Sabatini (1985, pp. 160-161) esclarece-se que o clítico *ci* originariamente possuía valor adverbial de lugar *qui* (aqui) e caracteriza-se por um uso muito amplo em união com os verbos *essere* e *avere* quando estes não desempenham função de auxiliar. O *ci* portanto acabou por perder seu significado original e assumiu uma função de reforço semântico e fônico para estas formas verbais.

Quando unido ao verbo *essere*, o *ci* mantém uma ligeira nuance de advérbio locativo, com um efeito que Sabatini (1985) chama de propriamente "atualizante". Isto, segundo o autor, acontece quando é descrito um acontecimento específico dos quais são evocados aspectos materiais e localizáveis. Citamos alguns dos exemplos de Sabatini para essa função:

- (29) *Oggi c'è sciopero dei giornali.*
Hoje acontece uma greve dos jornais.
- (30) *C'è il sole.*
Tem sol.

Para Sabatini, o uso do *ci* è definido como "normal e obrigatório com o verbo *essere*"³⁵ (SABATINI 1985, p. 160, tradução nossa) quando possui significado de existência, mesmo se não há uma referência concreta a um lugar, como no exemplo:

- (31) *C'era una volta un re.*
Era uma vez um rei.

Para o estudioso, no uso de *ci* com o verbo *avere* é ainda mais evidente a "função puramente atualizante do *ci* originariamente advérbio de lugar"³⁶ (SABATINI 1985, p. 160, tradução nossa).

Com outros verbos como *sentire* (ouvir) e *vedere* (ver), nas expressões *ci sento* (ouço bem) e *ci vedo* (vejo bem), segundo Sabatini, a função do *ci* é plenamente atualizante, pois não há referências nem de tipo locativo nem às condições ambientais. Já com verbos como *entrare* (entrar), *capire* (entender), *credere* (acreditar), *volere* (querer), em expressões como *c'entra* (tem a ver), *non ci credo!* (não acredito nisso!), *non ci capisco niente* (não consigo entender nada disso), *ci vuole poco* (falta pouco), o *ci* mantém em parte seu significado original de "aqui, nessa questão, a respeito disso". Eliminar o *ci* acabaria por modificar o sentido da frase.

35 "è normale e obbligatorio con il verbo essere".

36 "la funzione puramente attualizzante del ci originariamente avverbio di luogo".

Sabatini, portanto, distingue somente entre o valor ligeiramente "locativo" primitivo do clítico *ci* com verbos, que é dito função "atualizante", e uma outra função que acaba por modificar mais profundamente o verbo do ponto de vista semântico.

Esta segunda função é designada por Russi (2008) como "lexicalizante". Segundo a estudiosa, o *ci* acabou por lexicalizar uma série de verbos. São analisados um por um alguns destes verbos, que são discriminados entre aqueles cujo uso do *ci* é obrigatório, aqueles em que acontece uma especialização semântica e os verbos: *esserci*, *averci* e de percepção.

Com o verbo *essere*, Russi (2008) considera que o *ci* atingiu uma condição de plena gramaticalização, ou seja, tornou-se obrigatório quando o verbo *esserci* possui o valor apresentacional/existencial de "haver, existir". Segundo a estudiosa, há portanto uma dicotomia entre *essere*~*esserci*, o primeiro sendo cópula e verbo auxiliar em oposição ao segundo, que é apresentacional/existencial. Embora o *ci* de *esserci* não atue mais com a função de pronome locativo, pois não substitui um complemento locativo, "não esvaziou completamente sua referência espacial, enquanto expressa o lugar ou, talvez mais apropriadamente, a área da presença e/ou existência de uma dada entidade"³⁷ (RUSSI 2008, p. 161, tradução nossa).

No caso do verbo *avere*, segundo Russi, ainda não foi atingida a plena obrigatoriedade estrutural, pois alguns puristas ainda estigmatizam o *averci* no sentido de "ter, possuir". Na realidade, acreditamos tratar-se não de uma questão de purismo gramatical: ambas as formas tornaram-se aceitáveis, mas seu uso depende de uma questão de registro linguístico. Além disso, a omissão do *ci* ainda não leva, em todos os casos, a sentenças agramaticais: ainda não há a dicotomia entre *avere* e *averci* que é percebida em *essere* ~ *esserci*³⁸.

Segundo Sabatini (1986) alguns verbos de percepção, em específico *sentirci* (estar em condição de ouvir) e *vederci* (ser capaz de ver) encontram-se em um processo de gramaticalização. O *ci* parece ter perdido completamente sua referência locativo/espacial e tornou-se uma marca semanticamente vazia de reforço enfático. Russi (2008), por sua vez, discorda desta visão. Segundo a autora, o *ci* na realidade adquiriu a função de modificar o significado do verbo de base: a presença ou ausência do *ci* em *sentirci* e *vederci* não é uma simples questão de ênfase, mas o *ci* carrega em si um valor semântico, de estado, e gramatical, de marca intransitiva. O *ci* nestes casos pode ser visto como uma marca

37 "it has not been completely emptied of its spatial reference in that it expresses the place or, perhaps more appropriately, the realm of presence and/or existence of a given entity".

38 Para a discussão aprofundada sobre o assunto veja-se Russi (2008, pp. 162-167).

morfossintática que indica uma mudança de valência no verbo e uma passagem do verbo de um significado de percepção a um de puro estado.

Sentirci e *vederci* ainda não são considerados completamente lexicalizados pelo fato de que o *ci* ainda não se tornou um morfema estruturalmente obrigatório. No caso de *averci* o que acontece é ainda uma questão de registro: nas diastratias e diafásias mais altas usa-se o *avere*, enquanto naquelas mais baixas utiliza-se o *averci*.

O caso de maior divergência entre o verbo original e a nova forma com o clítico *ci* é, segundo Russi (2008), o verbo *volere* (querer)~*volerci* (ser necessário). As maiores distinções a serem feitas entre estes dois verbos relacionam-se a: posição do verbo, características semânticas, papel temático do sujeito e natureza do complemento³⁹. Russi (2008), concordando com De Mauro (1999-2000), afirma que *volerci* é um dos mais antigos verbos definidos pelo estudioso como *procomplementari*, portanto seu processo de lexicalização foi completado há muito tempo.

Os pares de verbos *mettere* (colocar)~*metterci* (empregar) e *entrare* (entrar)~*entrarci* (estar relacionado a), apresentam a menor discordância semântica e estrutural, pois as diferenças entre os verbos de base e aqueles com o clítico *ci*, segundo Russi (2008), residem somente em um processo de abstração conceptual, uma transição de um domínio mais concreto para um mais abstrato.

O verbo *stare* (estar, ficar) é discutido à parte. Um de seus significados com o clítico *ci* mantém o significado locativo do pronome, portanto *starci* significa "estar, ficar ali". Assim sendo, não houve ainda um processo de lexicalização concluído, mas somente o início do mesmo.

Contudo, *starci* possui outro significado que Russi (2008, p. 154) denomina "especializado", que é aquele de "ser fácil, especialmente para uma mulher". Este verbo e outros dois, *provarci*, no sentido de "tentar ter um encontro de tipo sexual com alguém" e *farci*, na conotação de "fingir", sofreram uma completa "obrigatorificação"⁴⁰ do *ci*, no sentido que todos os três desenvolveram uma conotação idiomática muito forte, em um processo que até ultrapassou a lexicalização.

Os significados "especializados" sugeridos por Russi podem ser contestados. Dependendo de usos regionais, o verbo *starci* não assume uma conotação tão específica, mas pode ser traduzido com "ficar com alguém", seja homem ou mulher.

39 Para maiores detalhes veja-se Russi (2008, pp. 143-149).

40 "obligatorification".

Como se pode observar, a questão da lexicalização de verbos em presença do clítico *ci* ainda é amplamente discutida e não há um consenso entre os estudiosos sequer sobre a terminologia a ser adotada. Para clareza de exposição, preferimos adotar aqui os vocábulos de uso mais comum e esclarecer como serão empregados: os verbos *esserci* e *averci* serão os únicos dois considerados neste estudo atualizantes, em concordância com o que a bibliografia acima afirma destes verbos; todos os outros que se unam ao clítico *ci*, serão por nós nomeados lexicalizantes, sem diferença de nuance entre aqueles em que o processo está ainda em sua fase inicial ou completamente concluída.

Finalmente, esclareceremos as funções *existencial* e *apresentacional* de *esserci*, citadas por Russi (2008). Quando *esserci* substitui outras formas lexicais que denotam diferentes formas de existência, exercita uma função *existencial*, como nos exemplos abaixo:

(32) *Alla festa c'erano molte persone.*

Na festa havia muitas pessoas.

O enunciado (32) pode ser perfeitamente substituído por outros que mantêm o mesmo significado como (33) e (34):

(33) *Alla festa parteciparono molte persone.*

Da festa participaram muitas pessoas.

(34) *Alla festa andarono molte persone.*

À festa foram muitas pessoas.

Outra função de *esserci* reside na estrutura *apresentacional*, formada pelo verbo *esserci*+SN+*che* pseudo-relativo, com função de tornar menos pesada do ponto de vista cognitivo uma estrutura totalmente nova e remática. O exemplo (35) é perfeitamente gramatical, mas pesado do ponto de vista informacional, pois introduz um elemento novo seguido por uma nova informação. Já a estrutura *apresentacional* do enunciado (36) atenua esse excesso de informações, tornando o primeiro elemento novo ponto de suporte já conhecido para um segundo elemento novo:

(35) *Un signore vuole conoscerti.*

Um senhor quer te conhecer.

(36) *C'è un signore che vuole conoscerti.*

Tem um senhor que quer te conhecer.

No estudo aqui apresentado não serão contemplados todos os clíticos, por razões de tempo e complexidade de análise, mas somente um pequeno grupo para comparação de

resultados com os estudos anteriores. Em específico, serão investigados os seguintes clíticos: os acusativos de terceira pessoa (*lo, la, li, le*), o *ci* atualizante, lexicalizante e locativo e o *ne* em suas funções argumental, partitivo e locativo de origem.

Dentro da sequência de aquisição da L1 portanto, a ordem destes pronomes seria a seguinte⁴¹:

- (3) *ci* atualizante e lexicalizante > *lo* flexionado > *ci* locativo > *ne* partitivo > *ne* argumental > *ne* locativo

Se compararmos esta sequência com aquela de aquisição de L2, veremos que estes pronomes ocupam a mesma posição, como mostra-se abaixo:

- (4) *ci* atualizante e lexicalizante > *lo* flexionado > *ci* locativo > *ne* partitivo > *ne* argumental > *ne* locativo

Observaremos no estudo se sua erosão será especular a esta escala.

2.3.2 A história linguística italiana e o sistema pronominal

A história linguística italiana é muito peculiar e não é essa a sede para resumí-la se não em aspectos relevantes para o trabalho presente⁴². Como é notório, o italiano de hoje é uma evolução do vulgar florentino dos séc. XIII-XIV. Contudo, trata-se de uma evolução separada daquela do próprio florentino. De fato, no séc. XVI, principalmente por obra do humanista veneziano Pietro Bembo e do dono da maior imprensa do século, veneziano de adoção, Aldo Manunzio, o florentino de dois séculos antes, ou seja, uma língua morta, se tornou o modelo para a prosa literária dos italianos: o *Decameron* de Boccaccio foi o modelo para a prosa e o *Canzoniere* do Petrarca o modelo para a poesia. A partir disso, enquanto o florentino falado continuava sua evolução natural, uma variedade diafasicamente e diastraticamente muito limitada de um florentino exclusivamente escrito e parado no tempo ia se difundindo na Itália. As conseqüências disso, em extrema síntese, foram que o italiano foi se impondo unicamente

41 Adaptação nossa da escala de aquisição de Berretta (1986).

42 Há hoje muitas boas sínteses sobre a história linguística italiana. Citamos apenas: Serianni-Trifone 1992-94 (3 voll.); Bruni 1992-2000 (10 voll.); Marazzini (2006); Benincà, 1993; De Mauro, 1963; Castellani, 1982; De Mauro-Vedovelli, 1999; Grassi, 1993; Berruto 1993^a; 2002; ISTAT, 2002; Berruto 1993b; Telmon 1993; D'Achille 1994; Sabatini 1985; Benincà, 1993.

como língua escrita, e portanto excluído da evolução normal devida ao uso comunicativo, e em camadas extremamente minoritárias da população, mas de maneira homogênea, com a exceção de Florença e da Toscana, que não podiam aceitar uma própria variedade morta como modelo. O problema de um italiano para a comunicação social se colocou somente após a unificação política da Itália, em 1861 (De Mauro, 1963; Castellani, 1982). Uma unificação linguística em que os falantes de italiano superam os falantes de dialeto acontece somente nos anos de 1970, com a escolarização de massa e graças ao impacto das tecnologias radiotelevisivas, entre outros fatores (De Mauro-Vedovelli, 1999).

Por *dialeto* na tradição italiana entende-se algo muito diferente daquilo que o mesmo termo indica na tradição anglo-saxônica: não é a variedade de um mesmo código (dialeto de uma região, dialeto caipira, dialeto de uma classe ou de um grupo), mas um outro código, geneticamente irmão (ou seja, filho da mesma língua, em nosso caso o latim), sociolinguisticamente menos prestigioso e territorialmente mais limitado (Grassi, 1993). Portanto, o *status* de dialeto se dá somente quando uma das variedades oriundas do latim, o florentino, adquire o *status* de língua em um determinado território, e as outras, os vários vernáculos ou variedades diferentes do florentino, adquirem uma posição sociolinguisticamente menos prestigiosa e fazem referência cultural àquela língua. Assim, o que se dá na Itália, até tempos recentes, é uma situação de diglossia, com um código, o italiano, usado como código alto (na escola, nas instituições, na escrita) e outro código, o dialeto, usado na comunicação familiar. Hoje o italiano pode ocupar tanto a função alta quanto a familiar, enquanto o dialeto ocupa somente os domínios familiares, gerando assim uma situação definida como *dilalia* (Berruto 1993^a; 2002).

O quadro sociolinguístico que vai se delineando na Itália é o seguinte: por um lado formas de bilingüismo italiano/dialetos, e por outro lado uma estratificação de registros do italiano que criam um complexo diassistema (ISTAT, 2002; Berruto 1993b). Nesse diassistema, o assim chamado italiano padrão (o *italiano standard*) se realiza unicamente na escrita, enquanto na oralidade (com exceção de profissionais da locução) temos principalmente três outras variedades:

O chamado italiano *regional* (Telmon 1993), um italiano mais ou menos influenciado pelo dialeto de baixo estrato, sempre na fonética, frequentemente no léxico, mas também na sintaxe. O que divide o italiano regional (mais ou menos marcado) do dialeto é basicamente a morfologia;

- O chamado italiano *popolar* ou *semiculto* (D'Achille 1994), ou seja, o italiano escrito ou falado em situações de alta formalidade, por pessoas que normalmente usam para a comunicação o dialeto e que não possuem uma aquisição completa do italiano. Trata-se de uma variedade nacional, mas muito permeável a interferências dialetais;
- O chamado italiano *neostandard* ou *de uso médio* (Sabatini 1985), que tem se tornado de fato o novo padrão tendencial em situação de formalidade média, inclusive na escrita.

O que acontece, nesse quadro complexo, com o sistema pronominal? Italiano e dialetos possuem sistemas pronominais diferentes como acontece com diferentes códigos. O italiano regional e o italiano *neostandard* não se diferenciam, quanto ao sistema pronominal, com formas de simplificação do sistema que tendem a ser aceitas fora de registros especialmente altos. Trata-se de simplificações (no sistema pronominal como naquele verbal e em outros pontos) que geralmente seguem, com séculos de atraso, processos que já ocorreram em línguas românicas de mais antigo uso na comunicação do dia-a-dia, como o espanhol, o francês e o português. O italiano popular, ao contrário, apresenta formas de simplificação mais extremas e muito estigmatizadas socialmente. Trata-se de uma variedade muito presente em emigrantes dialetófonos, mas completamente ausente em nossos informantes.

A variedade de referência de nossos informantes é o italiano *neostandard*, principalmente para o tipo de interação dos textos coletados. As simplificações do sistema pronominal no italiano *neostandard* tocam um único pronome entre os objetos do trabalho e em uma única função: trata-se do pronome *ci* com valor locativo que o italiano standard alternava com *vi*, forma completamente desaparecida no registro oral, a não ser em casos de altíssima formalidade. Todas as outras simplificações dizem respeito ao sistema pronominal tônico (perda de alomorfes) da terceira pessoa singular e plural, ou à terceira pessoa, singular e plural, do sistema clítico (o dativo *gli* que vale também para o plural eliminando o pseudo-clítico *loro*, e, em registros coloquiais, para o feminino *le*). Portanto, a reestruturação do sistema não afeta, de fato, as formas utilizadas em nossa pesquisa. Contudo, o movimento de re-estruturação mostra quão delicado e instável é esse setor do sistema, o que o torna um candidato muito interessante a pesquisas sobre erosão linguística (Benincà, 1993).

2.4 Aspectos psicolinguísticos e sociais da erosão linguística

Nesta nossa excursão sobre a erosão linguística, consideramos até agora dois pontos: a falta de insumo da L1, de muito difícil, quase impossível, investigação direta; e os efeitos do insumo de outros códigos, geralmente estudados através da análise do contato da L1 com uma L2. Falta agora uma explanação sobre outros fatores que levam à erosão, de natureza extralinguística, como aqueles sociais, cognitivos, emocionais ou biológicos. Para facilitar sua compreensão serão divididos em fatores de indução interna e fatores de indução externa.

2.4.1 Fatores de indução interna

Para se entender melhor o processo de erosão linguística, é necessário que se levem em conta as mudanças linguísticas que ocorrem. Assim como a aquisição é caracterizada por uma re-estruturação interna do sistema, da mesma forma, na erosão, há forças que são desencadeadas pelo insumo linguístico. Estas mudanças, como foi dito, podem ser induzidas internamente, seja por conta de princípios universais, seja por características peculiares à gramática da L1. Os principais fatores de indução interna são:

Simplificação: o termo é muito controverso, pois a noção de simplificar algo não esclarece o que é considerado complexo e o que é considerado simples. Segundo Köpke e Schmid (2004, p. 16) o vocábulo em si não explica nada e a noção somente tem sentido na medida em que se refere a algum traço linguístico que seja "complexo", como a oposição entre estruturas sintéticas *versus* analíticas, ou estruturas adquiridas tardiamente *versus* aquelas adquiridas cedo, entre outros processos. Apesar da polêmica gerada pelo termo, o conceito de simplificação é ainda amplamente usado pelos estudiosos, significando geralmente o procedimento pelo qual uma estrutura se torna "mais simples" em relação àquela da língua materna.

Generalização: é considerada como uma etapa específica da simplificação porque acarreta a redução da extensão de estruturas possíveis. Um exemplo⁴³ pode ser a tendência de imigrantes africanos a usarem o auxiliar *avoir*, mesmo em verbos que deveriam ser conjugados com *être*, na interlíngua do francês.

As mudanças induzidas internamente manifestam-se mais profundamente no domínio morfológico e morfo-fonológico. Geralmente, em sua elucidação, recorre-se a explicações que remetem à Gramática Universal⁴⁴ ou à consideração sobre traços linguísticos mais ou menos

43 Veronique (1984) citado em Appel R.; Muysken P. (1987, p. 91).

44 Sobre o contributo da GU nos estudos de erosão linguística verá-se logo em seguida.

marcados⁴⁵.

Seliger e Vago (1991) exemplificam bem o que se entende por fatores de indução interna nos vários níveis linguísticos:

Em nível analógico, um traço marcado ou padrões irregulares são eliminados e substituídos por traços não marcados, mais gerais ou regulares.

Em nível paradigmático, as alternâncias regulares de morfemas são reduzidas, e os paradigmas tendem a ser mais regulares.

Em nível categorial, ocorre uma neutralização das distinções categoriais em que o domínio de uma categoria acaba por ser estendido à outra.

Finalmente há o que os autores chamam de *category switch*: uma categoria é mantida conceitualmente, mas expressa em uma forma linguística diferente.

Schmid (2002) identificou quatro modelos linguísticos que têm tido ampla influência sobre a análise e explicação dos dados da erosão e que ocorrem internamente à língua:

A hipótese da regressão, a qual se terá oportunidade de discutir mais delongadamente adiante. A idéia de base é que a erosão é guiada por um "processo que tem suas raízes no todo do sistema cognitivo da memória, e somente de forma indireta no sistema linguístico⁴⁶" (KÖPKE; SCHMID, 2004, p. 16, tradução nossa). Inicialmente, pensava-se que a perda linguística da afasia espelhasse o desenvolvimento linguístico da criança. Porém, na verdade, a afasia tende a não progredir, e não é devida a fatores internos ao cérebro, mas a como este foi lesionado por fatores externos. Os estudos sobre erosão citam todos esta hipótese, mas a única pesquisadora que realmente a testou foi Keijzer (2007), confirmando em parte seus postulados⁴⁷.

O segundo ponto refere-se à noção de simplificação. Como foi visto acima, segundo a autora, o termo não possui nenhum significado, se não quando contextualizado em

45 A noção de marcação foi introduzida na fonologia, e o exemplo clássico, justamente de Jakobson, afirma que /p/ e /b/ distinguem-se pela marca de sonoridade, presente em /b/, que é marcado, e ausente em /p/, que é não marcado. Para ampliar os exemplos, em pares relacionados, aquele marcado é o que assinala de forma explícita a propriedade que o distingue do outro: por exemplo em *oste/ostessa*, *utile/inutile*, *cat/cats*, o segundo exemplo é formalmente marcado justamente por conter uma "marca" (-essa, in, -s) que o assinala como oposto aos outros (ser feminino e não masculino, negativo e não positivo, plural e não singular). De maneira mais geral, a forma não marcada é mais simples do ponto de vista articulatorio e morfológico e mais básica, seja pela sua maior frequência, seja porque se uma língua possui formas marcadas, isso pressupõe que há formas não marcadas.

46 *process that has his roots in the overall, cognitive nature of memory, and only indirectly in the linguistic system.*

47 Adiante será dedicado um parágrafo sobre a hipóteses e as descobertas da pesquisadora.

relação a algo que seja "complexo". A *Simplificação*, assim como a *redução*, são termos que devem ser usados de forma mais apropriada: *redução* no registro devido à *redução* nas funções; *redução* lexical quando afeta itens de baixa frequência; *redução* na morfologia quando o resultado é uma estrutura linguística mais analítica etc.

Interlíngua: como já discutido, na ausência de um insumo de L1, a L2 pode tornar-se fonte de evidência positiva indireta. Relacionado a este fator encontra-se a *transferência*. Contudo segundo alguns autores⁴⁸, deve ser considerada a distância tipológica entre as duas línguas em questão. A transferência acontece com maior facilidade se há similaridade⁴⁹ entre as línguas, como entre o italiano e o espanhol.

Gramática Universal: somente em anos recentes os teóricos do modelo gerativista mostraram interesse pela erosão linguística. O conceito mais utilizado por eles é aquele de "marcação", sendo que os parâmetros marcados seriam reconfigurados na L1 erodida para valores não marcados. Consequência disso seria a necessidade de reforço para evitar a desconfiguração dos parâmetros. Seliger e Vago (1991, p. 7) afirmam que os efeitos no desempenho devem ser extraídos da competência do falante (conhecimento implícito). Isto significaria que a erosão que realmente deveria interessar aos estudiosos seria aquela que atinge o nível da competência por permitir explicações mais contundentes sobre o processo.

Aparentemente cada um dos modelos acima tem um papel, maior ou menor, na erosão e, segundo Schmid (2004), deve ser levado em conta.

2.4.2 Fatores de indução externa

Como foi discutido acima há uma série de fatores internos que podem desencadear a erosão linguística. Contudo, é extremamente difícil conseguir separar as causas internas daquelas externas, pois o que o estudioso pode perceber é a produção linguística do falante e tentar isolar alguns aspectos de outros para poder analisá-los isoladamente. Em seu trabalho, o pesquisador pode achar um pouco mais simples discriminar a influência da L2 sobre a L1 e depois fazer conjecturas sobre os pontos sensíveis que foram atingidos. Contudo, até hoje os estudiosos não conseguiram ter instrumentos metodológicos satisfatórios sobre como distinguir a erosão devida à falta de insumo da L1 e aquela devida ao contato com a L2.

48 Altenberg (1991, 203); Giacalone Ramat (2004).

49 Justamente no sentido de menor distância tipológica.

Veremos adiante algumas propostas que são utilizadas para tentar minimizar este problema, e agora nos concentraremos nos vários fatores externos.

O início da erosão geralmente é evidenciado nas fases em que os falantes alcançam estágios avançados de bilinguismo ou de aquisição de L2⁵⁰, pois os elementos da L1 são moldados em analogia com a L2. Apesar disso, nunca ocorre o caso de um falante que não receba insumo da sua L1 e, após algum tempo, não adquira um estágio avançado da L2. A questão de qual é a causa da erosão, portanto, permanece. Retornando ao influxo da L2, podemos analisar alguns dos mais comuns; segundo Seliger e Vago (1991) são:

A *generalização* de regras, ou seja quando uma regra da L2 é estendida à L1;

Do ponto de vista semântico pode ocorrer a *extensão* de significado, se o significado de uma palavra da L1 é generalizado para incluir uma outra palavra da L1, em analogia com o significado da palavra equivalente da L2. Pode ocorrer também o que é chamado de *decalque*, quando uma expressão, geralmente uma expressão idiomática da L2, é traduzida literalmente para a L1, tornando-se agramatical nesta língua.

Köpke e Schmid (2004), respondem à eterna questão que vem acompanhando os estudos sobre erosão, ou seja, "é possível que alguém possa esquecer uma língua?" afirmando que, apesar de não haver ainda consenso sobre uma definição categórica do que é erosão, as evidências apontam para sua existência. Os autores questionam o quadro taxonômico proposto por van Els (1986)⁵¹ e que por longo tempo guiou os estudos sobre erosão, propondo outros aspectos extralinguísticos sociais mais relevantes, sendo eles:

A idade do falante: é importante diferenciar entre a erosão na idade infantil e aquela dos falantes adultos, claramente relacionando este aspecto ao bilinguismo simultâneo e àquele tardio. Vários estudos sugerem que é necessário levar em conta a idade de início do bilinguismo e aquela de início da erosão, pois o sistema da L1 pode ser erodido de forma bastante intensa se iniciado antes da puberdade⁵². Outrossim, quanto mais jovem a criança for quando a linguagem de seu ambiente é mudada, tanto mais

50 Veja-se Seliger e Vago (1991, p. 4).

51 O quadro, citado em Köpke e Schmid (2004, p. 9) é criticado por conter problemas metodológicos, como a dificuldade em distinguir a *reversion* (fenômeno em que imigrantes idosos, que por grande parte de sua vida utilizaram a L2, em idade avançada perdem esta língua e se tornam novamente mais proficientes na L1) de outros problemas linguísticos que podem ser detectados nos idosos, e pela inclusão da perda do dialeto, que segundo os autores é mais próxima do *shift* linguístico do que de uma erosão verdadeira.

52 Vejam-se os estudos de: Bode (1996); Isurin (2000); Kaufman & Aronoff (1991); Nicoladis e Grabis (2002); Schmitt (2001); Seliger (1989, 1991); Turian & Altenberg (1991); Vago (1991)

velozmente e profundamente ela estará sujeita à erosão⁵³.

O grau de instrução: este aspecto foi por longo tempo deixado de lado, também por uma certa dificuldade em defini-lo quando se lida com situações e culturas diferentes. Contudo, os estudos de Iaspert & Kroon's (1989), conseguiram mostrar que o grau de instrução dos informantes foi o fator extralinguístico mais determinante nos testes que evidenciaram a erosão. O estudo piloto dos pesquisadores, com 30 italianos imigrantes na Holanda, mediu através de vários testes seu nível de erosão linguística, e o fator extralinguístico que mostrou-se mais significativo foi justamente aquele que media seu grau de instrução: quanto menor seu grau de instrução, maior a erosão linguística por eles experienciada, enquanto aqueles que possuíam um nível escolar maior sofriam menos os efeitos da erosão.

Outros estudos⁵⁴ também são citados e todos mostram quanto este aspecto deva ser levado em conta, pois um maior grau de instrução geralmente possibilita uma menor erosão linguística ao falante. É importante lembrar que para as finalidades deste estudo, optou-se considerar este fator com extremo cuidado: uma das características mais importantes dos informantes contatados, como se verá adiante, é aquela de possuir um alto grau de instrução, que lhes permita dominar a língua materna, tendo completado sua aquisição, e ao mesmo tempo ter uma competência metalinguística que possibilite uma certa reflexão e um certo controle sobre a produção linguística. Isto porque a aquisição completa garantiria, em tese, que as evidências de perda linguística sejam devidas à erosão e não possam ser confundidas com outros fatores,

53 A título de exemplo citamos dois estudos que se concentram sobre a erosão linguística na primeira infância. O primeiro é de Turian e Altenberg (1991): o sujeito analisado é uma criança cujo insumo principal até os três anos e meio de idade foi o russo, e que depois foi exposta a um insumo sempre maior de inglês, restringindo o uso do russo a uma hora diária, quando o pai voltava para casa. O estudo, de tipo longitudinal, mostra que a erosão linguística da criança é bastante rápida e ela adota toda uma série de estratégias de compensação para contrabalançar as perdas linguísticas, das quais mostra estar consciente. Enfim as estudiosas observam que as estratégias de compensação usadas pela criança são as mesmas utilizadas por adultos na aquisição de L2.

O outro estudo que se refere à erosão em crianças é de Ventureyra e Pallier (2004): foi analisado um grupo de coreanos nativos que foram adotados na França na primeira infância e não receberam nenhum tipo de insumo da língua materna por muitos anos. Estudos anteriores tinham provado que eles não conseguiam mais reconhecer sentenças nem palavras em coreano. Este estudo se concentra sobre a fonologia e prova que os coreanos nativos não conseguem distinguir sequer a fonologia da própria língua materna, comprovando quão radical é a erosão linguística nas crianças.

54 Importantes a este respeito são os resultados de Yağmur (1997), que conseguiu provar que o desempenho no mesmo tipo de tarefa de fluência verbal teve a influência do nível de educação, seja naqueles sujeitos a erosão, que no grupo de controle, sendo que os sujeitos com um menor grau de instrução obtinham resultados inferiores em ambos os grupos.

como a aquisição incompleta, o que seria bastante comum no caso dos clíticos, que são objeto deste estudo.

Tempo de surgimento da erosão linguística: ainda não há consenso sobre quanto tempo demora para que a erosão se manifeste nos indivíduos. Em geral, os testes mostram que após um período muito breve de permanência no exterior há casos de erosão lexical na produção dos informantes⁵⁵. Vários estudiosos⁵⁶ concordam que, após cerca de dez anos de residência em outro país, os efeitos da erosão tendam a estagnar-se, e que o período de permanência tenha um efeito significativo somente quando o contato com a L1 seja nulo ou muito escasso. Hutz (2004), em um estudo longitudinal sobre 75 cartas escritas por uma emigrante alemã nos Estados Unidos à sua família, em um arco portanto de 57 anos, constatou que a erosão linguística atingiu primeiramente o léxico de forma bastante consistente, enquanto as estruturas morfológicas e sintáticas mantiveram-se suficientemente estáveis. O autor conclui que a erosão ocorre de forma mais marcante, especialmente do ponto de vista lexical, nos primeiros dez a quinze anos. Mas os dados por ele obtidos sugerem que a erosão continua, apesar de mais lentamente, em todos os domínios linguísticos, nas décadas seguintes. Hutz se pergunta também se a variabilidade na quantidade e continuidade, ou menos, da erosão encontrada nas diferentes pesquisas não seja devida a diferenças metodológicas.

Atitude em relação à língua usada: ou seja, fatores emocionais e motivacionais, além de afetivos, podem desencadear ou não o processo de erosão de forma mais ou menos forte.

Quantidade de contato com a L1: é um fator de difícil avaliação, e mesmo de medição, pois depende muito de dados fornecidos pelo próprio informante, mas parece exercer uma influência importante. Consiste na frequência por parte do falante de L1 de uma comunidade que compartilhe sua língua ou a frequência com que ele volta à sua terra natal. É importante ressaltar que o insumo, mesmo na L1, pode contribuir para o aumento da erosão, no caso em que este seja uma variedade já por si erodida. Isto é bastante comum nas comunidades de imigrantes, que acabam por sofrer um processo

55 Köpke e Schmid (2004, p. 10) citam os estudos de Lachman & Mistler-Lachman (1976) e Magiste (1979), que colocam o início da erosão por volta dos sete anos após a imersão em um ambiente linguístico diferente, enquanto Frenke-Mestre (1993) conseguiu resultados similares após somente três anos.

56 Vejam-se os trabalhos de Bouba *et alii* (2002); Bot de & Clyne (1994): especialmente no segundo, em que os informantes foram testados em 1970, e depois re-testados em 1987, é claramente afirmado que as habilidades linguísticas podem deteriorar-se na primeira década, mas após este período tendem a estabilizar-se.

de erosão em massa e, mesmo quando falam entre si em sua L1, utilizam uma variante não padrão.

Os últimos dois fatores são os mais problemáticos para o pesquisador, pois envolvem avaliações subjetivas e de difícil categorização, portanto nem sempre são levados em conta nas pesquisas.

2.4.3 Aspectos psicolinguísticos

Um dos aspectos que parecem ser mais relevantes do ponto de vista psicolinguístico é que se pode relacionar o início da erosão da L1 com estágios avançados de bilinguismo ou de aquisição de L2⁵⁷. No primeiro estágio do aprendizado da L2, o falante utiliza seus conhecimentos da L1 para formular hipóteses sobre a L2. É o que é conhecido como *transferência* linguística, e é caracterizada por um *continuum* de interlínguas. Em um segundo estágio, aquele da reestruturação, o falante separa as duas gramáticas, a da L1 e a da L2, e em seguida "recria"⁵⁸ regras para a língua-alvo que não possuem mais nenhuma referência à sua L1. O estágio final é parecido com o primeiro, mas de forma inversa: agora que adquiriu proficiência na L2 o falante utiliza estes conhecimentos para recuperar a L1 em que ele vai perdendo o domínio. É neste estágio que pode entrar em ação a erosão linguística da L1.

É importante ressaltar, no entanto, que nem todos os elementos da L2 na L1 são indicativos de um processo de erosão em curso. Podem ser simplesmente casos de *code mixing* ou *code switching*⁵⁹, que são estratégias constantemente usadas por falantes bilíngues em contextos que permitem o uso de um repertório multilíngue.

O ensaio de Köpke e Schmid (2004) aprofunda alguns aspectos psicolinguísticos importantes que devem ser considerados no estudo da erosão. O primeiro, que já foi citado como fator de indução externa, mas aqui é analisado sob uma outra ótica, é o fator idade. Em geral, há consenso entre vários estudiosos que a erosão nas crianças atinge níveis muito mais sérios do que nos adultos. A Hipótese do Período Crítico⁶⁰ postula que, devido a limitações da maturação cerebral, após uma certa idade, o aprendizado de uma L2 é mais difícil, pois são

57 Seliger & Vago 1991, reimpressão 1999, p. 4.

58 *Recreation* é o termo usado por Seliger & Vago, 1991.

59 Seliger & Vago 1991, reimpressão 1999, p. 6 ou mais extensamente o capítulo 10 de Appel e Muysken (1987); ou ainda Myers-Scotton (1993).

60 Lenneberg (1971): segundo o estudioso existiria um período crítico no aprendizado das línguas, seja a L1 que a L2, geralmente posto antes da puberdade. Ele afirma que a aquisição de L1 ou L2 seria mais fácil, devido a uma maior plasticidade neuronal.

necessários alguns anos para que a L1 estabeleça-se plenamente no cérebro do falante. Contudo, não há consenso sobre tal, e Harley e Wang (1997) preferem um conceito que é considerado mais apropriado por muitos, o "período sensível", que começa a declinar após os seis ou sete anos de idade. Isto implicaria que quanto mais facilmente uma criança aprende uma L2, com maior facilidade esquecerá sua L1. A idade em que isso ocorre porém, ainda não é claramente definida e talvez varie para cada nível descritivo da língua⁶¹.

Outro fator que deve ser levado em conta, e que já foi visto como determinante do ponto de vista externo, mas que aqui é visto sob outro ponto de vista, é aquele da educação. O que é sugerido é que não somente o grau de instrução tem um papel importante na resistência à erosão, mas também deve ser combinado ao fator idade, pois, justamente na fase em que parece haver uma forte propensão à erosão, na infância, é quando as crianças adquirem certas capacidades de leitura e escrita em pelo menos uma língua, e isto pode ajudar a "fixar" a L1 no cérebro, reforçando um insumo que seria somente oral.

Do ponto de vista da Gramática Gerativa, os estudos de Sharwood Smith (1983a e 1983b) são importantes por serem os primeiros a levarem em conta a diferença entre competência e desempenho na análise dos dados sobre erosão. O autor afirma que o que é afetado pela erosão é de fato o desempenho. Isto ocorre de duas formas: nas dificuldades de acesso ao léxico por parte do falante, e nas dificuldades de processamento: quando mudam os parâmetros entre as duas línguas, os sujeitos experienciam maior dificuldade em evitar o uso da L2 quando usam a L1. Sharwood Smith (1983a) afirma que na distinção por ele feita entre competência e desempenho, ocorreriam três estágios diferentes no processo de erosão:

O primeiro é caracterizado por desvios sistemáticos no desempenho, enquanto a competência permaneceria estável;

No segundo, de transição, o bilíngue possuiria uma nova variedade de contato, mas seria capaz de retornar àquela original;

O terceiro estágio seria caracterizado por um novo tipo de competência, reduzida e com um repertório mudado segundo as regras da gramática universal.

Finalmente um terceiro fator de ordem psicolinguística refere-se às noções de ativação

61 Estudos posteriores àqueles de Lenneberg propuseram a existência de vários períodos "sensíveis", para cada nível linguístico, como a fonética, a morfologia, a sintaxe e assim por diante. Estes estudos demonstraram que às vezes adultos e adolescentes são mais rápidos do que as crianças nas fases iniciais da aquisição, porque podem se valer de uma maior maturidade cognitiva, e portanto operam reflexões metalinguísticas e generalizações sobre o novo sistema linguístico de forma muito mais eficiente do que as crianças. A este respeito veja-se: Jacobs B. (1988); Singleton (1989).

e inibição, que estão ligadas à frequência de uso e reforço de um específico sistema ou estrutura linguística. Os itens que apresentam baixa frequência estariam mais sujeitos à erosão, segundo o princípio psicolinguístico da ativação e inibição que parece controlar o uso de mais de uma língua pelo cérebro.

Os fatores sociais, como pode ser visto, têm profundas consequências sobre o processo psicolinguístico que leva à erosão. Lembrando-os mais uma vez são eles: grau de instrução, idade da partida da terra natal e quantidade de contato com a L1.

2.4.4 Sequência de erosão linguística de L1

Os estudos sobre erosão linguística podem ser divididos em dois grandes grupos: aqueles de tipo longitudinal, em que um sujeito ou vários são monitorados por um arco mais ou menos longo de tempo, e aqueles sincrônicos, em que se escolhe um grupo de informantes que estão em situação de erosão há diferentes períodos. Ambos os tipos de estudo, como observa Keijzer (2007), almejam descobrir padrões de desenvolvimento da perda linguística, buscando saber o que é perdido, e em que momento. Assim como no processo de aquisição, chegou-se a uma hierarquia de perda, que na realidade espelha a aquisição. Keijzer (2007) propõe uma ordem hierárquica de perda, que ela mesma considera ainda sujeita a possíveis modificações em vista de estudos futuros, mas que é significativa das descobertas até agora feitas e de como a Hipótese da Regressão tenha ainda muito a dizer:

(5) fonética < fonologia < sintaxe < morfologia < léxico

Segundo a autora, a hierarquia deve ser interpretada na direção oposta à escala de aquisição: o léxico é o mais suscetível à perda, "talvez porque a aquisição do vocabulário seja um processo que dura a vida inteira"⁶² (KEIJZER, 2007, p. 14, tradução nossa), e, portanto, as influências da L2 sobre o léxico sejam mais visíveis, sob forma de *code switching*. Morfologia e sintaxe parecem ser mais resistentes ao processo de erosão⁶³, e são perdidos de forma mais lenta. A explicação dada por Keijzer (2007) é que talvez estes dois últimos sistemas estejam muito mais integrados na estrutura da linguagem do que o léxico. Finalmente, a erosão da fonética e da fonologia ainda são pouco estudadas mas parece que

62 *vocabulary acquisition is a lifelong process.*

63 Observa-se que estes resultados são muito parecidos com aqueles descritos em Hutz (2004).

sistemas fonéticos estrangeiros possam com o passar do tempo interferir na L1⁶⁴.

Segundo Keijzer (2007), a hierarquia acima apresentada pode sugerir, para cada âmbito linguístico, uma sub-hierarquia, como a seguinte:

- (6) formas fracas do passado < formas fortes do passado

Isto significa que as formas fortes do passado são menos expostas à erosão do que as formas mais fracas. Em italiano, por exemplo, podemos citar os passados remotos regulares como formas fortes e os irregulares como formas fracas.

2.5 A Hipótese da Regressão de Jakobson

Em sua monografia *Linguagem infantil, Afasia e Universais fonológicos* de 1941, Roman Jakobson afirma que a única oportunidade de verificar a formação da linguagem humana é através da observação da linguagem infantil, enquanto a oportunidade para observar sua dissolução é oferecida pelas falhas patológicas, de natureza cerebral, na fala. No ensaio, o autor opera um paralelo entre três processos: a mudança linguística diacrônica, a aquisição da primeira língua e a perda da linguagem devido a danos cerebrais. A grande intuição de Jakobson foi que o estudo destes três processos levasse à conclusão que existem mecanismos universais que regulam a linguagem. O seu trabalho concentrava-se sobre a fonologia e, efetivamente, ele conseguiu mostrar que os fonemas devem ser vistos como um conjunto, e em conjunto eles desenvolvem seu papel em cada uma das mudanças acima mencionadas. Jakobson percebeu que haveria relações de dependência nas hierarquias de aquisição do sistema fonológico, e que estas seriam unidirecionais⁶⁵, ou seja, o traço *x* não pode ser adquirido depois do traço *y*, assim como o traço *x* não será deteriorado antes que o traço *y* sofra um processo de erosão. Estas hierarquias sequenciais levam em conta as variantes que podem acontecer, e mesmo assim conseguiriam predizer paralelos testáveis entre aquisição e erosão⁶⁶.

Outra intuição importante de Jakobson na exposição da Hipótese da Regressão refere-se ao conceito de marcação. Nas línguas, há sempre uma oposição entre traços marcados e

64 É citado a este respeito o estudo de Major (1992). Lembramos, no contexto da erosão do italiano o estudo de Celata e Cancila (2008), que comprovou erosão fonológica no contato com o inglês.

65 Estas hierarquias são chamadas por Jakobson de *irreversible laws of solidarity*. Veja-se a este respeito Jakobson (1968, p. 64).

66 Veja-se a este respeito também as recapitulações das teorias de Jakobson sobre a Hipótese da Regressão em Bot de & Weltens (2004, p. 31 e segs.) e Keijzer (2007, p. 2-3).

não marcados. Os primeiros a serem adquiridos são os traços não marcados: prevê-se portanto, segundo esta lógica, que os últimos itens a serem adquiridos serão os primeiros a serem perdidos.

Segundo Keijzer (2007), o problema com o conceito da marcação *supra* citado é que ele pode funcionar para a fonologia, mas nem todos os sistemas gramaticais baseiam-se em oposições de pares, como por exemplo a ordem das palavras. Segundo a autora, a Hipótese da Regressão, assim como formulada por Jakobson, não foi devidamente comprovada por meio de dados.

Os estudos sobre regressão de Jakobson, e muitos outros posteriores a ele⁶⁷, inicialmente concentraram-se sobre o paralelo entre a aquisição linguística em crianças e a perda linguística patológica, a afasia. Entretanto, neste caso a hipótese não se mostrava convincente, e os resultados obtidos foram decepcionantes. Isto porque, na realidade, na afasia, os danos cerebrais atingem alguns aspectos específicos do sistema linguístico e não o sistema como um todo. Além disso, geralmente os danos cerebrais produzem um efeito imediato, e não gradual; portanto não há como fazer uma comparação inversa com o processo de aquisição linguística. Finalmente, como nota Keijzer (2007), os afásicos geralmente adotam estratégias de compensação em que provavelmente pode haver uma reestruturação do substrato neuronal e, portanto, qualquer comparação com a aquisição linguística não seria apropriada. Mesmo outras síndromes em que haja perda da linguagem, como o mal de Alzheimer, o de Parkinson ou a doença de Huntington, apresentariam os mesmos problemas metodológicos se comparadas com a aquisição infantil.

Inicialmente, estudiosos como De Bot & Weltens (1991) perceberam, em suas investigações, similaridades entre a linguagem de idosos e crianças, adaptando em seguida esta teoria à erosão linguística. Em sua visão, a Hipótese da Regressão não deveria ser aplicada à ordem de aquisição e perda, mas somente ao padrão da perda: habilidades utilizadas de forma mais inconsciente, portanto mais arraigadas na linguagem, seriam menos passíveis de erosão.

Um dos problemas de se estudar a regressão é que a investigação é possível somente com fenômenos linguísticos que são adquiridos com uma ordem mais ou menos fixa, isto para que seja possível detectar o "último adquirido, primeiro perdido".

O paralelo entre aquisição e erosão foi tentado também na comparação entre

67 Assinalamos a este respeito: Caramazza e Zyrif (1978); Grodzinsky (1990); Avrutin, Haverkort e van Hout (2001); Kolk (2001); Bastiaanse e Bol (2001).

linguagem infantil e erosão linguística não patológica, seja na L1 que na L2. Keijzer (2007) cita vários estudos⁶⁸ que tentaram comprovar a Hipótese da Regressão, mas observa que os resultados nem sempre são consistentes. Concordando com Schmid (2004), a autora identifica o problema nas deficiências metodológicas na coleta de dados, em que não há consenso entre os estudiosos sobre como deve ser feita.

A maior crítica feita por Keijzer (2007) aos estudos anteriores ao dela é de ter comparado a erosão ao desenvolvimento linguístico de crianças muito pequenas. Ela afirma ser, na prática, pouco provável que a erosão chegue a um estágio do sistema linguístico similar ao de crianças de quatro ou cinco anos. E, de fato, os estudos precedentes encontraram evidências muito sutis de erosão. Finalmente, a pesquisadora afirma que o que não aparece nestes estudos é uma explicação das sequências de erosão detectadas, ou seja, falta justamente uma investigação sobre o processo de erosão e os mecanismos que o causam. São estas as questões que Keijzer (2007) se colocou ao iniciar seu trabalho, que tinha como objetivo justamente testar, da forma mais minuciosa possível, a Hipótese da Regressão de Jakobson.

2.5.1 O teste de Keijzer sobre a Hipótese da Regressão

Como foi visto acima, apesar de sua fama e de ser constantemente citada nos trabalhos sobre erosão linguística, a Hipótese da Regressão de Jakobson nunca fora realmente testada até recentemente. Em sua tese de doutorado, Merel Keijzer (2007) observou, através de uma série de testes específicos, de que forma a ordem da perda linguística de um grupo de imigrantes holandeses no Canadá anglófono seria especular ou não àquela de aquisição linguística da L1 de um grupo de jovens holandeses em sua terra natal.

Uma das questões que mais instigaram a estudiosa, como foi discutido no parágrafo anterior, é que testar a Hipótese da Regressão na aquisição linguística de crianças durante a primeira infância engendraria uma série de problemas não tendo sido completado o processo de aquisição. Keijzer portanto, neste estudo, operou de forma radicalmente diferente, e examinou o paralelo entre a erosão de L1 e o estágio avançado de aquisição de jovens holandeses, que, segundo a estudiosa, se coloca na idade entre treze e catorze anos. A justificativa para tal escolha é que "(eles) teriam alcançado um domínio quase completo de seu sistema linguístico nativo, mas sua gramática poderia mesmo assim continuar a evidenciar alternativas que gramáticas totalmente adquiridas não exibem. Isto pode corresponder às

68 Citamos alguns entre aqueles sugeridos por Keijzer os de mais fácil obtenção: Andersen (1982); Jordens, de Bot, van Os e Schumans (1986); Schmid (2002), Hansen (1999).

mudanças sutis encontradas nos sistemas linguísticos sujeitos a erosão"⁶⁹ (KEIJZER, 2007, p. 6, tradução nossa).

O estudo testou 15 traços morfológicos e morfossintáticos, analisados e controlados através de tarefas linguísticas específicas e narrativas espontâneas⁷⁰. A escolha pela área sintático-morfológica foi feita por ela mostrar um processo gradual de aquisição, diferentemente do léxico, mais difícil de se testar.

O trabalho da pesquisadora dedica uma seção a cada item analisado, explicando como este é adquirido na L1 e na L2 de contato, portanto no holandês e no inglês. São feitas previsões de como poderia ocorrer a erosão segundo várias teorias e, em seguida, é feita a análise dos testes aplicados e dos dados coletados, nos grupos pesquisados: os holandeses imigrantes no Canadá anglófono, os adolescentes holandeses em pátria e o grupo de controle de holandeses adultos em pátria.

O estudo testou não somente os vários aspectos linguísticos: morfológico, sintático, da posição dos verbos; mas também buscou inserir a Hipótese no contexto de várias teorias, como a Gramática Gerativa, a *Usage-based Theory* (ou Teoria Construtivista), o desenvolvimento linguístico *Construction-based* e a *Dynamic System Theory*. Keijzer (2007) contextualizou a Hipótese Regressiva em cada teoria, mostrando até que ponto esta lhe se adaptaria, quais os prós e os contras.

Por exemplo, os modelos da Gramática Universal podem explicar as simetrias entre aquisição e erosão, porque ambas são limitadas pela GU: as teorias gerativistas podem reforçar a Hipótese da Regressão no sentido de que tanto a aquisição quanto a erosão são guiadas pela Gramática Universal. Além disso, parece que os parâmetros de reconfiguração (\pm

69 "have reached near-complete mastery of their native language systems, but their grammars may nonetheless continue to show optionality where mature grammars do not. This might correspond to the subtle changes found in attrited language systems".

70 A estudiosa questiona os métodos geralmente utilizados na coleta de dados: quando há a narração livre, o informante pode se valer de estratégias para evitar a utilização de construtos ou termos que ele não lembra mais, portanto há uma certa dificuldade em se encontrar determinadas estruturas. Apesar disto, é um tipo de teste em que o pesquisador pode ter acesso a uma produção de tipo natural. Os testes específicos de linguagem, como aquele de tradução ou de julgamento sobre gramaticalidade, por outro lado, permitem obter os dados necessários com uma certa facilidade, mas são pouco naturalísticos. Keijzer decidiu portanto utilizar um conjunto das várias tipologias, de forma a suprir a escassa naturalidade de um, e a possível ausência de dados relevantes do outro. Além destes testes, foi aplicado um questionário sociolinguístico para cada informante, foi utilizada a tarefa de recontar um trecho de filme ("Tempos Modernos", de Charlie Chaplin), foi aplicado o *C-test*, que consiste em preencher lacunas em textos originais, e o *Wug Test*, geralmente empregado com crianças pequenas, que utiliza palavras sem sentido para examinar qual seu grau de análise morfológica, e, enfim, o *Can-do Scales*, em que os informantes indicam qual seu grau de conhecimento linguístico. Para maiores detalhes sobre a metodologia utilizada remetemos ao capítulo 6 (pp. 140-181) de Keijzer (2007).

traços interpretáveis) podem ter algum impacto⁷¹. Contudo, o problema destas teorias é que não conseguem propor hipóteses que possam ser provadas na prática sobre o processo de aquisição e perda, e, além disso, não têm resultados satisfatórios quando lidam com efeitos parciais, que costumam caracterizar a aquisição e a erosão quando há influência de variáveis externas.

As Teorias Construtivistas, por outro lado, conseguem capturar o paralelismo entre aquisição e erosão, porque segundo estas teorias ambos não afetam, se não em estágios muito avançados, o inteiro sistema linguístico. Estas teorias enfatizam o papel do insumo da língua no processo de aquisição e levam em conta aspectos cognitivos e habilidades socioculturais. A aquisição aconteceria através de capacidades cognitivas que permitem a categorização: a criança forma categorias na base de objetos e acontecimentos que mostram similaridades entre si. Desta forma também a língua é adquirida, através de uma categorização do uso simbólico da linguagem, que é tipicamente humano. Na erosão linguística aconteceria o mesmo, mas de forma inversa: o quadro geral adquirido permaneceria intacto, mas seriam perdidos elementos específicos, assim como na aquisição a criança possui um quadro geral da língua mas adquire aos poucos os elementos através do insumo.

Outra teoria que segundo Keijzer (2007) mostra-se bastante convincente é a DST (*Dynamic Systems Theory*), que tem como base as explicações funcionalistas: o sistema linguístico é visto como um todo e não como a somatória de suas partes, "assim como o sistema bilíngue presente na erosão não é a soma de dois inventários monolíngues"⁷² (KEIJZER, 2007, p. 37, tradução nossa).

No estudo por ela apresentado, a pesquisadora se vale do conjunto de todas estas teorias, e outras mais⁷³, para buscar formar um quadro o mais completo possível do processo de aquisição e daquele de erosão.

As conclusões da pesquisa apontam para uma certa specularidade entre aquisição e regressão. Contudo, a regressão foi mais evidente no campo da morfologia, enquanto a sintaxe caracterizou-se por uma prevalência do influxo da L2. As evidências encontradas portanto, indicam quão sutil seja o fenômeno da regressão na explicação da erosão linguística. A hipótese jakobsoniana não pode ser o único fator de previsibilidade: Keijzer (2007) conclui que o sistema pode ser reestruturado seguindo como direção de base a regressão, mas a

71 Um estudo baseado em teorias gerativistas sobre a erosão é por exemplo: Tsimpli *et al.* (2004, pp. 257-277).

72 *just as the bilingual system present in attrition is not the sum of two monolingual inventories.*

73 Para um panorama mais detalhado sobre as teorias subjacentes e suas explicações, remetemos diretamente a Keijzer (2007, pp. 17-37).

influência de uma L2 específica, o grau de instrução e outros fatores externos podem ter um peso até maior.

2.6 O debate metodológico nas pesquisas sobre erosão linguística

Uma vez identificado o objeto de estudo, a erosão linguística, é importante uma discussão sobre os métodos de coleta e análise de dados, pois somente através de critérios pré-estabelecidos será possível obter resultados coerentes que comprovem ou menos as hipóteses a serem testadas, ou avaliar os aspectos metodológicos que é preciso implementar no futuro.

Como foi visto nos parágrafos anteriores, o estudo da erosão é um objeto relativamente recente e que ainda não consolidou definições precisas. Mesmo nas metodologias de estudo é importante ressaltar que ainda há divergências nas práticas. As críticas de Keijzer vistas acima referem-se em grande parte à metodologia e confirmam as incertezas que os pesquisadores ainda têm no momento de levar adiante suas pesquisas.

Inicialmente, os métodos utilizados dependiam do tipo de pesquisa a ser feito, isto é, variavam conforme o que o pesquisador estava procurando. A partir dos anos noventa, quando ocorreu a separação entre os estudos sobre erosão de L1 e de L2, houve também uma separação mais consistente sobre a metodologia. Enquanto as pesquisas sobre a L2 utilizavam os testes típicos do estudo da aquisição⁷⁴, aquelas sobre a erosão da L1 valiam-se da visão mais geral que a sociolinguística empregava nas investigações sobre multilinguismo e línguas minoritárias.

Pode-se observar em geral três técnicas de obtenção de dados nas pesquisas sobre erosão da L1:

A auto-avaliação, em que o informante é convidado, através de uma escala de valores, a indicar seu nível de proficiência. O problema é que este método é muito subjetivo e pode ser influenciado por inúmeros fatores não controláveis, portanto nos últimos anos este tipo de estudo foi completamente abandonado.

Técnicas formais de obtenção de dados: são utilizados vários tipos de testes, como testes de julgamento de categorias morfológicas ou semânticas, julgamento de gramaticalidade, produção de sentenças ou dados específicos, associações controladas, geração de determinadas estruturas sintáticas, entre outros. O único tipo de teste

⁷⁴ Alguns destes testes foram aqueles vistos acima e utilizados na pesquisa de Keijzer: vão desde a narração de um trecho de filme mudo (muito usado é o filme "Tempos Modernos" de Charlie Chaplin), até testes em que o informante deve indicar quão gramatical ou agramatical ele considera determinadas sentenças, ou exercícios em que devem ser preenchidas lacunas de textos originais e assim por diante.

controverso entre os vários aplicados e que é severamente criticado por vários autores⁷⁵ é aquele da tradução, que, por sinal, é cada vez menos usado. As vantagens destas técnicas são o fato de serem controláveis e reproduzíveis, e a facilidade de obtenção dos dados que o pesquisador necessita. Contudo, falta neles a naturalidade, porque são completamente guiados e não deixam margem ao informante de se expressar da forma que ele faria nas situações do cotidiano.

Técnicas que estimulam os informantes a produzirem dados que sejam os mais próximos possíveis da fala espontânea. Geralmente consistem de entrevistas e narrações de acontecimentos. Se por um lado este método permite medir todos os aspectos do desempenho, ou fazer uma pesquisa preliminar sobre as áreas sensíveis à erosão, por outro, como já foi visto, recai-se na questão do evitamento de determinadas estruturas e de como isto poderia ser quantificado⁷⁶.

Yağmur (2004) discute sobre como encontrar a metodologia mais adequada para cada pesquisa, fornecendo prós e contra de cada método e lembrando as perguntas de base que o pesquisador deve se fazer antes de iniciar o trabalho. O estudioso afirma que métodos divergentes de coleta de dados, diferentes classificações e instrumentos diversos são ainda o maior problema na área da pesquisa sobre erosão. Muitas vezes há ainda confusão sobre a finalidade de determinado teste, ou até mesmo alguns testes que deveriam medir a erosão de L1, na realidade foram projetados para medir a erosão de L2, e os resultados portanto não seriam confiáveis. Os estudos mais precisos são aqueles longitudinais, em que os informantes são testados antes que a erosão se manifeste, e ao longo dos anos vê-se o desenvolvimento do processo. Contudo, nem sempre é possível implementar tal tipo de pesquisa por problemas de logística, financeiros e de realizabilidade. Geralmente, para resolver este problema, o método usado é a implementação de um grupo de controle que vive no país de origem dos informantes estudados. Este procedimento entretanto pode ser utilizado no teste de informantes de primeira geração, mas não naqueles de segunda ou terceira geração.

Outra crítica feita por Yağmur (2004) refere-se à representatividade da comunidade ou do grupo testado em relação àquela língua, fato que nem sempre é levado em conta, ou pelo menos esclarecido, nos estudos. Em relação a isto, o autor ressalta o peso do fator idade, pois, para medir a erosão, é necessário que a aquisição tenha sido completada: nem sempre entre os mais jovens todo o processo foi concluído, e, nos mais idosos, pode haver perdas linguísticas

75 Veja-se por exemplo a crítica de Köpke e Schmid (2004, p. 27).

76 Köpke e Schmid, *ibidem*.

devido a fatores neuronais. Seria portanto importante que isso fosse levado em conta no planejamento da pesquisa.

O estudioso também alerta sobre os riscos dos testes aplicados para provar a erosão, pois algumas tarefas são muito questionáveis: em específico há fortes críticas às atividades que implicam a tradução. Finalmente Yağmur afirma que "o relato de 'grandes perdas' é muitas vezes causado por considerações insuficientes de tipo metodológico"⁷⁷ (YAĞMUR, 2004, p. 143, tradução nossa). Isto pode gerar resultados exagerados ou distorcidos. Os pesquisadores devem ter em mente um foco claro de pesquisa através de questões e hipóteses convincentes.

É o que Kopke e Schmid (2004) denominam de a "próxima fase" nas investigações sobre erosão linguística. Esta etapa consistiria em um maior cuidado por parte dos pesquisadores no rigor científico, levando em conta algumas questões importantes que foram acima citadas e que aqui serão brevemente resumidas:

- Esclarecer se os estudos da erosão devem limitar-se à análise da interferência ou se não seria necessário também verificar os dados do desempenho, como complexidade sintática ou o tipo de frequência dos tipos linguísticos pesquisados;
- Perguntar-se como a coleta de dados influencia os resultados e sobre a necessidade de um grupo de controle. Neste caso, estabelecer como este seria escolhido;
- Questionar sobre quais fatores extralinguísticos devam ser levados em conta e como controlar as variações dialetais e sociais, assim como o contexto, a história dos informantes e suas atitudes e motivações.

Na pesquisa aqui apresentada, dentro das limitações de tempo e recursos tidos, foi feita uma tentativa de seguir estas indicações e responder às questões da forma mais acurada possível.

2.7 Estudos sobre a erosão linguística do italiano

Os estudos sobre erosão linguística se difundiram especialmente no ambiente anglófono, onde nasceram e primeiramente foram testados; com isso entende-se que ou o inglês é a L1 sujeita a erosão, ou é a língua de contato L2 para outra L1 submetida a erosão. Na Europa, o maior centro de difusão continua sendo a Holanda, país tradicionalmente

⁷⁷ reports of "massive loss" are often caused by insufficient consideration of methodological issues.

multilíngue, e é relativamente fácil nos últimos anos, graças também aos fenômenos migratórios, encontrar trabalhos sobre a erosão não mais somente do inglês mas também, do holandês, do francês, do hebraico e até de línguas orientais como o coreano ou o japonês⁷⁸.

O estudo da erosão do italiano é um campo mais circunscrito. Poucos pesquisadores se dedicaram ao assunto, e isto aconteceu somente em tempos mais recentes. Razões históricas e sociopolíticas determinaram tal fato: a grande emigração em direção a outros países europeus ou às Américas e Austrália aconteceu no final do século XIX e foi bastante consistente até a segunda guerra mundial. Entretanto, tratava-se de italianos que possuíam o dialeto como primeira língua, portanto os estudos concentraram-se inicialmente sobre aspectos sociolinguísticos sobre as comunidades de migrantes e muitos sobre os imigrantes de segunda e terceira geração. Isto é muito diferente dos objetivos que nos propomos, que é estudar italianos nativos que possuam alto grau de instrução. A migração deste tipo de informante aconteceu somente após a segunda guerra mundial e ainda há pouquíssimos estudos sobre este assunto.

Apesar do grande interesse e dos avanços ocorridos nos vários campos da linguística a partir dos anos setenta do século passado, uma das maiores pesquisadoras italianas sobre o assunto, Camilla Bettoni (1990), ressalta que os estudos sobre a erosão do italiano L1 são muito heterogêneos. A estudiosa observa que são poucos os projetos de estudo da língua de filhos de imigrantes na Alemanha, na Suíça e na Holanda⁷⁹, e em raros casos houve uma sistematização dos dados: muitas observações são esporádicas e falta o rigor interpretativo de uma teoria de base.

Os trabalhos de Bettoni⁸⁰ são mais rigorosos em sua metodologia. A pesquisadora dedicou longos anos ao estudo da língua falada pela comunidade italiana na Austrália. As pesquisas sobre o que é chamado de "austritaliano"⁸¹ conjugam a análise dos comportamentos que levam à erosão, continuando através das gerações sucessivas de falantes com as modalidades do *shifting* em direção ao inglês.

Deve ser notado, tanto nos trabalhos de Bettoni, quanto nos próximos que serão citados, a forte presença dos dialetos italianos na língua dos migrantes, fato que já foi

78 Para não nos delongarmos com exemplos aqui desnecessários, remetemos às bibliografias de referência: Schmid M. (2004) fornece um precioso trabalho com uma vasta bibliografia sobre erosão linguística, anotada por tipo de estudo, línguas erodidas, línguas de contato, entre outros dados interessantes; além desta citamos o site <http://www.lara.ox.ac.uk/firstlang.php> que contém uma, não completa mas bastante vária, bibliografia sobre erosão até o ano de 2008.

79 São citados os trabalhos coordenados por Berruto (1991; 1998) na Suíça e Ludi (1986), sempre na Suíça.

80 Vejam-se por exemplo os volumes Bettoni; Rubino (1996) e Bettoni; Rubino (1995) entre outros.

81 Veja-se Bettoni (1991).

explicado anteriormente. Muitas das pesquisas sobre o italiano no exterior possuem esta dicotomia: apresentam-se como estudos sobre erosão, mas acabam por abordar situações em que o peso do dialeto é muito forte. No caso de Bettoni, no próprio título de alguns trabalhos é presente este fator e a estudiosa fala de "trilinguismo", pois estão em jogo o italiano, os dialetos vêneta e siciliano e o inglês australiano⁸². É importante lembrar que vários trabalhos referem-se à variedade do italiano submetido à erosão dos migrantes de segunda ou terceira geração o que, a rigor, caracterizaria uma situação de *shift* e não propriamente de erosão linguística⁸³.

Além disso, o que pesa no desequilíbrio dos estudos sobre a erosão do italiano tem a ver com a distribuição geográfica das realidades estudadas. Apesar da forte migração italiana em direção à América Latina, especialmente à Argentina, ao Brasil e à Venezuela, poucos são os estudos feitos até hoje nesta área, fora aqueles citados. Um pouco mais rica e mais recente é a bibliografia sobre os italianos e sua língua nos Estados Unidos. Os primeiros trabalhos concentraram-se sobre a costa leste (New York e Chicago). O interesse de tais estudos, entretanto, era mais direcionado aos fenômenos de hibridação recíproca, geralmente fonética e lexical⁸⁴.

Um dos estudos mais recentes sobre esta área é o de Scaglione (2000). Trata-se de uma análise sobre a erosão fonológica e morfológica de um grupo de migrantes da província de Lucca na cidade de São Francisco nos Estados Unidos. É feita uma investigação sobre a primeira e a segunda geração de falantes, portanto passa-se da questão da erosão para aquela do *shift*, e alguns fenômenos de erosão que ocorrem nos falantes, como a mudança de gênero, são atribuídos não ao contato com o inglês, mas ao substrato dialetal.

Outras estudosas que se interessaram pelo mesmo fenômeno, a erosão linguística de imigrantes da cidade de Lucca na cidade de San Francisco, nos Estados Unidos, foram Celata e Cancila (2008). Sua pesquisa investigou a erosão do traço de alongamento consonântico na comunidade de ítalo-descendentes. Para verificar as diferenças entre os imigrantes de primeira

82 Vejam-se por exemplo os volumes Bettoni C; Rubino A. (1996) e Bettoni C.; Rubino A. (1995).

83 Referimo-nos aqui, por exemplo, ao *cocoliche* rio platense, que é a língua mista dos italianos do Rio de La Plata, uma mistura de italiano com interferências espanholas e um espanhol com interferências italianas; vejam-se a este respeito os trabalhos de Meo Zilio (1959; 1960). Outro caso que é significativo do forte influxo dos dialetos que os imigrantes trouxeram consigo é o *talian*, uma *koiné* de base vêneta falada na região da Serra Gaúcha, no estado do Rio Grande do Sul e no oeste de Santa Catarina; veja-se Luzzato (1994) e Corrà (1998).

84 Veja-se Scaglione (2000) que na introdução aborda esta questão e fornece uma bibliografia sobre alguns dos principais estudos. Vejam-se também alguns dos trabalhos de Haller, como Haller (1993; 1998; 1999; 2000, 2003; 2006).

e segunda geração, utilizou-se um grupo de controle na cidade de Lucca, na Itália. Foi dada muita atenção à questão metodológica, utilizando um teste que discriminaria as palavras, e outro baseado em não-palavras. Os resultados confirmam a erosão do traço em questão. A explicação aponta tanto para variáveis individuais como também para motivos extralinguísticos, como o nível de escolaridade, e ainda para uma hierarquia de perda, que parece seguir os mesmos critérios que correspondem ao comparecimento das habilidades na aquisição de uma L2.

Em Tsimpli *et al.* (2004) é relatada uma pesquisa sobre falantes L1 de grego e italiano, ambas línguas PRO-DROP, que utilizam sua L1 diariamente e possuem uma proficiência quase nativa do inglês L2. O estudo se valeu de dois testes de identificação do sujeito, expresso e não expresso, em enunciados encaixados em que, em alguns casos, o pronome era coreferencial ao sujeito, e em outros não. Os informantes, e seus respectivos grupos de controle, deveriam interpretar estes enunciados, em um dos testes com a ajuda de figuras e no outro produzindo um enunciado a partir de uma imagem. O sujeito não expresso foi corretamente interpretado e os índices de acerto foram os mesmos do grupo de controle, portanto não houve sinais de erosão linguística neste contexto. Contudo, os mesmos informantes comprovaram indícios de erosão nos casos das frases em que o sujeito era expresso: a interpretação não coincidia com aquela do grupo de controle. A conclusão dos autores é de que há sinais de erosão linguística não na simples questão sintática, mas quando há uma interface entre sintaxe e pragmática.

Para concluir esta revisão a respeito da produção referente aos trabalhos sobre a erosão linguística do italiano, citamos algumas considerações de Sorace (2004). A estudiosa italiana dedica sua pesquisa mais à aquisição que à erosão, mas no ensaio a que nos referimos opera uma série de considerações sobre Montrul (2004) que podem ser úteis para qualquer estudo sobre erosão. Fazendo referências a seu outro trabalho⁸⁵, Sorace observa que, no estudo de Montrul (2004) os informantes, por serem de segunda geração, necessariamente devem ser considerados pouco adequados à pesquisa, já que a aquisição certamente não foi completa, e portanto os dados obtidos podem ser confundidos com este fator e não serem devidos à erosão. Finalmente, e de grande importância para o nosso trabalho, lembramos que, segundo Sorace (2004), deve ser feita uma diferenciação entre a erosão individual e aquela de uma comunidade, e para determinar os efeitos da erosão teria sido fundamental certificar-se do que

85 Tsimpli, Sorace, Heycock & Filiaci (2004).

exatamente era conhecido pelos informantes antes do início do processo de erosão. A estudiosa afirma que maiores informações sobre os informantes poderiam ter ajudado na compreensão de como foram obtidos os resultados.

Como se pode observar, mais uma vez a questão metodológica é debatida, e um adequado projeto de construção da pesquisa é fundamental para o sucesso da mesma. No parágrafo final deste capítulo serão debatidas as escolhas que guiaram e guiam o grupo coordenado por Raso, e do qual este trabalho faz parte, e será exposto de que forma elas foram tomadas.

2.8 Estudos sobre a erosão linguística do Português do Brasil

Se excetuarmos os estudos sobre línguas minoritárias, o interesse pela erosão linguística no Brasil ainda possui pouquíssima tradição. São escassos os trabalhos que retratem a erosão do PB em contato com outras línguas ou o contrário. Entre aqueles de nosso conhecimento, citamos: Major (1998), que aborda os efeitos fonéticos no inglês dos americanos em contato com o PB; as pesquisas de M. Schoenmakers-Klein Gunnewiek (1989 e 1997) que versam sobre a perda linguística em emigrantes de língua portuguesa na Holanda⁸⁶ e de emigrantes holandeses no Brasil⁸⁷.

No Brasil, além dos já citados estudos de Raso (2003, 2009), Calvo Capilla (2007) realizou uma pesquisa⁸⁸ em que analisa a erosão linguística de hispanofalantes espanhóis adultos residentes no Brasil, cuja L2 de contato é o PB. O objetivo da pesquisa era estudar justamente as interferências da L2 sobre a L1 dos informantes. A pesquisadora debate qual seria o elemento mais importante no desencadear da erosão, citando a Hipótese da Ilha Deserta⁸⁹ e o influxo da L2, em seus fatores internos⁹⁰ e externos, ou seja, a transferência da L2⁹¹.

A pesquisa foi de tipo qualitativo: foram investigados oito informantes espanhóis nativos adultos, residentes desde longa data no Brasil. A coleta dos dados foi efetuada através de entrevistas gravadas semi estruturadas, depois transcritas, e um questionário sociolinguístico. O *corpus* coletado foi analisado pela pesquisadora à procura de desvios da

86 Schoenmakers-Klein Gunnewiek (1989 e 1998).

87 Schoenmakers-Klein Gunnewiek (1997).

88 Na realidade sua dissertação de Mestrado.

89 Já citada acima em 2.1.2.

90 Capilla se refere em específico à Teoria da Simplificação, como apresentada por Py e Grosjean (2002) e Silva-Corvalán (1994), e ao Princípio de Redução da Redundância de Seliger (1989, p. 173).

91 Neste caso os autores de referência da pesquisadora são Schmid e de Bot (2004, p. 212); Schmid (2006) e Scharwood Smith (1989, p. 185).

norma, e foi feita uma triangulação dos dados através da análise do mesmo *corpus* por parte de dois linguistas espanhóis, que verificaram os mesmos erros. Ao invés de utilizar um grupo de controle, foram empregados dois *corpora* linguísticos de referência da língua espanhola⁹².

Calvo Capilla (2007) conclui que os desvios da norma foram, em sua maioria, devidos a transferências da L2. Portanto, segundo ela, o PB é determinante na erosão linguística do espanhol. Foram encontrados sinais de erosão no nível léxico (extensões ou decalques semânticos); no nível morfossintático, houve forte presença de desvios no âmbito das regências verbais, advérbios, conjunções e preposições; na sintaxe, o maior número de desvios foi relacionado aos pronomes e a uma maior frequência de expressão do sujeito pronominal. A autora reitera a opinião de que todos estes desvios sejam devidos ao influxo do PB, e que a conservação observada da L1 seja devida a atitudes positivas em relação à língua e cultura nativa que os informantes demonstraram nos questionários sociolinguísticos.

2.9 Os estudos coordenados por Raso

Os únicos estudos de nosso conhecimento que deliberadamente buscam evitar uma aquisição incompleta do italiano são aqueles feitos no Brasil, no contato com o português brasileiro, que se iniciaram junto à USP de São Paulo sob a supervisão de Loredana de Stauber Caprara: trata-se de doze entrevistas e uma conversação livre a italianos cultos publicadas na *Revista de Italianística* (1997). O estudo citado acima de Tsimpli *et al.* (2004) utilizou informantes que possuíam uma proficiência quase nativa da L2 e utilizavam sua L1 diariamente e podemos deduzir que sua aquisição da L1 seja similar àquela dos informantes dos trabalhos brasileiros, mas no artigo este fator não foi citado explicitamente.

Raso (2003) fez uma análise preliminar do *corpus* de São Paulo e detectou vários pontos interessantes para futuras pesquisas. Em específico, ficou evidenciado que os efeitos do PB sobre o It não se limitam a interferências de tipo lexical, como os decalques semânticos e lexicais e interferência no sistema dos afixos; morfossintáticas, como nas locuções preposicionais, adverbiais ou das conjunções decalcadas do PB ou usadas impropriamente, além de desvios nas regências verbais, nominais e adjetivais; ou sintáticas, com uso impróprio do gerúndio. Raso (2003) notou que os desvios mais interessantes referem-se a um âmbito mais sutil e menos evidente, mas mais profundo da língua, que é aquele que se situa entre a sintaxe e a pragmática, ou seja a estrutura informacional do discurso.

92 Para maiores detalhes remetemos à pesquisa da autora e a seu artigo Calvo Capilla (2007 b).

A diferença entre os trabalhos coordenados por Raso (Raso, 2003, Vale, 2007, Raso e Vale 2009, Raso 2009) e aquele de Calvo Capilla, que é muito similar àquele de Raso (2003), é que a pesquisadora identifica os pontos de evidência de erosão na base da gramaticalidade, mas não aprofunda os fenômenos não visíveis, como a evitação de estruturas. Este tipo de pesquisa, mais sutil, requer outras metodologias para que seja possível detectar tal tipo de erosão, como técnicas estatísticas e utilização de *corpora* comparáveis, como será visto adiante.

A partir do primeiro artigo, Raso coordenou algumas investigações sobre o It em contato com o PB⁹³, do qual também esta dissertação de mestrado faz parte, especialmente sobre o sistema pronominal, que deverá provar como a erosão linguística não se limita aos traços morfossintáticos, mas toca, como foi intuído por ele na análise preliminar do *corpus* de São Paulo, a estrutura informacional do discurso. Destes trabalhos se falará detalhadamente adiante. O que é necessário ressaltar, e que impõe-se como caráter de novidade, é a importância dada à escolha dos informantes. Não se trata mais de migrantes cuja língua principal muitas vezes era o dialeto, mas de falantes cultos, o que garante, em teoria, que a aquisição da L1 foi completada e portanto os sinais de erosão não serão confundidos com outros fatores.

93 Veja-se a respeito: Raso (2003), Vale (2007), Raso (2009), Raso e Vale (2009).

3 METODOLOGIA

O presente capítulo é dedicado à justificação das escolhas metodológicas adotadas na pesquisa. Serão explicadas a seleção dos informantes e dos fenômenos analisados e a opção pela adoção da linguística de *corpus* para este tipo de estudo. Serão fornecidos os dados sociolinguísticos sobre os informantes e serão descritos o *corpus* utilizado e aqueles de comparação. Finalmente, serão expostas as reflexões sobre os limites metodológicos da pesquisa e as implementações necessárias no futuro.

Esclarecemos desde já, como foi dito no capítulo anterior, que uma das diferenças fundamentais deste trabalho em relação a outros sobre erosão linguística está na escolha por um determinado tipo de informante: optou-se por informantes cultos, significando com isto que seu percurso de instrução somou pelo menos treze anos de estudo na Itália, sendo neste país oito anos de ensino fundamental e cinco de ensino médio. Além disso, outra característica deste informante é possuir título superior, obtido na Itália ou outrem. Isto lhe permite possuir uma aquisição completa da língua materna, pelo menos em teoria, e portanto os desvios que porventura sejam encontrados em sua fala, seriam mais provavelmente sinais de erosão do que devidos a outros fatores. A pesquisa se concentrará sobre o contato prolongado por parte dos informantes com o PB, isto quer dizer que eles devem residir no Brasil há pelo menos oito anos. Desta forma é possível considerar que a aquisição do português esteja em um estado avançado, ao ponto de poder considerar sua competência como quase nativa.

Enfim, a novidade deste trabalho, que se insere no projeto geral de Raso sobre a erosão linguística de italianos cultos em contato com o PB, consiste na opção pela utilização da metodologia da linguística de *corpus* que, segundo nosso conhecimento, raramente foi utilizada nos trabalhos sobre erosão⁹⁴.

3.1 A importância dos estudos de *corpora*

A partir do início dos anos sessenta do século passado, a linguística de *corpus* tem se afirmado como uma metodologia extremamente útil no estudo da língua e como instrumento de validação de teorias⁹⁵. A arquitetura de criação de um *corpus* requer antes de mais nada que

94 O único trabalho que conhecemos que se valeu deste tipo de metodologia é o já citado de Calvo Capilla (2007). Contudo, apesar de ter coletado um *corpus* de fala para seu estudo, a autora não o comparou a outro *corpus* de monolíngues mas, após ter detectado os desvios na fala de seus informantes certificou-os consultando dicionários de referência do espanhol (veja-se Capilla, 2007, p. 60). Portanto a pesquisa em si não pode a rigor ser considerada como de linguística de *corpus*.

95 Citamos aqui alguns dos textos de referência na linguística de corpus: Biber (1993), Atkins; Clear ; Ostler (1992); Biber; Conrad; Reppen (1998); Mc Enery; Xiao; Tono (2006); McEnery; Wilson (1996); Meyer (2004); Sardinha (2004).

sejam claros os objetivos: há *corpora* extremamente grandes, cujos metadados permitem a busca automática de vários itens⁹⁶, assim como é possível criar também *corpora* menores para finalidades específicas⁹⁷.

Uma diferenciação entre os tipos de *corpora* refere-se àqueles escritos e aos orais. Os primeiros podem incluir textos, integrais ou trechos, de diferentes tipologias como: jornalísticos, acadêmicos, correspondências públicas ou particulares, literários, entre outros. A dificuldade de coleta que se apresenta na compilação deste tipo de *corpus* é relativa aos textos de tipologia literária mais recentes, que implicariam no pagamento de direitos autorais, o que acaba por impossibilitar sua inclusão, a não ser de trechos muito curtos. Os *corpora* orais se distinguem por abrangerem exemplos de língua falada como: conversações e diálogos formais e informais, conversações telefônicas, conferências e debates, entrevistas, entre outros. Estes *corpora* podem apresentar ou não o áudio, portanto podemos encontrar *corpora* somente das transcrições e outros das transcrições e gravações. Os problemas que o pesquisador pode enfrentar neste tipo de coleta são, por um lado, conseguir a autorização de uso das informações dos falantes, por outro, os altos custos econômicos e de tempo envolvidos na gravação e transcrição da fala.

Há uma série de passos a serem seguidos na criação de um *corpus*: os estudos linguísticos exigem uma grande quantidade de dados, que devem ser coletados, compilados e organizados de forma sistemática para análise posterior. O que é mais importante, grande ou pequeno que seja o *corpus*, é a representatividade. Esta é a diferença entre um *corpus* e um simples arquivo: o *corpus* deve ser pensado para representar uma língua ou uma variedade linguística específica. Para atingir uma representatividade mínima, utilizam-se os conceitos de amostragem e balanceamento. Dois fatores influem sobre estes: a amostragem cobre as variáveis de uma população, que supõem-se representativas daquela língua específica. A amostragem inclui desde os diferentes tipos de texto escolhidos para aquela língua, até o grau de distribuição linguística dos itens pesquisados na língua. Estas categorias textuais são geralmente balanceadas proporcionalmente, para que no momento da inclusão no *corpus* elas sejam um modelo em pequena escala daquela língua. Como pode ser intuído, o conceito de representatividade é circular, pois apoia-se nas amostras e no balanceamento que, por sua vez,

96 As dimensões dos *corpora* podem variar mas geralmente entende-se por *corpus* de grande dimensão aquele que supera o milhão de palavras. *Corpora* deste tipo requereram esforços de anos e de um grande número de pesquisadores. Geralmente são *corpora* de língua inglesa como o *Brown Corpus*, o primeiro a atingir o milhão de palavras, ou o *BNC (British National Corpus)*, que coleta cem milhões de palavras.

97 "*specific purpose corpora*".

sustentam-se na representatividade. Os teóricos concordam que na realidade não há uma maneira científica de medir o balanceamento, e estas noções estão fundamentadas "na intuição e melhor estimativa"⁹⁸ (Mc Enery; Xiao; Tono, 2006, p. 16, tradução nossa).

Outra questão de grande importância que deve ser levada em conta na criação de um *corpus* é a validação de seus dados. Esta serve para garantir o rigor na coleta e a precisão do *corpus*. Geralmente opera-se uma validação interna, que consiste nas várias revisões, geralmente feitas por transcritores e analistas diferentes que verificam erros de transcrição, formatação, espaçamento, etiquetagem entre outros possíveis. Eventualmente pode haver também uma validação externa posterior, que garante ainda mais a acurácia do *corpus*.

Dependendo do tipo de pesquisa a ser realizada, e mesmo para futuras análises, são feitas anotações no *corpus*. A anotação básica que fornece os metadados essenciais, como língua, criador do *corpus*, dados sociolinguísticos, entre outros, é a anotação estrutural (*structural mark-up*), que geralmente é encontrada sob forma de cabeçalhos. Existe além desta, a anotação morfossintática (*Part Of Speech Markup*), feita através do uso de etiquetas, colocadas através de programas específicos ou manualmente, que podem ser postas junto a todos os itens morfológicos ou sintáticos, ou somente naqueles de interesse do pesquisador. Estas anotações servem para facilitar o trabalho de varredura do *corpus* por parte do estudioso, muitas vezes através de programas específicos que fazem a busca em automático.

Como foi dito acima, os estudos sobre erosão linguística baseados em *corpora* são escassos. Isto talvez deva-se ao fato de que montar um *corpus* envolve um volume de trabalho muito grande e nem sempre os itens linguísticos que se procuram podem aparecer. Os testes de que se falou acima, ao contrário, são concebidos justamente para induzir determinados tipos de estrutura. Contudo o que é uma grande vantagem do *corpus*, em relação a qualquer teste, é a possibilidade de obter dados o mais próximos possível do que é efetivamente o uso real do informante. Isto resulta ser ainda mais importante quando se trata de analisar uma língua em processo de erosão, pois na escrita o informante seria muito mais inclinado a controlar sua produção, enquanto na fala este controle pode ser reduzido. Enfim, para corroborar ainda mais esta escolha por um trabalho baseado em *corpus*, foram seguidas as sugestões de Biber (1990 e 1993), que propõe que, caso não seja possível criar um *corpus* de amplas dimensões, como desejável, deverão ser privilegiadas as mais várias tipologias e um número maior de informantes e de textos, e além disso, afirma que para itens frequentes,

98 Mc Enery; Xiao; Tono (2006, p. 16): *on intuition and best estimates*.

como boa parte dos pronomes, excertos de cerca de 1000 palavras fornecem uma amostragem suficientemente confiável.

3.2 O *corpus* analisado por Raso e Vale

Os trabalhos coordenados por Raso a partir de sua análise preliminar⁹⁹ do *corpus* compilado junto à USP de São Paulo nos anos noventa¹⁰⁰, mencionado acima, concentraram-se sobre aspectos específicos do sistema pronominal. Para tal não foi utilizada a totalidade do *corpus* de São Paulo, que perfaz 54.810 palavras, mas foram extraídos alguns trechos, complementados com um pequeno *corpus* montado junto à UFMG nos anos de 2004 a 2006, perfazendo um total de 18.080 palavras. Serão descritas em seguida as características do *corpus* de São Paulo e aquele da UFMG.

3.2.1 Características do *corpus* de São Paulo

O *corpus* de São Paulo foi montado nos anos noventa sob a coordenação de Loredana de Stauber Caprara, que supervisionou o trabalho de um grupo de estudantes de mestrado na parte de entrevistas e transcrições. Ele é constituído por doze entrevistas e uma conversação livre realizadas entre 1994 e 1996. Destas doze entrevistas, dez são presenciais e duas são telefônicas, enquanto o último texto é uma conversação livre entre um médico, uma professora aposentada e uma professora em atividade. A maioria das entrevistas foi feita pelos estudantes de mestrado, e portanto brasileiros, na casa dos informantes, os quais foram contatados sob o pretexto de uma pesquisa sobre a emigração dos italianos de classe média alta no Brasil. Os entrevistadores eram guiados por um questionário pensado para verificar os mais variados registros linguísticos, dependendo dos tipos de respostas dadas pelos informantes. O questionário abrangia desde as motivações que levaram os entrevistados a deixarem a Itália, até o impacto com a nova realidade brasileira, o uso da língua italiana, as atividades na comunidade italiana, a educação dos filhos, os interesses culturais, a frequência das viagens à Itália e como eles se adaptaram à nova realidade. Os entrevistadores deveriam intervir o menos possível e somente nas conversações telefônicas houve uma maior presença de sua fala.

Algumas das características dos entrevistados são comuns: todos são italianos nativos, viveram na Itália até a idade adulta, possuem título superior e concluíram na Itália pelo menos

99 Raso (2003).

100 *Revista de Italianística*, 5, 1997, pp. 29-272.

o segundo grau, alguns até a graduação. Isto, como foi já discutido, garante que sua competência na L1 seja completa e que eles possuam também um boa capacidade de reflexão metalinguística. Além disso, todos os entrevistados, no momento da entrevista, residiam no Brasil há pelo menos vinte anos, alguns até há trinta ou mais anos, portanto podem ser considerados plenamente bilíngues e ótimos candidatos ao estudo da erosão ¹⁰¹.

As normas utilizadas para a transcrição são aquelas adotadas nos volumes da *Gramática da língua falada*, Campinas, Unicamp (1993). Um dos grandes problemas deste *corpus* é que as gravações foram perdidas após a transcrição. Surgem portanto uma série de questões de difícil solução: não é possível verificar eventuais discrepâncias que poderiam despontar devido a uma transcrição imprecisa (coisa ainda mais importante já que foi feita por estudantes brasileiros), assim como é impossível uma análise fonética. É impraticável também explicitar através das gravações pontos onde há margem para mais de uma interpretação, ou avaliar a curva entonacional, necessária para análises sintático-pragmáticas específicas. Finalmente, apesar de o *corpus* possuir uma dimensão significativa e este ser um fator positivo, outra desvantagem apresentada é que as situações comunicativas estão circunscritas praticamente a uma única tipologia, a entrevista, o que limita a pesquisa, seja por não contemplar outras modalidades, seja por ser pouco espontânea, em outras palavras, trata-se de uma tipologia que não é nem verdadeiramente espontânea nem uma tipologia textual codificada. Esta questão merece uma reflexão, pois múltiplas situações comunicativas, que podem ir desde discussões entre conhecidos, a compras em lojas, instruções sobre como proceder em diferentes situações, comentários sobre assuntos diversos e assim por diante, permitem obter uma variedade muito maior de estruturas e de léxico, sendo portanto possível vislumbrar um panorama mais amplo e mais confiável da língua. Lembramos contudo quão difícil é conseguir entrar em contato com informantes que possuam características tão específicas quanto aquelas requeridas para esta pesquisa e, ainda por cima, criar situações o mais espontâneas possível em uma língua que não é a L1 do lugar.

O *corpus* montado na UFMG sob a coordenação de Tommaso Raso tem por objetivo continuar e aprofundar as análises iniciadas no *corpus* anterior. Tendo em vista as limitações do *corpus* de São Paulo, devidas sobretudo à ausência das gravações, decidiu-se coletar novos dados, seguindo critérios mais rigorosos. Inicialmente, entre 2004 e 2006, foi feita uma série de gravações de situações comunicativas diferentes de italianos cultos que tivessem as

101 Para maiores informações sobre o perfil de cada informante remetemos a Raso (2003, p. 13-17).

mesmas características que foram estabelecidas para o trabalho anterior. Devido à qualidade destas gravações, nem sempre suficientemente nítidas, foram transcritos somente quatro monólogos (trechos de conferências) e um diálogo. As gravações são conservadas, portanto é possível consultá-las. O trabalho de transcrição ficou a cargo de um grupo de alunos da Iniciação Científica da UFMG, cotejado em seguida por uma equipe de professores italianos nativos. Os critérios de transcrição seguiram o formato CHAT (Codes Human Analysis of Transcripts)¹⁰², do projeto CHILDES CLAN¹⁰³. Este formato tem a vantagem de:

- Permitir transcrições sucessivas com objetivos e modalidades diferentes;
- Possuir um sistema de codificação conhecido internacionalmente;
- Possuir várias ferramentas de análise;
- Possibilitar a disponibilização na internet do *corpus* junto ao site do projeto CHILDES (no endereço <<http://childes.psy.cmu.edu>>)

Apesar dos cuidados tidos na compilação deste *corpus*, sua desvantagem consiste em seu tamanho bastante reduzido e no fato de as situações comunicativas serem limitadas.

Uma segunda parte deste *corpus* foi constituída para o presente trabalho, iniciado em 2008, que é a contribuição desta Dissertação à pesquisa coordenada por Raso e ao *corpus* por ele implementado. Este novo *corpus*, que amplia o já existente, contém modalidades de comunicação mais variadas, e permite uma análise mais aprofundada dos fenômenos encontrados.

3.3 O novo *corpus* Raso-Ferrari

Discutiremos agora sobre como foi feito o trabalho de coleta e transcrição do novo *corpus*, quais as escolhas envolvidas e quais as características básicas do mesmo.

Gostaríamos de lembrar uma informação que acreditamos ser importante para entender como foi feito este trabalho. A pesquisadora é falante nativa do italiano: não obstante tenha nascido no Brasil, cresceu e efetuou todo seu ciclo de estudos na Itália ou em escolas italianas. Apesar de não possuir todos os requisitos para ser considerada também como informante, este fato foi muito relevante nos contatos com os sujeitos pesquisados. Isto porque todos os contatos entre a pesquisadora e os informantes foram feitos em língua italiana, o que os deixou mais à vontade, também no momento das gravações de cujas interações muitas vezes a pesquisadora participou, descartando em seguida seus dados. Esta

102 MacWhinney (1994).

103 Mac Whinney (2000).

participação possibilitou uma espontaneidade, mesmo nos assuntos ou na expressão de opiniões pessoais, muito maior do que se a pesquisadora não fosse identificada pelos participantes como italiana.

3.3.1 Coleta de dados e transcrições

O novo *corpus* constitui-se de 21.298 palavras, agrupadas em oito textos, com oito informantes diferentes. Dois destes últimos são os mesmos que compareceram na primeira etapa do *corpus* da UFMG, pois um dos textos foi revisado e reaproveitado.

As características dos informantes deste novo *corpus* seguiram aquelas indicadas na Introdução e basicamente as mesmas adotadas nos *corpora* da USP e da UFMG. A amostragem foi não-probabilística: foram escolhidos indivíduos considerados adequados pela pesquisadora e disponíveis para pesquisa, em uma combinação daquilo que é definido como *judgement e convenience*. Dada a especificidade da pesquisa, o número de informantes que possuem as características adequadas é muito reduzido, então seria impossível se fazer uma amostragem proporcional. As gravações efetuadas e transcritas, para um total de quase 35.000 palavras, na realidade não foram totalmente utilizadas neste *corpus* justamente para não comprometer o já tênue balanceamento e proporcionalidade das amostras, pois eram trechos muito longos de interações e alguns informantes acabariam por prevalecer no *corpus*. Em futuro, quando será coletado um *corpus* maior e mais variado, é possível que partes destas transcrições e gravações agora descartadas possam ser reaproveitadas.

Houve uma certa dificuldade em encontrar os informantes que possuíssem as características estabelecidas para a pesquisa. Os motivos principais foram dois: em primeiro lugar constatou-se que o número de italianos nativos que possui título universitário é menor do que se esperava, ou os que o possuem encontram-se no Brasil há menos dos oito anos previstos para poderem ser incluídos na pesquisa. Este fator por si só limitou bastante a totalidade dos possíveis informantes. Em segundo lugar, entre as pessoas contatadas, muitas se recusaram a participar da pesquisa, ou por embaraço, ou por temor de serem reconhecidos, apesar das garantias de sigilo das gravações. Este conjunto de acontecimentos atrasou o início das gravações e conseqüentemente das transcrições, portanto prejudicou de certa forma os resultados obtidos.

Os contatos foram feitos pela própria pesquisadora em seu círculo de conhecidos, e através de indicações por estes feitas, e as gravações aconteceram em sua residência, ou na

residência dos informantes e em seus locais de trabalho. Privilegiaram-se as situações comunicativas mais diversas, evitando aquela da entrevista, já amplamente desfrutada anteriormente. As tipologias que comparecem no *corpus* são as seguintes:

cinco diálogos

dois monólogos

uma conversação

Os diálogos acontecem entre os seguintes informantes¹⁰⁴ e nas seguintes situações e ambientes:

Título	número total de palavras	número de palavras por informantes	Tipologia e descrição da interação
Pranzo	4408	MRC: 1843 MAS: 2565	Diálogo entre os dois participantes durante o almoço em casa de MAS: os assuntos versam opiniões sobre futebol, o vício do cigarro e uma longa discussão sobre vinhos. (LAF está presente mas fala pouco). Gravado com microfone omnidirecional.
Partita 1	1284	UCR: 533 MAS: 751	Diálogo entre UCR e MAS e depois entre UCR e LAF enquanto aguardam que inicie um jogo de futebol na televisão, na casa de MAS. O primeiro diálogo trata de comentários sobre pessoas conhecidas e futebol; o segundo diálogo sobre o que é feito nos finais de semana. Gravado com microfone omnidirecional.
Partita 2	3578	UCR: 1348 MAS: 2230	Conversação entre MAS, UCR, ALG (informante nativo não adequado à pesquisa) e LAF: são comentários feitos durante um jogo de futebol na televisão, na casa de MAS. Gravado com microfone omnidirecional.
Sorelle	3691	PAT: 1194 LIV: 2497	Diálogo, dividido em duas partes, entre duas irmãs. O primeiro diálogo versa sobre livros e aulas de italiano. O segundo sobre a mudança de apartamento de LIV e a nova decoração da casa. Gravado com gravador de fita cassete.
Missionari	1764	MON: 1764	Monólogo: é uma entrevista a uma missionária sobre sua experiência no Brasil (LAF, a entrevistadora, intervém muito pouco). A entrevista foi feita na sede da paróquia. Gravado com microfone omnidirecional.
Genitori	1480	ANG: 1480	Monólogo, com algumas esporádicas intervenções de outros participantes que não são informantes adequados à pesquisa. ANG relata sua mudança com a família para o Brasil, as motivações que os trouxeram e as dificuldades iniciais. Gravado na

104 Serão utilizadas aqui suas siglas do código CHAT. Esclarecemos que a sigla LAF se refere à pesquisadora cujas falas não foram computadas nem analisadas.

			paróquia onde ele trabalha. Gravado com microfone omnidirecional.
Medici	2144	GIC: 2144	Diálogo entre GIC e LAF (a interação de GIC é muito superior à de LAF), na residência de GIC, em que GIC comenta sobre suas experiências negativas com os médicos e relata vários episódios sobre o assunto. Gravado com microfones de lapela.
Cena	2949	MAS: 2949	Diálogo entre MAS e LAF (a presença de LAF é bastante grande) em sua residência. Enquanto preparam o jantar comentam sobre os ingredientes e a forma de cozimento, o resultado da receita e vários assuntos. Gravado com microfones de lapela.

Quadro 2: informações sobre o *corpus* e cada um dos textos.

As gravações foram realizadas em formato .wav utilizando aparelhos sofisticados que garantissem uma boa qualidade acústica:

Gravador digital Marantz PDD660 com cartão de memória Compact Flash de 2 *gigabytes*; *Kits wireless* Sennheiser Evolution EW100 G2 (*receiver*, *transmitter*, microfone de lapela) com dois *kits* bateria/carregador adaptados para o *receiver*, ou solução alternativa com bateria própria e seis microfones completos;

Microfone omnidirecional Sennheiser MD 421 com pedestal Hunter PMP103 e cabos RCL303569 de 6 metros, ou sistema *wireless*;

O microfone de lapela permitiu obter gravações com um bom índice de espontaneidade, pois o fato de não impedir os movimentos e não serem minimamente desconfortáveis possibilitou a descontração das interações. Também o microfone omnidirecional, utilizado em algumas das gravações por ser de mais fácil instalação, garantiu uma boa qualidade de gravações e, aparentemente, não condicionou os diálogos entre os informantes que praticamente esqueciam de sua presença.

As transcrições foram efetuadas seguindo o formato CHAT, pelos mesmos motivos ilustrados acima. As anotações foram mínimas: reduziram-se a um cabeçalho simplificado que contém os metadados dos informantes e o local e data de gravação. Optou-se, após várias tentativas, que vão ser explicitadas adiante, por não fornecer o *corpus* de nenhuma anotação morfossintática. As intervenções da pesquisadora foram descartadas no momento da análise e da contagem dos dados, por ela não se configurar nos requisitos necessários.

3.3.2 Perfil sociolinguístico dos participantes do novo *corpus*

Os entrevistados possuem características comuns: todos são italianos nativos, mudaram-se para o Brasil após a idade adulta e concluíram pelo menos o segundo grau na Itália. Três deles iniciaram os estudos universitários na Itália, cursando-os por pelo menos dois anos, mas não chegaram a concluí-los, os outros os concluíram na Itália ou no Brasil. O tempo de permanência no Brasil é, excetuando-se dois casos, muito inferior em relação aos informantes do *corpus* de São Paulo, sendo em média de dez anos ou pouco mais. Também a origem dos informantes é bastante variada. Não foi aplicado um questionário sociolinguístico aos informantes para não criar situações de inibição no momento das gravações, mas muitos dados foram coletados em contatos anteriores ou posteriores.

Todos os informantes autorizaram o uso de suas contribuições através da assinatura de um *Termo de Consentimento* (Anexo A) aprovado pelo Comitê de Ética em Pesquisa (COEP) da Universidade Federal de Minas Gerais. Este Termo garante a confidencialidade das informações coletadas: a identidade do informante é mantida em sigilo. Os resultados do estudo são apresentados como o retrato de um grupo e não de uma pessoa. Na parte da gravação publicada e nas transcrições os informantes são indicados com uma sigla que não permite de maneira alguma sua identificação. Relacionados à sigla são disponibilizados somente a idade, dividida por faixas, o nível de estudo, o sexo, a tipologia de trabalho e a cidade de proveniência. Somente a pesquisadora tem conhecimento da identidade de cada um dos informantes e de sua relativa sigla.

Abaixo indicaremos as principais características sociolinguísticas dos informantes, lembrando que as faixas etárias são: de 18 a 25 anos faixa 1; de 26 a 39 anos faixa 2; de 40 a 60 anos faixa 3; mais de 60 anos faixa 4:

MAS. Sexo: masculino. Idade: faixa 2. Estado civil: casado. Tempo de residência no Brasil: 8 anos. Proveniência: Brescia. Estudos: segundo grau (*maturità scientifica*) e faculdade de Letras concluídas na Itália. Atividade Profissional: professor de italiano e tradutor. Relações com a Itália: volta de férias para a Itália a cada um ou dois anos.

UCR. Sexo: masculino. Idade: faixa 3. Estado civil: casado. Tempo de residência no Brasil: 12 anos. Proveniência: Varese. Estudos: segundo grau técnico concluído na Itália, faculdade de História concluída no Brasil. Atividade profissional: professor de

- italiano. Relações com a Itália: volta raramente, nos doze anos em que esteve no Brasil voltou três ou quatro vezes.
- GIC. Sexo: masculino. Idade: faixa 3. Estado civil: separado. Tempo de residência no Brasil: 15 anos. Proveniência: Torino. Estudos: concluiu o segundo grau (maturità classica) na Itália onde iniciou a Faculdade de Arqueologia, que não chegou a concluir. Atividade profissional: gráfico e tradutor. Relações com a Itália: voltou para a Itália somente uma vez nesses anos.
- MON. Sexo: feminino. Idade: faixa 3. Estado civil: solteira. Tempo de residência no Brasil: 11 anos. Proveniência: Oristano. Estudos: concluiu o segundo grau e a faculdade de Ciências Religiosas na Itália. Atividade profissional: missionária. Relações com a Itália: retorna de férias a cada dois-três anos.
- ANG. Sexo: masculino. Idade: faixa 4. Estado civil: casado. Tempo de residência no Brasil: 20 anos. Proveniência: Bergamo. Estudos: concluiu o segundo grau (maturità classica) na Itália onde iniciou a Faculdade de Lettere, que cursou por dois anos mas que não chegou a concluir. Atividade profissional: contador e missionário. Relações com a Itália: retorna quase todos os anos, mesmo por alguns meses.
- MRC. Sexo: masculino. Idade: faixa 3. Estado civil: casado. Tempo de residência no Brasil: 19 anos. Proveniência: Milano. Estudos: concluiu o segundo grau (maturità scientifica) na Itália onde iniciou a faculdade de Geologia mas não a concluiu; no Brasil graduou-se em matemática. Atividade profissional: professor de matemática e italiano. Relações com a Itália: retorna muito raramente.
- LIV. Sexo: feminino. Idade: faixa 3. Estado civil: casada. Tempo de residência no Brasil: 33 anos. Proveniência: Chiasso (Suíça italiana). Estudos: concluiu o segundo grau na Itália e graduou-se no Brasil. Atividade profissional: professora de italiano. Relações com a Itália: retorna de vez em quando de férias.
- PAT. Sexo: feminino. Idade: faixa 3. Estado civil: separada. Tempo de residência no Brasil: 33 anos. Proveniência: Chiasso (Suíça italiana). Estudos: concluiu o segundo grau na Itália e graduou-se no Brasil em Letras. Atividade profissional: professora universitária de italiano. Relações com a Itália: retorna de vez em quando de férias.

Para uma melhor visualização destes dados, abaixo foi inserido um quadro com as informações principais:

Informante	Sexo	Idade	Proveniência	Título de estudo	Tempo de residência	Profissão	Retorna à Itália
MAS	M	faixa 2	Brescia	segundo grau e faculdade na Itália	8 anos	professor e tradutor	a cada 1 ou 2 anos
UCR	M	faixa 3	Varese	segundo grau na Itália, graduação no Brasil	12 anos	professor	raramente
GIC	M	faixa 3	Torino	segundo grau e faculdade (incompleta) na Itália	15 anos	tradutor e gráfico	raramente
MON	F	faixa 3	Oristano	segundo grau e faculdade na Itália	11 anos	missionária	a cada 2 ou 3 anos
ANG	M	faixa 4	Bergamo	segundo grau e faculdade (incompleta) na Itália	20 anos	contador e missionário	quase todos os anos
MRC	M	faixa 3	Milano	segundo grau e faculdade (incompleta na Itália); graduação no Brasil	19 anos	professor	raramente
LIV	F	faixa 3	Chiasso	segundo grau na Itália e graduação no Brasil	33 anos	professora aposentada	de vez em quando
PAT	F	faixa 3	Chiasso	segundo grau na Itália, graduação e pós-graduação no Brasil	33 anos	professora	de vez em quando

Quadro 3: dados sobre os informantes investigados na pesquisa.

3.3.3 Método de análise dos dados: erros e acertos

A análise dos dados de um *corpus* geralmente é feita através de *softwares* específicos que "leem" os vários tipos de anotações que foram postas, sejam elas estruturais, morfossintáticas ou de outro tipo. As próprias anotações, na maioria dos casos, podem ser colocadas de forma automática por programas específicos.

Inicialmente era nosso objetivo fornecer este novo *corpus* de algum tipo de anotação morfossintática que facilitasse o trabalho de extração de dados. O programa que melhor se adaptaria a isto e que mostrara-se suficientemente confiável para a língua italiana (Tamburini, 2000) resultou ser o software *TreeTagger*¹⁰⁵. Foram selecionados alguns trechos do *corpus*

105 Disponível em <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/> acesso em 10 de outubro de 2008.

para um estudo piloto que verificasse a efetiva utilidade do programa. Contudo, após algumas tentativas, percebeu-se que o etiquetador automático não distinguia de forma suficientemente confiável os pronomes, que eram nosso objeto de estudo, de outras partes do discurso homófonas e homógrafas como, por exemplo, os artigos. Isto acontece porque o It é uma língua em que vários morfemas de alta frequência podem possuir mais de uma função e significado, e este fato acabou por confundir o programa. Este problema poderia ser parcialmente resolvido operando um "treinamento" específico do *Tree Tagger*, mas para tal seria necessário um *corpus* muito maior e, mesmo assim, não teríamos nenhuma garantia de obter os resultados esperados. Optou-se portanto por deixar por enquanto o *corpus* sem etiquetas e fazer a busca pelos pronomes manualmente. Pretende-se em futuro, em colaboração com Eckhard Bick autor do parser *PALAVRAS*¹⁰⁶ (BICK, 2000), adaptar o parser a esse tipo de pesquisa. Contrariamente ao *TreeTagger* (que é um software estatístico) o *PALAVRAS* é composto de regras e possui uma adaptabilidade maior.

Outra tentativa de utilizar ferramentas computacionais, o que é prática comum na linguística de *corpus*, foi o de fazer a varredura através de um programa muito simples e gratuito, o *Texstat 2*, que realiza buscas de índices de frequência e de itens específicos, com opção de visualização do contexto. Antes de inserir os dados no programa e antes de analisá-los, foi feita uma limpeza dos textos, retirando-se os metadados e a fala da pesquisadora. Este programa se mostrou bastante útil para uma análise preliminar, mas mesmo assim optou-se, após várias tentativas, por não utilizá-lo por dois motivos: em primeiro lugar a contagem manual mostrou-se mais prática, sendo um *corpus* reduzido, e em segundo lugar, como este é um trabalho sobre erosão, é importante assinalar a ausência de certos itens, portanto esta etapa poderia ser feita somente de forma manual.

A única ferramenta informática que foi utilizada por sua simplicidade na análise do *corpus* foi o programa *Notepad ++*, que serviu para a contagem das ocorrências, limpeza dos textos, e recontagem do *corpus*, e para a averiguação das ocorrências em caso de dúvida. No entanto, todo o trabalho de busca dos pronomes, dada a separação por funções e contextos significativos, seja neste *corpus* seja naquele de comparação, foi feito manualmente.

3.3.4 Limitações do *corpus* Raso-Ferrari

O trabalho de compilação de um *corpus*, como pôde ser visto, é muito demorado e

¹⁰⁶Disponível em <http://visl.sdu.dk/visl/pt/parsing/automatic/>> acesso em 10 de outubro de 2008.

complexo. Serão enumeradas algumas das dificuldades encontradas neste trabalho, que não tinham sido levadas em conta suficientemente no início da pesquisa, dificuldades que poderão explicar parcialmente alguns dos limites do *corpus*, que serão explanados em seguida.

O primeiro obstáculo que se fez presente foi encontrar os informantes que tivessem as características necessárias para a pesquisa. Alguns contatos tinham sido feitos antes do início do trabalho, mas no momento em que se solicitou a colaboração dos possíveis informantes, como foi dito acima, alguns se sentiram inibidos ou temerosos de que o fato de serem gravados pudesse de alguma forma prejudicá-los, e mesmo com todas as garantias de confidencialidade, preferiram não participar da pesquisa. Da mesma forma, como já discutimos, o número de graduados foi inferior ao que se tinha suposto. Isto acabou por dificultar ainda mais o trabalho, pois novos contatos tiveram que ser feitos para se encontrar um número mínimo de informantes dispostos a colaborar.

Além disso, a opção por evitar a tipologia comunicativa da entrevista em favor de diálogos e conversações não foi de fácil solução, porque foi bastante árduo conseguir criar situações em que as pessoas se sentissem à vontade para conversarem livremente de vários assuntos sem um questionário-guia. Outra dificuldade foi juntar mais de um informante que conversasse com o outro em uma mesma situação em italiano, pois as situações naturais do dia a dia previam a presença de algum brasileiro, o que mudaria o código de comunicação. Para evitar este tipo de circunstância, em alguns casos optou-se por ir até a residência ou local de trabalho dos informantes, certificando-se de que não haveria informantes que não fossem italianos, em outros casos foram organizados pequenos eventos sociais como almoços ou jantares em que participassem somente os informantes escolhidos.

Quanto à dificuldade de encontrar ferramentas computacionais adequadas a este projeto, ou seja, sobre o fato de que nenhum *software* pôde ser de efetiva ajuda na etapa de análise, e que esta teve que ser realizada de forma totalmente manual, já foi dito acima.

Vejamos agora as limitações deste *corpus*. O que ressalta é seu tamanho, bastante reduzido. Após o término das transcrições, o número de palavras era de cerca de 35.000, mas os textos não estavam equilibrados, pois alguns eram demasiadamente curtos e outros extremamente longos. Foi, portanto, necessário operar uma seleção e vários cortes, para se chegar a estes oito textos que compõem o *corpus* e que perfazem em média cerca de 1.500 a 3.000 palavras cada.

Outro problema que surgiu foi que, mesmo após os cortes operados nos diálogos e

interações, das 21.298 palavras que compõem o *corpus*, 8.495 pertencem a um único informante, o que pode comprometer o balanceamento e a representatividade.

Estes fatores foram levados em conta no momento da análise e da interpretação de dados e serviram para compreender melhor as dificuldades em se trabalhar com *corpora*. Uma das conclusões, de que se falará adiante, é que para resolver estes problemas será necessário continuar o trabalho aqui iniciado, aumentando o *corpus* em tamanho e em número de textos e informantes.

3.4 Os *corpora* de comparação

O que caracteriza a erosão linguística em um indivíduo é a série de desvios que ocorrem em sua fala. Estes desvios devem ser de alguma forma medidos e quantificados, e isso é possível só comparando sua fala com aquela de alguém que não esteja sujeito à erosão. Schmid e de Bot (2004) consideram que um estudo longitudinal seria a melhor forma de medir a erosão em ato. O ideal seria seguir e testar os informantes antes que a erosão se inicie e possivelmente em intervalos regulares, para detectar o processo em andamento. Entretanto, este tipo de estudo raramente é utilizado, pois requer um enorme dispêndio de tempo e energia, além do fato que nem sempre os informantes, ou os próprios pesquisadores, estão disponíveis por um período tão longo, visto que está-se falando de anos entre uma medição e outra.

Um outro método utilizado em substituição aos estudos longitudinais é a implantação de um grupo de controle de nativos, residente no país de origem, que deveria fornecer os dados da língua, livre de um processo de erosão¹⁰⁷. Este método é aplicado com uma maior frequência em relação ao primeiro citado, contudo não deixa de oferecer dificuldades já que não é fácil encontrar um grupo de controle com características similares às dos falantes com erosão ou, em caso de migração de longa data, a língua falada no país de origem pode não ser mais idêntica àquela praticada pelos informantes quando migraram¹⁰⁸.

Devido a estas dificuldades, as pesquisas coordenadas por Raso se valeram de uma comparação através de *corpora* já prontos, outra metodologia bastante usada. Em específico, nos estudos de Raso e Vale, o *corpus* de comparação utilizado foi o *BADIP (Banca Dati dell'Italiano Parlato)*, na parte referente ao *Lessico di frequenza dell'italiano parlato (Corpus*

107 É o que foi feito no estudo de Keijzer (2007), entre outros.

108 Veja-se a este respeito o parágrafo 2.3 relativo à questão da história da língua italiana.

LIP)¹⁰⁹, enquanto neste nosso trabalho, optou-se por outro *corpus* de língua falada, o *C-ORAL-ROM italiano*¹¹⁰. Os motivos destas escolhas e os detalhes destes dois *corpora* serão explicitados abaixo.

3.4.1 O *BADIP*

O *BADIP* (*Banca Dati dell'Italiano Parlato*) é um site¹¹¹ gratuito da [Karl-Franzens-Universität Graz](http://www.karlfranzens-universitaet-graz.at) (Austria) destinado à publicação de *corpora* para a análise e o estudo do italiano falado. Atualmente o único *corpus* disponível para consulta é uma versão *online* do *Lessico di frequenza dell'italiano parlato (Corpus LIP)*, mas a intenção de seus criadores é poder ampliá-lo com um número maior de *corpora* de acesso gratuito.

O *corpus LIP* foi realizado entre 1990 e 1992 por um grupo de linguistas sob a supervisão de Tullio De Mauro e serviu para a construção do primeiro léxico de frequência do italiano falado, graças à colaboração da Fundação IBM Itália e de um grupo de estudiosos (Vedovelli, Mancini, Voghera e o próprio De Mauro).

O *corpus* é constituído por cerca de 500.000 palavras, sendo considerado portanto de tamanho médio. O número de ocorrências de formas gráficas é de 475`883, enquanto o número de ocorrências de lexemas é de 496.335. Tal diferença é devida à contagem em separado das preposições articuladas e dos clíticos, o que aumenta o número de palavras.

A questão da representatividade geolinguística do *corpus* foi amplamente debatida pelos autores no momento de sua arquitetura. Como seria impossível, dadas as diferenças impostas pela história linguística italiana, limitar-se a uma coleta de dados de uma única localidade, surgiu o problema de quais cidades escolher. Incluir todos os grandes centros urbanos teria sido impossível do ponto de vista prático e pouco representativo de cada realidade. Optou-se, portanto, por selecionar as cidades que teriam um maior peso para a história linguística italiana, seja nos níveis sociolinguísticos mais altos, como Florença e Roma, seja nas camadas da burguesia, como Milão, seja nas faixas mais populares, como Nápoles¹¹².

Quanto à representatividade textual, o *LIP* constitui-se, como foi dito, de textos de língua falada. Foram escolhidas cinco tipologias, que segundo os autores cobririam uma escala de naturalidade da mais baixa à mais alta. Para cada um destes graus foram

109 De Mauro *et al.* (1993).

110 Cresti, Moneglia (2005).

111 disponível em <<http://badip.uni-graz.at/>> acesso em 15 de março de 2009.

112 Para esta discussão veja-se De Mauro *et al.* (1993).

selecionadas 100.000 palavras, 25.000 para cada cidade representada. Os textos somam um número total de 469. As tipologias de que se falou são as seguintes:

Tipologia A: comunicação bidirecional com fala livre entre os participantes, presencial (inclui conversações em casa, nos locais de trabalho, em ambiente escolar e universitário, em locais de lazer ou nos meios de transporte);

Tipologia B: comunicação bidirecional com fala livre entre os participantes, não presencial (inclui conversações telefônicas normais, gravadas no rádio e mensagens gravadas em secretária eletrônica);

Tipologia C: comunicação bidirecional com fala direcionada entre os participantes, presencial (inclui discursos em assembleias legislativas, debates culturais, assembleias estudantis, assembleias sindicais, encontros entre trabalhadores, arguições na escola, exames universitários, interrogatórios processuais, entrevistas no rádio ou televisão);

Tipologia D: comunicação unidirecional em presença do destinatário (inclui aulas escolares ou universitárias, relações em congressos ou convenções políticas, sindicais ou científicas, comícios políticos, homilias, conferências não especialísticas, discursos judiciários);

Tipologia E: comunicação unidirecional não presencial à distância ou relatada sobre um texto escrito (inclui transmissões radiofônicas ou televisivas).

Como foi visto pela descrição, o *LIP* é um *corpus* bastante variado, além de oferecer a vantagem de estar disponível gratuitamente. Entretanto, possui um grande inconveniente que é a ausência das gravações. Isto, como já foi discutido no caso do *corpus* da *USP*, desencadeia uma série de questões de difícil solução: não é possível verificar eventuais incertezas que possam surgir devido a uma transcrição supostamente imprecisa; ou em caso de ambiguidades.

3.4.1.1 Textos utilizados na pesquisa de Raso-Vale

Para finalidades de comparação com o primeiro *corpus* de erosão¹¹³ que fora montado no âmbito dos projetos coordenados por Raso, escolheram-se, entre todos os textos do *LIP*, alguns que fossem os mais próximos possíveis daqueles a serem comparados. Foram selecionados dois textos para a cidade de Milão (tipologia C e D), cinco para a cidade de Roma (tipologias A, B e D), um texto para a cidade de Nápoles (tipologia A) e dois textos

¹¹³ Constituído, como foi dito, de 18.080 palavras, selecionadas entre uma parte do *corpus* da *USP* e outra das novas gravações efetuadas entre 2004 e 2006.

para a cidade de Florença (tipologia A e D). Quanto à preferência pelas cidades do centro-norte da Itália, isto foi devido ao fato de que os informantes sujeitos à erosão provinham em sua maioria destas regiões da Itália, então seria mais coerente manter uma escolha geográfica quanto mais próxima à realidade a ser comparada.

O *corpus* assim montado perfaz um total de 18.080 palavras, comparável com aquele colacionado para a erosão.

3.4.2 O NURC

O NURC *Projeto Norma Urbana Culta*, é um *corpus* de língua falada do PB. Em 1968, em um relatório solicitado pela Comissão de Linguística Iberoamericana do PILEI, o professor Néelson Rossi (UFBA) ressaltava o interesse de se estender ao Brasil a execução do *Proyecto de Estudio Coordinado de la Norma Lingüística Culta de las Principales Ciudades de Iberoamérica y de la Península Ibérica*, de que participavam países de língua espanhola. O relatório propunha que fossem estudadas as normas cultas de cinco grandes centros urbanos brasileiros, seguindo o modelo da pesquisa hiberno-americana. Foram escolhidas as cidades de Recife, Salvador, Rio de Janeiro, São Paulo e Porto Alegre, por proporcionarem uma amostra significativa da população urbana do país na época.

O projeto no Brasil foi chamado NURC, iniciou em 1970 e a etapa de coleta de dados foi concluída na década de noventa. Seu objetivo era registrar a modalidade culta da língua falada nos centros urbanos, ou seja, procurava-se estudar a pluralidade de normas objetivamente comprovadas no uso oral, o que era a verdadeira língua falada. Para tal eram gravadas interações de vários tipos como: elocuições formais, diálogos entre informante e documentador, diálogos entre informantes. Os vários textos eram transcritos e subdivididos por tipo de texto e em que constava a faixa etária dos participantes, o sexo e o tema da interação. Os informantes deveriam possuir todos nível escolar superior e ter nascido, ou vivido desde a primeira infância, no centro urbano onde seus dados eram coletados.

A partir do *corpus* do NURC foram publicados vários trabalhos: em um primeiro momento estes versaram sobre morfologia, sintaxe e léxico; entre o final da década de oitenta e o início da década de noventa os dados serviram para o início da implementação do *Projeto Gramática do Português falado*, cujo objetivo era preparar uma gramática referencial da variante culta urbana brasileira¹¹⁴.

114 Para os detalhes sobre os dados do NURC e a produção científica que foi elaborada a partir deste corpus remetemos às seguintes páginas da web: <<http://www.lettras.ufjf.br/nurc>> acesso em 2 de agosto de 2009;

Até pouco tempo atrás as gravações do NURC não eram consultáveis, pois haviam sido feitas com gravadores de rolo, e assim eventuais dúvidas sobre as transcrições não poderiam ser conferidas. Atualmente a UFRJ disponibilizou *online* em seu site¹¹⁵ o *corpus* com os metadados, as transcrições e as gravações digitalizadas, o que pode auxiliar no trabalho do pesquisador.

3.4.3 O C-ORAL-ROM italiano

O C-ORAL-ROM Cresti Moneglia (2005) é um *corpus* multilíngue de fala espontânea, formal e informal, das principais línguas românicas europeias: italiano, francês, espanhol e português de Portugal, com um total de 1.200.000 palavras. O *corpus* portanto constitui-se de quatro conjuntos de gravações comparáveis nestas quatro línguas¹¹⁶, cada uma perfazendo cerca de 300.000 palavras. Contribuíram com a coleta e trabalho de montagem dos *corpora* nas várias línguas as seguintes instituições:

Università di Firenze (Dipartimento di Italianistica, LABLITA);

Université de Provence (Description Linguistique Informatisée sur Corpus);

Fundação da Universidade de Lisboa/Centro de Linguística da Universidade de Lisboa

Universidad Autónoma de Madrid (Departamento de Lingüística, Lenguas Modernas, Lógica y F. de la Ciencia, Laboratorio de Lingüística Informática).

As transcrições foram feitas em formato CHAT, com anotações prosódicas das quebras terminais e não terminais¹¹⁷. Os textos são precedidos de cabeçalhos que contêm metadados com as informações principais como participantes, data, língua, tipo de interação, qualidade acústica da gravação entre outros. As linhas de texto da transcrição ortográfica são divididas de duas formas: verticalmente, seguindo os turnos dialógicos (introduzidos por uma etiqueta para cada participante); e horizontalmente, com a análise prosódica e a divisão em limite de enunciados, representando as quebras prosódicas terminais e não terminais do contínuo da fala. As transcrições e os vários tipos de etiquetagem são alinhados às gravações através do *software WinPitch* (MARTIN, 1996-2004)¹¹⁸. Como as análises que seriam feitas no *corpus* requeriam uma qualidade acústica muito alta, grande atenção foi dada a esta etapa: foram

<<http://www.fflch.usp.br/dlcv/nurc/index.html>> acesso em 2 de agosto de 2010.

115 Disponível em <<http://www.lettras.ufjf.br/nurc-rj/>> acesso em 2 de agosto de 2010.

116 Está sendo montado, junto à UFMG, sob a coordenação de Raso e Mello a quinta parte deste *corpus*, com gravações do PB; veja-se no site <http://www.c-oral-brasil.org/>.

117 Veja-se a este respeito Cresti e Moneglia (1997).

118 Disponível em <http://www.winpitch.com>.

utilizados microfones de alta qualidade e os textos inadequados foram descartados. Os arquivos de áudio, com sinal mono ou estéreo de acordo com o tipo de microfone utilizado, são arquivados em formato .wav.¹¹⁹

A arquitetura do *corpus* tem como objetivo assegurar uma grande variedade de ocorrências de atos de fala de tipologias as mais diferentes, buscando reproduzir ao máximo o que é encontrado na fala espontânea. Há então uma seção dedicada à fala informal (diálogos, monólogos e conversações em ambiente familiar ou particular) e uma destinada à fala formal, seja em contextos naturais como debates políticos, sermões, aulas, explicações no trabalho, conferências entre outros; seja em contextos midiáticos como reportagens, programas esportivos, culturais, científicos e assim por diante.

Todos os *corpora* que fazem parte do *C-ORAL-ROM*, inclusive o de língua italiana, o *C-ORAL-ROM* italiano, dividem-se portanto em duas partes, uma informal e outra formal. Em específico o *corpus* em língua italiana contém:

Textos informais, para um total de aproximadamente 150.000 palavras e 87 textos diferentes. Os textos variam de um tamanho de 1.500 a um máximo de 4.500 palavras;

Textos formais, para um total de mais ou menos 150.000 palavras divididas em mais de 50 textos de aproximadamente 3.000 palavras cada um.

3.4.3.1 Textos utilizados na presente pesquisa

Para a presente pesquisa, preferiu-se utilizar como *corpus* de comparação o *C-ORAL-ROM* italiano pelos seguintes motivos:

Estão disponíveis as gravações e por isso é mais fácil desfazer incertezas em pontos de pouco clara interpretação, ou valer-se da prosódia para interpretação das funções dos clíticos pesquisados;

É um *corpus* mais moderno e mais recente, com uma grande quantidade de metadados que facilitam o trabalho do pesquisador na escolha dos textos mais adequados;

Apesar da variedade de fala prevalente ser aquela de Florença, isto não comporta um problema para a presente pesquisa, pois os pronomes sujeito não são objeto de nossa investigação neste momento¹²⁰.

¹¹⁹ Para os detalhes técnico do projeto remetemos diretamente à bibliografia dos autores: Cresti, Moneglia (2005).

¹²⁰ A variedade florentina da língua italiana é não PRO-DROP, portanto expressa obrigatoriamente o sujeito.

É perfeitamente comparável com o C-ORAL-BRASIL¹²¹, em fase final de construção e que segue os mesmos critérios de arquitetura do *C-ORAL-ROM*, e poderá, no futuro, ser útil comparar o PB com o *corpus C-ORAL-BRASIL*.

Foram escolhidos 14 textos entre conversações, monólogos e diálogos que fossem o mais parecidos possível com aqueles do *corpus* de erosão analisado, para um total de 21.224 palavras. Os textos utilizados, todos de interações informais em ambiente familiar, foram assim divididos:

- 6 conversações: ex-colegas de escola que relembram os velhos tempos; a organização de uma festa em casa; uma discussão entre amigos enquanto vão assistir a um jogo de futebol no estádio; uma conversa entre amigos em um bar; um grupo de amigos que joga a póquer; uma troca de ideias entre estudantes de arquitetura sobre alguns projetos;
- 6 diálogos: o relato entre amigas de como foi um concerto a que uma delas foi; a preparação de um jantar entre amigas; as dicas de como preparar a receita de um doce; uma discussão sobre política entre amigos; o relato de uma viagem a Cuba entre dois amigos; o relato do presentes ganhos para a formatura;
- 2 monólogos: o relato da viagem de núpcias de uma noiva; a narração da história de um aquecedor que passou de casa em casa.

Desta forma pretendeu-se abranger situações que fossem similares às aquelas coletadas no *corpus* de erosão. Entre os textos escolhidos, tentou-se buscar aqueles com uma maior participação de informantes de localidades diferentes da Itália, tentando assim obter uma maior variação diatópica, além de optar por aqueles em que os participantes tivessem um nível escolar mais alto, similar portanto àquele dos sujeitos pesquisados no *corpus* de erosão.

Como o italiano padrão é uma língua PRO-DROP, utilizar a variedade florentina no estudo dos pronomes sujeito, como tinha sido cogitado no início deste trabalho, tornaria-se inviável e por isso optou-se por excluí-la.

121 O C-ORAL-BRASIL é um projeto cujo principal objetivo é o estudo do PB de fala espontânea a partir da formação de um *corpus* que siga os moldes do C-ORAL-ROM (Cresti; Moneglia 2005), projeto do qual é uma ramificação. O projeto do C-ORAL-BRASIL é coordenado por Mello e Raso e uma equipe de estudantes de graduação e pós-graduação da UFMG. O *corpus* é segmentado em enunciados que seguem a Teoria da Língua em Ato (Cresti, 2000), segundo a qual o enunciado é a menor unidade linguística que pode ser interpretada pragmaticamente. Para maiores informações veja-se o site <http://www.c-oral-brasil.org/>, no qual se encontra também uma bibliografia sobre o assunto.

4 ANÁLISE DOS RESULTADOS

Este capítulo será subdividido em duas seções. Na primeira, serão apresentados os resultados obtidos na pesquisa de Raso e Vale (2009) sobre os clíticos e serão analisados alguns pontos em aberto. Na segunda parte, serão expostos os resultados obtidos no trabalho aqui descrito. Como se verá, há uma certa discrepância nos dados e nas análises e interpretações entre as duas pesquisas, o que será discutido na parte final da Dissertação.

4.1 Os resultados obtidos nos estudos de Raso e Vale (2009)

Como foi dito acima, a pesquisa feita por Raso e Vale (2009) debruçou-se no estudo da erosão linguística do It em contato prolongado com o PB, analisando 18.080 palavras de um *corpus* de bilíngues montado em parte a partir do *corpus* da USP e em parte completado com um *corpus* organizado na UFMG¹²². Este *corpus* foi comparado com um de monolíngues extraído do *LIP*, para um total de 18080 palavras. A análise se concentrou sobre um grupo de clíticos: os pronomes acusativos de terceira pessoa, o *ci* atualizante, lexicalizante e locativo, e o *ne* em suas várias funções, exceto aquelas lexicalizantes. Serão aqui examinados os pronomes acima, em um estudo comparativo feito sobre um *corpus* de fala de italianos sujeitos a erosão diferente do de Raso e Vale (2009) e um *corpus* de comparação também distinto.

O que se verá das conclusões das análises de Raso e Vale (2009) é que a pesquisa identificou algumas funções pronominais que, como desempenham um papel fundamental nas estruturas de interface entre sintaxe e pragmática, indicam características interessantes da estruturação do enunciado no paralelo entre a língua falada pelos brasileiros, pelos italianos e pelos bilíngues sujeitos a erosão linguística. Além disto, os indícios de erosão linguística dos pronomes analisados permitiram uma discussão sobre a morfologia e seu uso nos vários grupos comparados.

O quadro geral das análises é o que resulta abaixo:

122 Veja-se seção 3.2.1.

Tabela 1 - Ocorrência dos clíticos nos *corpora* de italianos bilíngues e monolíngues.

Clíticos	Italianos bilíngues		Italianos monolíngues
	Absoluta	Variação percentual	
acusativo de terceira pessoa <i>lo, la, li, le l'</i>	166	(- 45,4)	304
<i>ci</i> atualizante e lexicalizante	120		117
<i>ci</i> locativo	24	(-38,4)	39
<i>ne</i> partitivo, locativo, argumental	14	(-52,8)	29
Total	324	(-33,8)	489

Os resultados mostram uma redução geral dos clíticos de cerca de 34%, de forma mais marcada para o *ne*, que, contudo, apresenta menor número de ocorrências em termos absolutos, e para o pronome acusativo de terceira pessoa. Há uma erosão consistente também do pronome *ci* com valor locativo, enquanto o *ci* atualizante e lexicalizante parece não sofrer erosão.

Vale (2007, p. 20) afirma que não é possível atestar com estes dados que a sequência de erosão, que confirmaria a hipótese jakobsoniana, é exatamente oposta àquela de aquisição, pois a ordem de alguns pronomes resulta inversa. Contudo a amostra analisada é relativamente pequena para se fazer tal comparação e a estudiosa adverte que, quanto a estes pronomes, somente em aparência não comparecem na ordem exatamente oposta.

Lembramos que quanto aos pronomes analisados, a ordem de aquisição de L1 e de L2 coincidem, sendo:

(1) *ci* atualizante e lexicalizante > *lo* flexionado > *ci* locativo > *ne* partitivo > *ne* argumental > *ne* locativo

A hipótese de Jakobson preveria portanto uma ordem de erosão especular a esta, ou seja a seguinte:

(2) *ne* locativo < *ne* argumental < *ne* partitivo < *ci* locativo < *lo* flexionado < *ci* atualizante e lexicalizante

Organizando os dados coletados por Raso e Vale (2009) e seu grau de erosão, e reunindo os pronomes *ne* em um único grupo, por sua exiguidade, teríamos a seguinte sequência de erosão:

- (3) *ne* (totais) < *lo* flexionado < *ci* locativo < *ci* atualizante e lexicalizante

Efetivamente, há uma inversão entre o *lo* flexionado e o *ci* locativo.

A partir deste quadro geral Raso e Vale (2009) decidiram isolar cada clítico objeto de estudo e analisar melhor a situação.

4.1.1 O pronome *ci* com valor locativo

Como foi visto na tabela 1 acima, o *ci* locativo apresenta uma redução de ocorrências de 38,4% na fala dos bilíngues em relação aos italianos nativos. É, com certeza, uma redução significativa e equiparável à redução total detectada nos clíticos em geral, mas bem inferior ao que era esperado, justamente porque a sequência de erosão previa que sua diminuição fosse maior em relação àquela do *lo* flexionado. Raso e Vale (2009) observam que, provavelmente, a redução do *ci* locativo poderia estar sendo subestimada. Isto seria devido à tipologia de textos utilizados no *corpus*: o tema das entrevistas aos bilíngues aborda em grande parte a viagem da Itália para o Brasil, as idas e vindas dos informantes, as comparações entre as duas nações, acabando por induzir uma produção maior de enunciados de tipo locativo, e portanto de sua retomada anafórica com o pronome locativo *ci*. Isto causaria um aumento no uso deste pronome em relação a outros tipos de texto.

4.1.2 O pronome *ne* em suas várias funções

O pronome *ne* compareceu em um número de ocorrências muito baixo em ambos os *corpora*, como se pode observar na tabela 2 abaixo:

Tabela 2 - Ocorrência do clítico *ne* em suas várias funções nos *corpora* de italianos bilíngues e monolíngues.

Clíticos	Italianos bilíngues		Italianos monolíngues
	Absoluta	Varição percentual	
<i>ne</i> partitivo	6	62,5	16
<i>ne</i> argumental	8	(-20)	10
<i>ne</i> locativo	0	(-100)	3
Total	14	(-52,8)	29

Apesar de apresentar uma diminuição consistente no número de ocorrências, à exceção do *ne* argumental, Vale (2007, p. 21) afirma que o número é pouco representativo da porcentagem de perda de cada categoria e portanto não é possível afirmar de maneira definitiva qual seria a sequência em função da perda (que lembramos seria: *ne* locativo < *ne* argumental < *ne* partitivo). Segundo ela, a hipótese da regressão não pode ser descartada, pois a erosão é incontestável e forte, mas a sequência de erosão nas diferentes funções não pode ser atestada com certeza com números tão diminutos.

4.1.3 O clítico *ci* em suas funções atualizante e lexicalizante

Examinaremos agora o clítico *ci* nos valores atualizante e lexicalizante. Como foi visto, o valor atualizante refere-se à ligação do clítico com os verbos *essere* (ser) e *avere* (ter), nas formas *esserci* e *averci*, e mantém ainda o estatuto de pronome, apesar de modificar ligeiramente o significado do verbo e estar ligado a ele. Nos lexicalizantes, os pronomes formam com o verbo uma unidade lexical nova, modificando o significado do verbo¹²³.

Observemos os dados obtidos na análise do *corpus* Raso-Vale e do *corpus* do LIP, ambos de 18080 palavras, apresentados pelos estudiosos:

Tabela 3: Ocorrência do clítico *ci* com valor lexicalizante nos *corpora* de italianos bilíngues e monolíngues.

Types lexicaux	Italianos bilíngues		Italianos monolíngues
	Absoluta	Variação percentual	
<i>Esserci</i>	107	(+ 87,75%)	57
<i>Averci</i>	10	(-80%)	50
Total	117		107

É evidente o enorme aumento das ocorrências da forma *esserci* nos italianos sujeitos a erosão, que quase dobra em relação aos italianos monolíngues. A frequência de *averci*, ao contrário, é extremamente baixa, apresentando uma diminuição de 80% em relação aos monolíngues. Quanto aos *types* que foram encontrados nos dois *corpora* comparados, pode-se observar a tabela 4 abaixo:

¹²³ Por exemplo o verbo *volere* significa "querer", mas quando é lexicalizado em *volerci*, assume o significado de "ser necessário". Veja-se a este respeito: Berretta (1991).

Tabela 4 - Ocorrências absolutas dos *types* lexicais e *tokens* do clítico *ci* com verbo.

<i>Types</i> lexicais	<i>Tokens</i> italianos bilíngues	<i>Tokens</i> italianos monolíngues
<i>Entrarci</i>	0	3
<i>Metterci</i>	1	1
<i>Guadagnarci</i>	0	1
<i>Volerci</i>	0	1
<i>Impiegarci</i>	0	3
<i>Rimanerci</i>	0	1
<i>Tenerci</i>	1	0
<i>Saperci fare</i>	1	0
TOTAL TYPES	3	6
Total tokens	3	10

A tabela mostra que, nos bilíngues, os *types* com *ci* lexicalizante são três, para um total de três *tokens*, enquanto nos monolíngues há seis ocorrências para um total de dez *tokens*, mostrando uma significativa diminuição seja de *types* lexicais seja de *tokens* nos italianos sujeitos à erosão, apesar de os números totais serem abaixo de valores estatisticamente significativos.

4.1.3.1 Análise dos resultados

A erosão de alguns pronomes era prevista pelos estudiosos, contudo os dados surpreenderam. Uma diminuição tão drástica de *averci* e um aumento tão sensível de *esserci* foram inesperados. Raso e Vale (2009) contudo, explicaram estas variações da seguinte forma:

A forma *averci*, no italiano falado pelos imigrantes cultos, no momento de sua partida da Itália, era estigmatizada, tornando-se aceitável na variedade padrão somente após esta época¹²⁴. O *corpus* de comparação fornece uma versão do italiano mais moderno em relação àquele falado pelos informantes bilíngues estudados. Isto explicaria uma diminuição tão radical de suas ocorrências.

O aumento da forma *esserci* é interpretado como devido à erosão lexical. Isto porque *esserci* pode ser utilizado como forma genérica para muitos verbos de estado, como nos exemplos citados por Raso e Vale (2009), em que o enunciado (1) pode substituir os enunciados (2)-(4):

(1) Alla riunione *c'erano* anche due americani.

124 Veja-se a este respeito Sabatini (1985).

- Na reunião, havia também dois americanos.
- (2) Alla riunione **parteciparono** anche due americani.
À reunião participaram também dois americanos.
- (3) Alla riunione **erano presenti** anche due americani.
Na reunião estavam presentes também dois americanos.
- (4) Alla riunione **furono invitati** anche due americani.
À reunião foram convidados também dois americanos.

O uso de *esserci* como estratégia de simplificação lexical pode ser confirmado, segundo Raso e Vale (2009), pelo fato de que *esserci* é usado pelos bilíngues sobretudo com valor apresentacional.

De fato, *esserci* pode apresentar dois tipos de função:

Função existencial: quando substitui outras formas lexicais que denotam diferentes formas de existência, como em:

- (5) *In classe c'erano cinque alunni.*
Na sala havia cinco alunos.

O enunciado (5) poderia ser substituído por outras formas como os exemplos (6) e (7):

- (6) *In classe erano presenti cinque alunni.*
Na sala estavam presentes cinco alunos.
- (7) *In classe studiavano cinque alunni.*
Na sala estudavam cinco alunos.

Função apresentacional: torna cognitivamente menos pesada uma estrutura completamente nova e remática; ao invés da forma (8), pode-se utilizar a forma (9), em que é apresentada primeiro a informação sobre a existência de um referente novo e depois, sobre essa informação que se torna de alguma maneira dada, a segunda informação, dividindo em duas micro-estruturas uma estrutura cognitivamente pesada:

- (8) *Un signore vuole parlare con te.*
Um senhor quer falar com você.
- (9) *C'è un signore che vuole parlare con te.*
Tem um senhor que quer falar com você.

A contagem da forma *esserci* com valor apresentacional mostra 15 ocorrências no *corpus* dos bilíngues e 11 no de monolíngues, confirmando, segundo Raso e Vale (2009), a hipótese da erosão lexical. Como veremos adiante, estes valores foram revistos recentemente por Vale (2009) e os números originais foram um pouco modificados.

4.1.4 Os acusativos de terceira pessoa

Uma primeira varredura dos *corpora* indicou, no que diz respeito aos acusativos de terceira pessoa (a dizer, as formas *lo, la li, le l'*) uma significativa diminuição no *corpus* dos bilíngues em relação àquele dos monolíngues, como se pode notar na tabela 5 abaixo:

Tabela 5 - Ocorrência dos clíticos acusativos de terceira pessoa nos *corpora* de italianos bilíngues e monolíngues.

Clíticos	Italianos bilíngues		Italianos monolíngues
	Absoluta	Varição percentual	
Clíticos acusativos de terceira pessoa (<i>lo, la li, le, l'</i>)	166	(-45,4)	304

A tabela mostra que o número total de ocorrências encontrado no *corpus* de italianos bilíngues é de 166, enquanto no *corpus* de monolíngues há uma presença de 304 ocorrências, para um total de 18080 palavras, nos dois *corpora* comparados. À primeira vista, portanto, a hipótese de que haveria uma perda dos clíticos na fala dos italianos sujeitos a erosão linguística é sustentada por estes números absolutos. Inclusive, a diminuição é bastante significativa, visto que, quando observados os números em percentual, a erosão representa uma perda de 45,4% em relação ao *corpus* de comparação.

Os autores, contudo, quiseram investigar mais a fundo estes dados e desagregaram os clíticos acusativos de terceira pessoa em base às diferentes funções que podem ocupar no enunciado, ou seja, quando desenvolvem uma função anafórica ou catafórica de um constituinte deslocado.

A tabela 6 abaixo explicita estes dados:

Tabela 6: Ocorrência dos clíticos acusativos de terceira pessoa nos *corpora* de italianos bilíngues e monolíngues total e em diferentes posições.

Clíticos	Italianos bilíngues		Italianos monolíngues
	Absoluta	Varição percentual	
Clíticos acusativos de terceira pessoa (<i>lo, la li, le, l'</i>)	166	(-45,4)	304
Em retomada de deslocamentos à esquerda	29	25	23
Em catáfora de deslocamentos à direita	13	(-54)	28

Quando foram contadas somente as ocorrências de pronomes que retomam constituintes deslocados à esquerda, portanto em função anafórica, as ocorrências dos bilíngues chegam a ser superiores em 25% àquelas dos italianos monolíngues. Isto quer dizer que, além da perda média, nesta posição, os clíticos analisados apresentam números absolutos superiores àqueles dos monolíngues. Raso e Vale (2009) notam portanto que isto significa que a perda dos clíticos acusativos, excluindo-se os casos de deslocamento à esquerda, deve ser

ainda maior dos 45,4% detectados na contagem total, e que acontece um comportamento diferente quando há deslocamento à esquerda, sendo necessário explicar esta característica.

Entretanto, quando o pronome se caracteriza por uma função catafórica de constituinte deslocado à direita, a perda é ainda maior que a média, sendo de 54% em relação aos monolíngues.

Outra descoberta de Raso e Vale (2009) refere-se ao fato de que, apesar do aumento dos pronomes acusativos com função anafórica de constituintes deslocados, pode no entanto ocorrer erosão pronominal nesta mesma função. De fato, uma varredura do inteiro *corpus* de São Paulo, de cerca de 54000 palavras, retornou 12 casos de omissão do pronome com função anafórica que apresentam deslocamento. Citamos abaixo um exemplo fornecido por Raso e Vale (2009, p. 11): o enunciado (10) é agramatical, faltando a retomada pronominal anafórica; o exemplo (11) mostra como seria a forma correta.

(10) **Io mi sono resa conto che l'italiano lui poteva mantenere in casa.*

Eu percebi que o italiano ele podia manter em casa.

(11) *Io mi sono resa conto che l'italiano lui **lo** poteva mantenere in casa.*

Eu percebi que o italiano ele **o** poderia manter em casa.

A seguir, reportaremos as explicações dos pesquisadores para estes fenômenos aparentemente anômalos.

4.1.4.1 Análise dos resultados

4.1.4.1.1 As tematizações

Para interpretar um aumento tão alto de clíticos em deslocamento à esquerda, apesar de terem sido documentadas no *corpus* perdas do mesmo pronome anafórico, Raso e Vale (2009) optaram por verificar a estruturação do enunciado nas duas línguas em contato e observar a predileção de cada língua quando mais soluções estão disponíveis para a mesma função. Como a análise referia-se ao pronome anafórico deslocado à esquerda, os pesquisadores quiseram verificar as possíveis estratégias de tematização de constituintes postos à esquerda no enunciado.

Por *tematização* entende-se a sinalização de um elemento como tema. O *tema* é o elemento sobre o que se diz algo, em contraposição ao *rema*, que é aquilo que é dito a respeito

do tema. Tematizar um elemento equivale portanto a destacar um constituinte temático sobre o qual será construída a ilocução, ou seja, o âmbito semântico dessa referência.

Em italiano, as tematizações de constituintes colocados à esquerda do enunciado podem acontecer das seguintes formas¹²⁵:

Deslocamentos à esquerda:

- (12) *Mario, lo incontro domani.*
Mario, o encontro amanhã.

A estrutura canônica desta frase seria: *Incontro Mario domani*/Encontro Mario amanhã. No deslocamento à esquerda, o elemento "Mario" é deslocado de sua posição normal para uma posição marcada que o põe em evidência, tornando-o o tema do enunciado. Neste caso o elemento deslocado deve ser retomado obrigatoriamente por um pronome anafórico, isto é: *lo/o*;

Tema suspenso:

- (13) *Mario, gli parlo più tardi.*
Mario, lhe falo mais tarde

Nesta construção, o elemento tematizado, "Mario", é posto no começo do enunciado como componente autônomo, que não possui marca de caso mas continua sendo ligado pelo pronome ao resto da frase: é uma estrutura parecida com o deslocamento à esquerda mas não há marca de caso como seria se fosse:

- (14) *A Mario, gli parlo più tardi.*
Ao Mario, lhe falo mais tarde.

Anacoluto¹²⁶:

- (15) *Il regalo, non ho ancora pensato a niente.*
O presente, ainda não pensei em nada.

Neste caso, falta qualquer ligação sintática entre o constituinte tematizado e o resto do enunciado. A ligação é puramente funcional e veiculada prosodicamente.

Deve-se ter em mente que, em italiano, quando o constituinte deslocado é um objeto direto, não é permitida a omissão do pronome anafórico. Portanto, o exemplo (14) seria agramatical caso fosse produzido como em (16), em que o primeiro constituinte tem valor temático, e não remático:

125 Exemplos nossos.

126 A vírgula indica quebra prosódica, que nos primeiros dois casos pode não ocorrer mas no terceiro é obrigatória.

(16) **Mario, incontro domani.*

Mario, encontro amanhã.

No PB a predileção é por outros tipos de estrutura de tematização. Podem ocorrer três possibilidades, exemplificadas abaixo¹²⁷:

(16) João, vejo amanhã.

(17) João, vejo ele amanhã.

(18) João, o vejo amanhã.

As estruturas mais utilizadas são a (16), sem retomada pronominal, e a (17), com o uso do pronome tônico. A (18), com o clítico anafórico, que seria aquela mais utilizada no It, em PB é a menos usada, pertencendo a um registro alto e quase exclusivo da escrita formal.

Há, portanto, diferenças substanciais em relação a como as duas línguas se comportam nas tematizações. Os pesquisadores quiseram verificar mais a fundo este fator, para poder entender o que ocorre no contato do PB com o It.

4.1.4.1.2 As topicalizações temáticas nos três grupos de falantes

Raso e Vale (2009) confirmaram que no PB há uma forte presença de tematizações e a quase totalidade delas são topicalizações¹²⁸. Este fato foi verificado na análise comparativa de um novo *corpus* por eles montado de 33.000 palavras. O *corpus* era sub-dividido em três *subcorpora*, cada um de 11.000 palavras, um para o italiano de monolíngues, extraído do *BADIP*, outro de italianos bilíngues, extraído do *corpus* Raso-Vale, e um de PB, extraído do *NURC*¹²⁹. Abaixo podem-se observar os resultados obtidos:

127Exemplos nossos.

128 As *topicalizações* são um tipo de tematização em que o elemento tematizado é deslocado no começo do enunciado sem retomada pronominal e com marca prosódica.

129 UNIVERSIDADE FEDERAL DO RIO DE JANEIRO. Faculdade de Letras. *Projeto norma linguística urbana culta - RJ*: banco de dados. Disponível em: <<http://www.lettras.ufjf.br/nurc-rj/>>. Acesso em: 21 mar. 2007.

Tabela 7 – Ocorrência (normalizada para 11.000 palavras) dos diferentes tipos de tematização nos *corpora* de monolíngues italianos, monolíngues brasileiros e italianos bilíngues.

Tipo de tematização	Italianos monolíngues	Italianos bilíngues	Brasileiros monolíngues
Deslocamentos à esquerda.....	12	17	0
Temas suspensos.....	1	2	4
Topicalizações temáticas.....	2	7	38
Total	15	26	42

Como se pode notar, o PB mostra uma altíssima frequência de tematizações, quase tripla em relação ao italiano. Os bilíngues ocupam uma posição intermediária. Contudo, o que é mais interessante, segundo os pesquisadores, é a análise das diferentes estratégias utilizadas pelos falantes para tematizar um constituinte à esquerda.

No PB, a absoluta maioria de tematizações é composta por topicalizações ou seja, estruturas em que se perde completamente a relação sintática entre o elemento tematizado e o resto do enunciado, e é mantida somente uma relação de tipo funcional. Comparecem alguns casos de tema suspenso, mas nunca ocorre o pronome anafórico do constituinte deslocado, ou seja, o deslocamento à esquerda. Abaixo, a título de exemplo, reportamos algumas topicalizações temáticas citadas por Raso e Vale (2009, p. 14):

- (19) As freiras, a gente morria de rir.
- (20) O resto do comércio, havia loja de móveis, havia farmácia.
- (21) Corrida de cavalo, eu nunca fui ao Jockey Club.
- (22) Porque um primário bem feito, tudo fica mais fácil, não?

No caso dos italianos monolíngues, ao contrário, quase todas as tematizações compõem-se por deslocamentos que, como foi visto, nunca são usados pelos monolíngues brasileiros. Em geral, "pode-se dizer que o italiano prefere marcar sintaticamente a relação entre constituinte deslocado e resto do enunciado" (RASO; VALE, 2009, p. 14).

No caso dos bilíngues, as ocorrências indicam que eles ocupam uma posição intermediária entre os italianos monolíngues e os monolíngues brasileiros. Contudo, Raso e Vale (2009) fazem notar que, apesar de haver um certo aumento das topicalizações e uma tendência ao enfraquecimento das relações sintáticas, a estratégia preferida permanece o deslocamento.

As conclusões de Raso e Vale (2009) a respeito do aparente aumento dos pronomes

clíticos objeto em posição anafórica do constituinte deslocado são de que, na realidade, não há um aumento dos clíticos nos deslocamentos, mas um aumento expressivo dos deslocamentos (e portanto dos clíticos), causado por um aumento geral das estratégias de tematização. Este fato faz com que o número absoluto dos clíticos em deslocamento acabe sendo superior àquele dos italianos monolíngues.

4.1.4.1.3 Os deslocamentos à direita

A segunda questão investigada por Raso (2009) a respeito dos clíticos acusativos de terceira pessoa refere-se à diminuição dos casos de deslocamentos à direita, superior até mesmo à redução geral dos pronomes. A explicação reside, segundo o autor, mais uma vez, no comportamento dos clíticos dependendo de sua posição.

Em italiano, pode ocorrer deslocamento à direita com ou sem quebra prosódica antes do constituinte deslocado, como nos exemplos abaixo¹³⁰:

(23) *L'ho mangiata / la mela.*

A comi / a maçã.

(24) *L'ho mangiata la mela.*

A comi a maçã.

O PB, ao contrário, sinaliza o deslocamento à direita somente através da quebra prosódica, pois não há catáfora pronominal e a quebra prosódica torna-se a única marca que possa distinguir a estrutura deslocada daquela canônica. Observe-se o exemplo abaixo:

(25) *Eu comi / a maçã.*

Raso (2009) aproveita do exemplo com o verbo *mangiare* (comer) para discutir sobre a estrutura argumental de alguns verbos e sua importância na questão dos deslocamentos à direita. O pesquisador explica que verbos com mais estruturas argumentais, como *mangiare*, podem ser caracterizados por uma possibilidade monoargumental, como em (26) e uma a dois argumentos como em (27):

(26) *Hai mangiato o non hai avuto tempo?*

130 Exemplos de nossa autoria.

Você comeu ou não teve tempo?

(27) *Hai mangiato il frutto esotico che ti ho portato?*

Você comeu a fruta exótica que eu te trouxe?

Quando o verbo é usado com a estrutura a dois argumentos, "parece que o elemento deslocado deve ser necessariamente precedido do pronome e não pode ser marcado apenas através da quebra prosódica. Não parece aceitável encontrar quem nos tenha trazido a fruta exótica e pronunciar (28), ao passo que é perfeitamente aceitável (29)"¹³¹ (RASO 2009, p. 390, tradução nossa):

(28) **A proposito, poi ho mangiato / il frutto esotico.*

Falando nisso, eu acabei comendo / a fruta exótica.

(29) *A proposito, poi l'ho mangiato / il frutto esotico.*

Falando nisso, eu acabei a comendo / a fruta exótica.

Quando, ao contrário, a estrutura utilizada é monoargumental é plausível uma resposta como:

(30) *Ho mangiato / il pollo.*

Eu comi / o frango

A não ser nesse caso, não seria possível deslocar à direita um constituinte com valor temático sem a catáfora pronominal. No entanto Raso (2009) vai além, afirmando que o deslocamento à direita pode ser confundido com o eco marcado prosodicamente. Verbos como *attraversare* (atravessar), que possui somente a estrutura a dois argumentos, pode ser empregado em um enunciado como (31), permitindo-o o contexto, como no diálogo (32) reportado por Raso (2009, p.391):

(31) *Ha attraversato / la strada.*

Ele atravessou / a rua.

(32) *A: Per favore, avvisami appena Mario ha attraversato la strada.*

B: Ecco, ha attraversato / la strada

A: Por favor, me avise logo que Mario atravessar a rua.

B: Pronto, ele atravessou / a rua.

131 "*sembrebbè che l'elemento dislocato debba essere necessariamente preceduto dal pronome e non possa essere marcato solo attraverso il break prosodico. Non sembra accettabile incontrare chi ci ha portato il frutto esotico e profferire (39), mentre è perfettamente accettabile (40)*".

É evidente que no deslocamento à direita o falante deve considerar vários fatores: a semelhança com o eco que pode confundi-lo, a possibilidade de haver ou não quebra prosódica e a alternativa da catáfora pronominal, enquanto o deslocamento à esquerda admite somente uma opção: a retomada anafórica pronominal.

Raso (2009) fornece uma explicação para a diferença nas ocorrências dos clíticos dependendo da posição e da função que ocupam no enunciado, ou seja, se são catafóricos ou anafóricos de constituintes deslocados ou não. Como no PB o clítico não é expresso em ambos os casos, o papel da interferência é igual. No entanto, os bilíngues italianos parecem seguir o modelo do PB quando o pronome é catafórico, mas quase nunca quando o clítico é anafórico e o constituinte deslocado está à esquerda.

É fato que o deslocamento à esquerda sem retomada pronominal em italiano é agramatical, mas na forma do deslocamento à direita a necessidade do pronome depende da estrutura argumental do verbo, já que a semelhança entre deslocamento e eco é muito grande.

Portanto, o bilíngue pode ter a impressão de que o modelo do PB possa servir também para o It, como de fato acontece no caso do eco. Isto, segundo Raso (2009), explicaria a alta perda do pronome no caso em que "a omissão ou não compromete a aceitabilidade do enunciado ou gera estruturas cuja aceitabilidade ou menos deve ser avaliada no contexto individual, e não, como no caso do deslocamento à esquerda, com base na estrutura em si"¹³² (RASO, 2009, p. 391).

4.1.5 Conclusões de Raso e Vale (2009)

Raso e Vale (2009) constataram em seus trabalhos que há um enfraquecimento morfossintático no italiano falado pelos bilíngues, interpretado como sinal de erosão linguística. Contudo, o que eles consideram mais relevante é que a reestruturação morfossintática é ligada a uma reestruturação do enunciado, que leva na direção de uma estruturação mais pragmática. Isto permite compreender os deslocamentos com retomada anafórica e catafórica que foram discutidos nos parágrafos acima. O que os estudiosos questionam é o peso da interferência do PB. Se por um lado o aumento das estruturas tematizantes reproduziriam o modelo do PB, por outro há de se perguntar até que ponto ao contrário não revelam estratégias típicas das primeiras fases de aquisição, e da própria erosão,

132 *"l'omissione o non compromette l'accettabilità dell'enunciato oppure genera strutture la cui accettabilità o meno deve essere valutata nel singolo contesto, e non è, come nel caso della dislocazione a sinistra, dipendente tout court dalla struttura in sé".*

na qual a organização com base pragmática precede aquela com base sintática. A conclusão final dos pesquisadores é que é necessário ampliar os dados analisados para se ter um quadro mais completo. É o que se propôs o presente trabalho e que será explicitado nas próximas páginas.

4.2 Os resultados obtidos no presente estudo

Serão agora apresentados os dados obtidos na análise do *corpus* Raso-Ferrari, coletado durante o período 2008-2009 e composto por 21.298 palavras. Esclarecemos desde já que este *corpus* novo foi comparado com um *corpus* diferente daquele de referência do *corpus* Raso-Vale. Portanto, espera-se encontrar divergências nos resultados, não tanto, como será visto, pela diferença entre os *corpora* de bilíngues, mas pelas diferenças entre os *corpora* de comparação.

Por tratar-se de uma pesquisa comparativa, os clíticos analisados são os mesmos estudados por Raso e Vale (2009), ou seja, o pronome *ci* em seus valores atualizante, lexicalizante e locativo, os pronomes *ne* em suas funções partitiva, argumental e locativa e os pronomes acusativos de terceira pessoa, a dizer: *lo, la, li, le, l'*. A tabela 8 abaixo mostra as ocorrências contadas no *corpus* para cada clítico objeto de estudo.

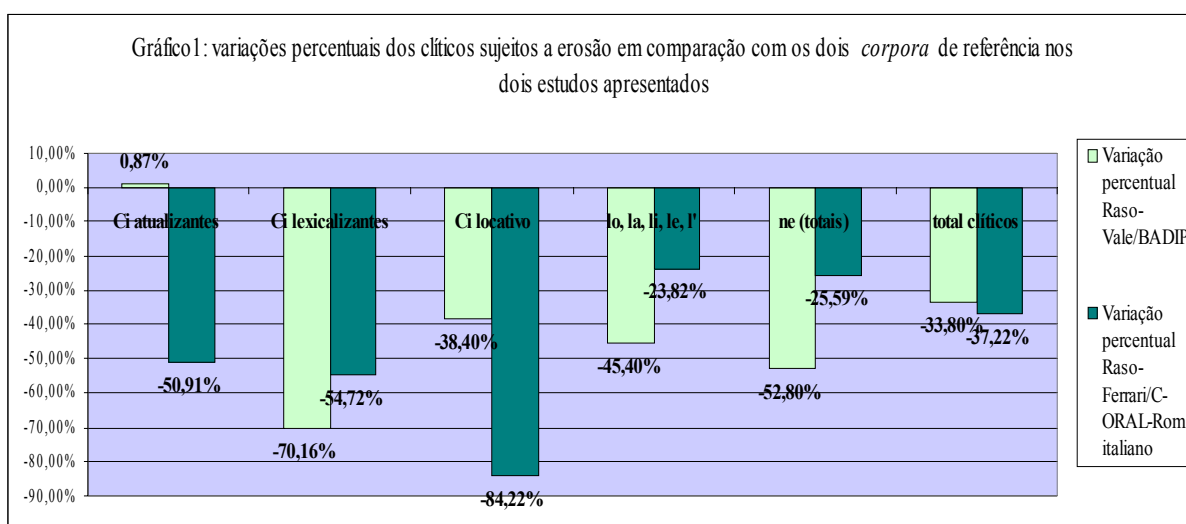
O *corpus* de comparação foi extraído, como dito anteriormente, do *C-ORAL-ROM italiano*, por um total de 21.224 palavras. Não havendo correspondência precisa do total das palavras dos dois *corpora*, foi necessário normalizar os dados ou seja, as ocorrências foram divididas pelo número de palavras do *corpus* a que pertenciam, e depois multiplicadas por 10.000 para se obter um número minimamente relevante. Os dados normalizados são evidenciados na tabela para que o cotejo seja mais prático:

Tabela 8: Ocorrência dos clíticos nos *corpora* de italianos bilíngues e monolíngues.

Clíticos	Corpus Raso-Ferrari		C-ORAL-ROM It		Variação percentual
	Absoluta	Normalizada	Absoluta	Normalizada	
<i>Ci</i> atualizantes	135	63,38	274	129,09	(-50,91)
<i>Ci</i> lexicalizantes	10	4,69	22	10,36	(-54,72)
<i>Ci</i> locativo	3	1,4	19	8,95	(-84,22)
<i>lo, la, li, le, l'</i>	237	111,21	310	146,06	(-23,82)
<i>Ne (totais)</i>	32	15,02	43	20,26	(-25,59)
TOTAL	407	191,09	646	304,37	(-37,22)

A tabela acima mostra uma erosão linguística de todos os clíticos objeto de estudo. A redução das ocorrências é particularmente evidente no caso do *ci* locativo, que apresenta uma diminuição de 84,22% e do *ci* nos valores atualizante e lexicalizante, com uma perda de 50,91% por parte dos bilíngues em relação ao *corpus* de monolíngues. A variação total dos clíticos utilizados na fala dos informantes sujeitos a erosão em relação àqueles monolíngues é de -37,22%, ou seja, uma perda até maior em relação àquela detectada por Raso e Vale (2009).

Vejamos o gráfico que compara as variações percentuais entre os dois estudos:



Como é possível observar nesta figura, separando o clítico *ci* atualizante daquele lexicalizante, obtemos resultados diferentes em ambos os estudos. No *corpus* Raso-Vale, que foi comparado com o *BADIP*, o *ci* atualizante apresenta um não significativo aumento de 0,87%, enquanto com o *ci* lexicalizante há uma diminuição drástica de -71,16%. No *corpus* Raso-Ferrari, que foi comparado com o *C-ORAL-ROM italiano*, o *ci* atualizante apresentou uma considerável diminuição de 50,91% e o *ci* lexicalizante uma perda de 54,72%. A diminuição das ocorrências do *ci* locativo também é mais do que o dobro maior em relação aos resultados do estudo de Raso e Vale, mas como se verá adiante isto era esperado. A perda total dos clíticos, utilizando um *corpus* de comparação diferente mostrou-se superior às análises anteriores: enquanto os estudos de Raso e Vale tinham advertido uma redução de 33,8% no uso dos clíticos na fala dos informantes sujeitos a erosão, no presente estudo a redução total é de 37,22%. Entretanto, no caso dos acusativos de terceira pessoa e do pronome *ne* em sua várias funções, houve uma redução bem inferior àquela detectada por Raso e Vale:

enquanto nesta pesquisa a redução dos clíticos acusativos de terceira pessoa soma 25,59%, Raso e Vale acharam uma redução de 45,4%. No caso do *ne* nas várias funções, Raso e Vale detectaram uma diminuição de 52,8%, praticamente o dobro daquela, de 25,59%, revelada no presente estudo. Contudo, há de se lembrar que os números absolutos das várias funções de *ne* eram extremamente exíguos, portanto os dados percentuais devem ser lidos com uma certa cautela.

Nos parágrafos a seguir serão separados os vários clíticos e cada um será analisado isoladamente, para em seguida retomar o estudo contrastivo com os resultados obtidos por Raso e Vale (2009).

4.2.1 O pronome *ci* com valor locativo

O pronome *ci* com valor locativo foi aquele, dentre os estudados, que sofreu o maior grau de erosão linguística, com uma perda de ocorrências de 84,22% em relação ao *corpus* de comparação *C-ORAL-ROM italiano*. Contudo, uma diminuição tão grande era já esperada. Mesmo Raso e Vale (2009) afirmaram que a perda por eles detectada de 38,4% poderia ter sido subestimada devido aos temas abordados nas entrevistas, que estimulavam o uso de enunciados de tipo locativo e portanto a conseguinte retomada destes através do uso do clítico.

O *corpus* agora apresentado possui um grau de informalidade¹³³ muito mais alto, e mesmo os monólogos presentes não se configuram como entrevistas: os informantes não são guiados pelo pesquisador, discorrem sobre vários temas, e não há nenhuma indução ao uso de enunciados de tipo locativo. Acreditamos que a perda deste pronome tenha sido tão forte por este motivo.

4.2.2 O pronome *ne* em suas várias funções

Quanto à variação do pronome *ne*, a análise do *corpus* Raso-Ferrari destacou uma diminuição do número de suas ocorrências de cerca de 25% em relação ao *C-ORAL-ROM italiano*. Desmembrando os dados pode-se observar a tabela abaixo; lembramos que os dados foram normalizados por 10.000 e são evidenciados na tabela para que o cotejo seja mais prático:

133 Veja-se a respeito das características das situações comunicativas a seção 3.3.2.

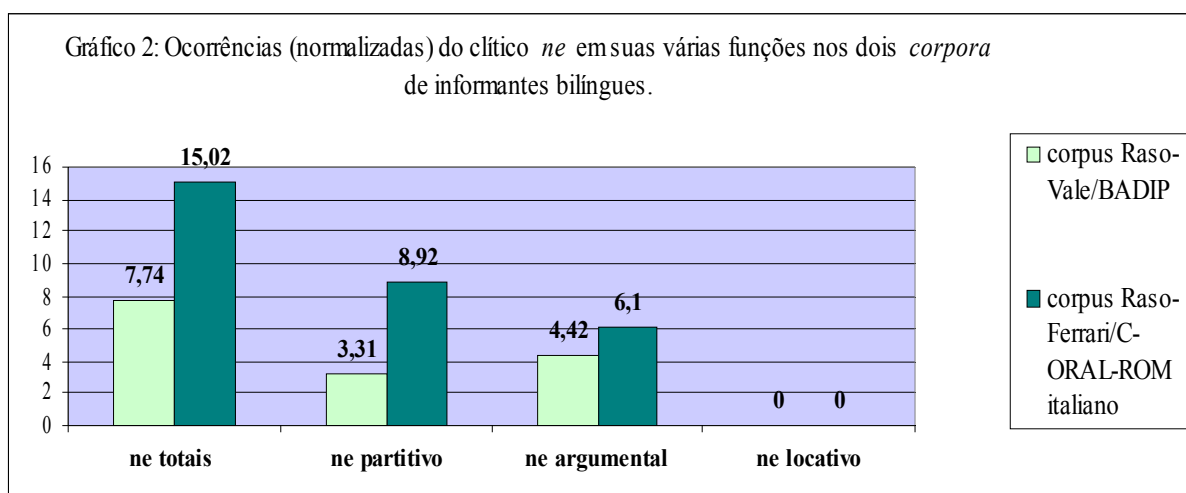
Tabela 9: Ocorrência do clítico *ne* em suas várias funções nos *corpora* de italianos bilíngues e monolíngues.

Clítico	Corpus Raso-Ferrari		C-ORAL-ROM Italiano		Variação percentual
	Absoluta	Normalizada	Absoluta	Normalizada	
<i>Ne</i> partitivo	19	8,92	28	13,19	(-32,15)
<i>Ne</i> argumental	13	6,1	15	7,06	(-13,34)
<i>Ne</i> locativo	0	0	0	0	0
Total	32	15,02	43	20,26	(-25,59)

Pode-se observar que a variação maior, em termos de perda, se há com o *ne* com função partitiva, apresentando este uma diferença de -32,15%. O *ne* argumental não mostra uma diminuição tão radical, com uma perda de -13,34%. Finalmente, não foram encontradas ocorrências do *ne* locativo nem no *corpus* Raso-Ferrari, nem naquele extraído do *C-ORAL-ROM italiano*, portanto não há mudança entre o grupo de controle e os falantes sujeitos a erosão.

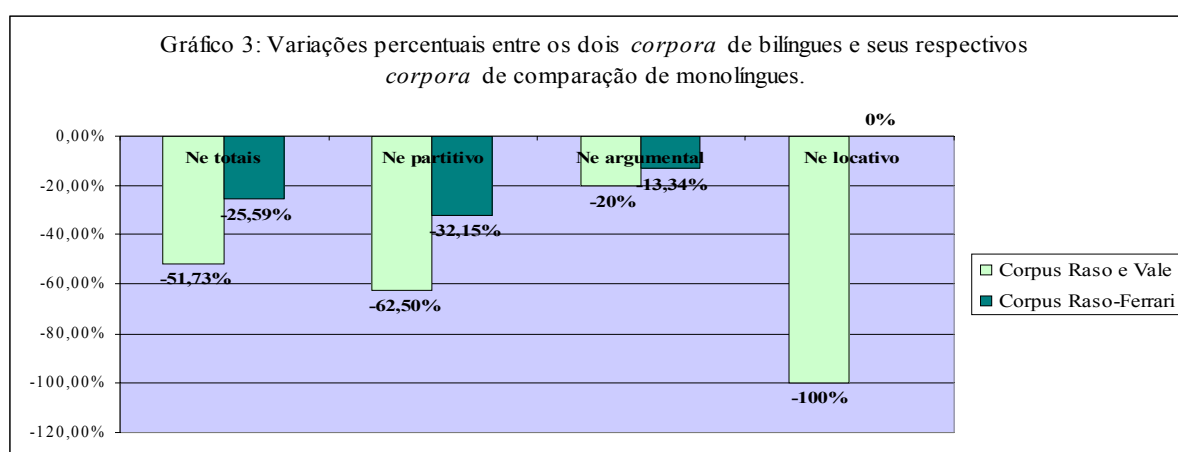
4.2.2.1 Análise dos resultados do clítico *ne*

Para poder compreender melhor estes dados, comparamo-os com os resultados obtidos nos estudos de Raso e Vale (2009). Lembramos que os pesquisadores consideraram os dados relativos ao clítico *ne* em suas várias funções pouco representativos por sua exiguidade. Entretanto, achamos interessante compará-los para poder formular algumas hipóteses. O gráfico abaixo mostra os números absolutos encontrados nos dois *corpora* de italianos bilíngues.



Em números absolutos, Raso e Vale (2009) computaram 14 clíticos *ne* no total, enquanto nossa contagem, com um *corpus* de bilíngues diferente e um pouco maior, fez 32 ocorrências. Se normalizarmos os dados, teremos cerca de 8 clíticos *ne* totais para cada 10.000 palavras no *corpus* Raso-Vale, contra 15 no *corpus* Raso-Ferrari, ou seja, praticamente o dobro de ocorrências. Estes números indicariam uma erosão linguística menor entre os informantes do novo *corpus*. Esta tendência é confirmada no caso do *ne* argumental, com 3,31 ocorrências a cada 10.000 palavras no *corpus* Raso-Vale e 8,92 no *corpus* Raso-Ferrari. Se forem observados os números que comparecem do *ne* argumental no entanto a diferença entre os dois *corpora* não é tão grande, sendo de 4,42 a cada 10.000 palavras no caso do *corpus* Raso-Vale e de 6,1 no *corpus* Raso-Ferrari. Ambos os *corpora* não apresentam nenhuma ocorrência do clítico *ne* com valor locativo. Contudo, como foi dito acima, o próprio *C-ORAL-ROM italiano* não apresenta ocorrências deste pronome. A hipótese mais convincente é que o *ne* locativo é extremamente pouco usado mesmo entre os monolíngues, e isto pode ser comprovado pela lexicalização de muitos verbos através deste pronome¹³⁴. Naturalmente, para poder comprovar esta hipótese, seria necessário mais uma vez um *corpus* maior para uma análise mais detalhada.

Observemos agora os mesmos dados vistos sob outra ótica ou seja, nas variações de porcentagem que acontecem entre os *corpora* de bilíngues e seus respectivos *corpora* de comparação de monolíngues:



134 Em Russi (2008, p. 62), lê-se a este respeito: *Locative ne [...] does not have a widespread use in contemporary Italian. It occurs almost exclusively in combination with the clitic si, in verbs such as andarsene 'go away, leave' and venirsene 'come away', venirne fuori 'come out' where [...] ne must be considered a fully grammaticalized element.* (O Locativo *ne* [...] não é muito difundido no italiano contemporâneo. Comparece quase exclusivamente em combinação com o clítico *si*, em verbos como *andarsene* "ir embora" e *venirsene* "sair", *venirne fuori* "sair dessa" onde [...] o *ne* deve ser considerado um elemento totalmente gramaticalizado).

As variações percentuais confirmam o que os números absolutos tinham mostrado de forma visualmente mais clara. No caso do clítico *ne* em todas as funções estudadas, as análises de Raso e Vale (2009) detectaram uma redução de mais de 50%, na comparação com o *corpus* de monolíngues *BADIP*, à medida que na presente pesquisa a perda é da metade, tendo sido utilizado como *corpus* de referência o *C-ORAL-ROM italiano*, comprovando o que foi sugerido acima, de que a erosão linguística do grupo de italianos bilíngues deste nosso estudo é inferior àquela percebida nos estudos de Raso e Vale (2009).

Quanto aos pronomes analisados em separado: o *ne* partitivo apresenta uma variação de -62,5% no estudo de Raso e Vale (2009) e de -32,15% naquele aqui apresentado, e o *ne* argumental mostra uma diminuição de 20% no *corpus* Raso-Vale e de 13,34% no *corpus* Raso-Ferrari. Estes dados corroboram a tese de que a erosão em ato nos bilíngues do *corpus* aqui montado é inferior àquela apresentada nos estudos anteriores.

Finalmente, há o caso do *ne* locativo: no *corpus* de Raso-Vale a perda é de 100%, mas é importante lembrar que no *corpus* por eles analisados não constava nenhuma ocorrência deste pronome, e no *corpus* de comparação do *BADIP*, havia somente três ocorrências em total, número muito exíguo para que se possa elaborar qualquer hipótese. Por outro lado, o fato que no *BADIP* o *ne* compareça, mesmo em números absolutos muito baixos, enquanto isto não acontece no *C-ORAL-ROM italiano*, pode ser considerado como mais uma evidência sobre a diferença entre os dois *corpora* de comparação, depondo sobre a maior idade do *BADIP* e sobre sua menor coloquialidade e espontaneidade.

A escala de erosão, retirando-se os locativos que como vimos não podem ser considerados, não é confirmada, pois deveria ser *ne* argumental < *ne* partitivo e neste caso mostrou-se oposta: *ne* partitivo < *ne* argumental nos dois estudos.

4.2.3 O pronome *ci* com valor atualizante e lexicalizante

Vejamos agora as variações do pronome clítico *ci* com valor atualizante em união com o verbo *essere*, subdividido entre o *esserci* com valor existencial e aquele com valor apresentacional, no *corpus* Raso-Ferrari em comparação com o *C-ORAL-ROM italiano*, em números absolutos e com os dados normalizados para cada 10.000 palavras:

Tabela 10 - Ocorrência do clítico atualizante *ci*+verbo (*esserci*) em suas funções existencial e apresentacional nos *corpora* de italianos bilíngues e monolíngues

Clítico	<i>Corpus Raso-Ferrari</i>		C-ORAL-ROM Italiano		Variação percentual
	Absoluta	Normalizada	Absoluta	Normalizada	
<i>Esserci</i> existencial	103	48,43	167	78,67	(-38,43)
<i>Esserci</i> apresentacional	16	7,51	16	7,53	(-0,27)
Total	119	55,87	183	86,22 66,9*	(-35,21) (-16,48)*

O asterisco* refere-se aos dados fornecidos por Panunzi (comunicação oral) sobre o corpus inteiro C-ORAL-ROM italiano de 300.000 palavras.

O total de ocorrências em números absolutos do *esserci* é de 119 no *corpus* Raso-Ferrari. Com os dados atualizados obtemos 55,87 ocorrências a cada 10.000 palavras contra 86,22 no *corpus* de comparação. O *esserci* com função existencial é muito superior àquele com função apresentacional nos dois *corpora* comparados: no *corpus* Raso-Ferrari há 48,43 ocorrências a cada 10.000 palavras de *esserci* existencial, e 7,51 do *esserci* apresentacional, enquanto no C-ORAL-ROM italiano foram encontradas 78,67 ocorrências do primeiro e 7,53 do segundo.

No total de *esserci* há evidências de erosão linguística: no *esserci* com função existencial a diminuição das ocorrências é de 38,43%, pouco acima dos 35,21% dos *esserci* em total. Não é detectada perda do *esserci* apresentacional.

Panunzi (2010) fez um estudo específico sobre o verbo *esserci* analisado a partir de todo o C-ORAL-ROM italiano. Dados fornecidos pelo autor¹³⁵ informam que a normalização de *esserci* a cada 10.000 palavras representa 66,9¹³⁶, ocorrências contra aquelas por nós calculadas de 86,22. Isto confirma a necessidade de estudos sobre *corpora* maiores (o *corpus* analisado por Panunzi soma cerca de 300.000 palavras), pois a variação percentual mostrou-se bem diferente, subindo de -35,21% quando confrontamos nosso *corpus* com aquele de 21.224 extraído do C-ORAL-ROM italiano, para -16,48% quando a comparação foi feita com o C-ORAL-ROM italiano inteiro, de 300.000 palavras.

Quanto às ocorrências do clítico atualizante *ci* com o verbo *avere* os resultados foram bastante diferentes. A tabela 11 abaixo mostra estes resultados em detalhes:

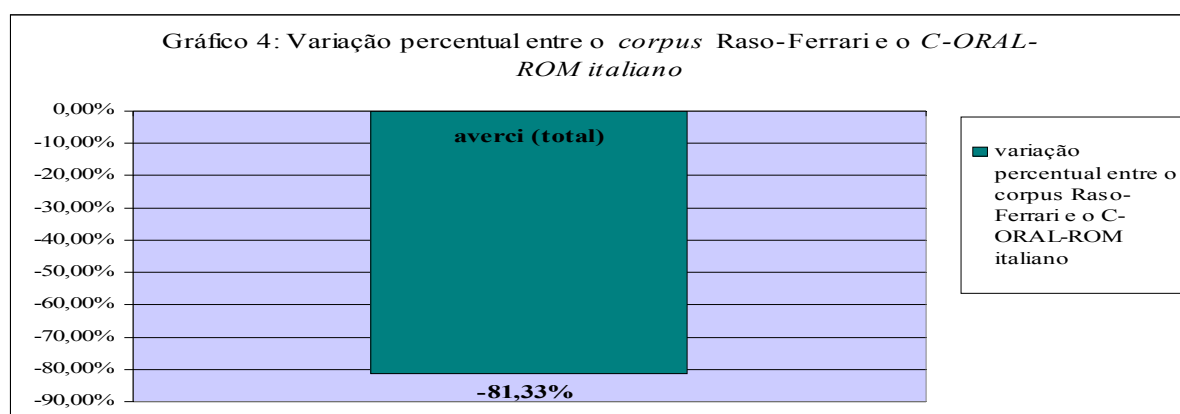
135 Informação verbal.

136 Evidenciadas na tabela com um asterisco.

Tabela 11 - Ocorrência do clítico *ci* atualizante com o verbo *avere* nos *corpora* de italianos bilíngues e monolíngues

Clítico	<i>Corpus Raso-Ferrari</i>		C-ORAL-ROM Italiano		Variação percentual
	Absoluta	Normalizada	Absoluta	Normalizada	
<i>Averci</i>	16	7,51	91	40,21	(-81,33)

É claro, pelos dados, que a diminuição de *averci* é muito maior do que aquela de *esserci*; as ocorrências de *averci* apresentam uma diminuição extremamente sensível: os dados normalizados computam 7,51 ocorrências no *corpus* dos bilíngues e 40,21 naquele dos monolíngues. Vejam-se no gráfico abaixo as diferenças em termos percentuais:



A perda total de *averci* é de 81,33%. Uma possível explicação para estas diminuições pode ser devida ao fato de que o *corpus* de comparação, o C-ORAL-ROM italiano, é muito mais coloquial do que nosso *corpus*. Isto explicaria a grande quantidade de *esserci* e de *averci*. A confirmação deste fato pode ser observada no fato que o *esserci* apresentacional, ao contrário dos outros, praticamente não apresenta variação.

4.2.3.1 Análise contrastiva dos resultados

Serão agora comparados os resultados acima apresentados com aqueles obtidos nos estudos de Raso e Vale. Em um trabalho recente, Vale (2009)¹³⁷ revisou os dados do verbo *esserci* nas funções existenciais e apresentacionais. Os novos resultados (coluna “2009” da tabela 12) são evidenciados em comparação com aqueles anteriormente divulgados (coluna “2007”):

137 Apresentação oral: *A erosão dos clíticos verificada em um novo corpus: ESSERCI, AVERCI e CI lexicalizante*. In XIII Congresso da ABPI Associação Brasileira de Professores de Italiano, Universidade Federal do Pará, Instituto de Letras e Comunicação, Belém, de 14 a 19 de setembro de 2009.

Tabela 12 - Ocorrência de *esserci* existencial e apresentacional nos *corpora* de italianos bilíngues e monolíngues

Clíticos	Corpus Raso-Vale				Corpus BADIP				Variação percentual	
	Absoluta		Normalizada		Absoluta		Normalizada		2007	2009
	2007	2009	2007	2009	2007	2009	2007	2009		
<i>esserci</i> existencial	96	90	53,09	49,77	44	44	24,33	24,33	118,2	104,56
<i>esserci</i> apresentacional	11	17	6,08	9,04	15	13	8,29	7,19	(-26,65)	25,73
Total	107	107	59,17	59,17	59	57	32,63	31,52	81,33	87,71

Fonte: Vale (2007; 2009)

A tabela evidencia divergências entre os dados: as ocorrências totais permanecem as mesmas para o verbo *esserci*, mas a nova análise fez salientar variações. Nos estudos anteriores, a presença da forma existencial no *corpus* de bilíngues era superior àquela dos monolíngues em mais de 118%. Os novos cálculos abaixaram de alguns pontos percentuais esta diferença. O que, todavia, é mais evidente, é a questão do *esserci* apresentacional: anteriormente haviam sido computados 11 ocorrências desta forma no *corpus* de bilíngues, contra 15 no de monolíngues (lembramos que os dois *corpora* comparados possuem o mesmo número de palavras). Portanto, supunha-se que nos bilíngues havia uma perda de 26,65%. As novas contagens apresentam 17 ocorrências do *esserci* apresentacional, contra 13 do *corpus* de monolíngues. Assim sendo, há um aumento desta função, de 25,73%, no *corpus* dos bilíngues. É importante lembrar que estes números absolutos são muitos reduzidos em um panorama geral da situação e portanto uma variação percentual é facilmente revertida com poucas ocorrências a mais ou a menos. O que mais ressalta nos dados anteriores e naqueles agora apresentados é que o *esserci* apresentacional apresenta uma certa estabilidade em todas as contagens feitas, da qual discutiremos mais a fundo adiante e que é muito importante nesta pesquisa.

Estes novos dados serão agora comparados com aqueles encontrados neste nosso estudo. Para facilitar a visualização, utilizaremos somente os elementos normalizados, pois os *corpora* possuem números de palavras diferentes:

Tabela 13 – Ocorrência normalizada da forma *esserci* nas funções existencial e apresentacional nos diferentes *corpora* de italianos bilíngues e monolíngues

Clítico	Corpus Raso-Vale	Corpus Raso-Ferrari	BADIP	C-ORAL-ROM Italiano	Corpus Raso-Vale e BADIP	Corpus Raso-Ferrari e C-ORAL-ROM Italiano
					Variação percentual	Variação percentual
<i>Esserci</i> existencial	49,77	48,43	24,33	78,67	104,56	(-38,43)
<i>Esserci</i> apresentacional	9,04	7,51	7,19	7,53	25,73	(-0,27)
Total	58,81	55,87	31,52	86,22	87,71	(-35,21)

Pode-se observar que as ocorrências normalizadas entre os dados dos dois *corpora* de bilíngues são bastante parecidas e não apresentam divergências consistentes: no *corpus* Raso-Vale as ocorrências de *esserci* em total a cada 10.000 palavras são de 58,81, enquanto no *corpus* Raso-Ferrari são de 55,87; o *esserci* existencial apresenta praticamente o mesmo número de ocorrências, 49,77 no primeiro *corpus* contra 48,43 no segundo. Diferenças ligeiramente maiores mas significativas dado o pequeno número de ocorrências, foram encontradas entre o *esserci* apresentacional: 9,4 ocorrências no *corpus* Raso-Vale contra 7,51 no *corpus* Raso-Ferrari.

Observemos agora os resultados comparados da forma atualizante *averci* nos dois estudos:

Tabela 14 – Ocorrência normalizada do clítico *ci* atualizante com o verbo *avere* nos diferentes *corpora* de italianos bilíngues e monolíngues

Clítico	Corpus Raso-Vale	Corpus Raso-Ferrari	BADIP	C-ORAL-ROM Italiano	Corpus Raso-Vale e BADIP	Corpus Raso-Ferrari e C-ORAL-ROM Italiano
					Variação percentual	Variação percentual
<i>Averci</i>	5,53	7,51	27,65	40,21	(-80)	(-81,33)

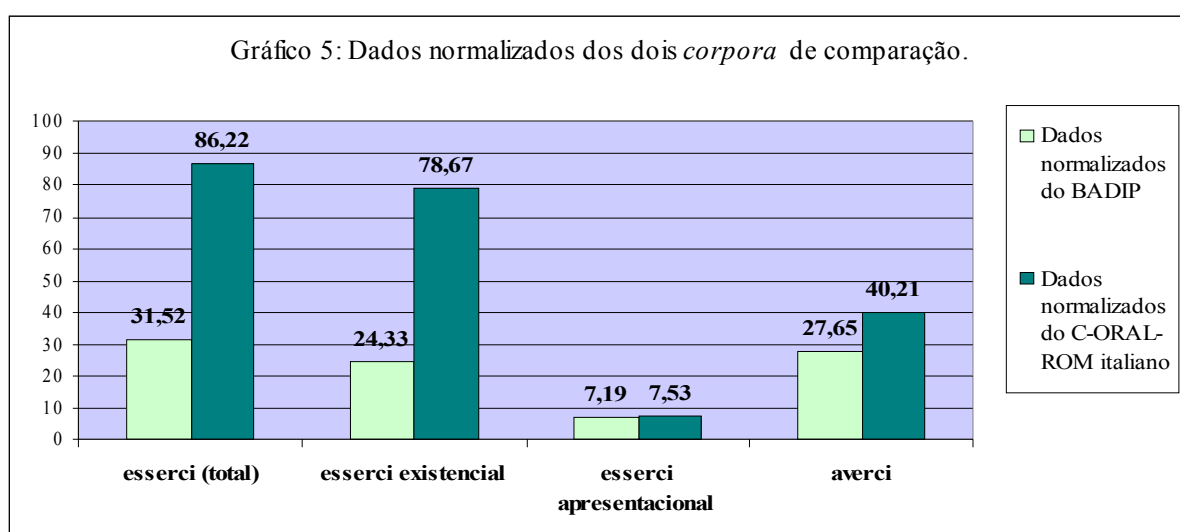
Esta comparação evidencia resultados similares aos apresentados para as formas de *esserci*: como já foi dito, não há variações consistentes no número de ocorrências normalizadas entre os dados encontrados por Raso e Vale (2009) e aqueles desta pesquisa. Também no caso do *averci* não há diferenças significativas: foram evidenciadas 5,53 ocorrências a cada 10.000 palavras no *corpus* Raso-Vale e 7,51 no *corpus* Raso-Ferrari.

A este ponto, é interessante observar com atenção as tabelas para constatar onde aparecem as divergências: se excluimos o caso do *averci*, em que a diminuição percentual em relação aos respectivos *corpora* de comparação praticamente coincide (-80% no caso do *corpus* Raso-Vale e -81,33% no *corpus* Raso-Ferrari) há várias disparidades entre os estudos

de Vale e Raso e aqueles atuais. Em primeiro lugar, é evidente que o aumento da forma *esserci* em total, detectada nos estudos anteriores, de 87,75%, aqui caminha em sentido oposto, pois há uma diminuição sensível, de 35,21%. Mesmo a forma existencial do verbo, que apresentava um aumento de 51,12%, nos estudos de Raso e Vale (2009), aqui apresenta uma perda de 38,54%.

Finalmente, o *esserci* apresentacional, que evidenciava um aumento de 23,52% em relação ao *corpus* de comparação, em nossa pesquisa não apresenta praticamente alguma variação, com uma diminuição de 0,27%. Contudo é importante lembrar que os números absolutos são muito diminutos e portanto as variações percentuais poderiam sofrer alterações muito grandes caso o número de ocorrências varie, mesmo que de pouco.

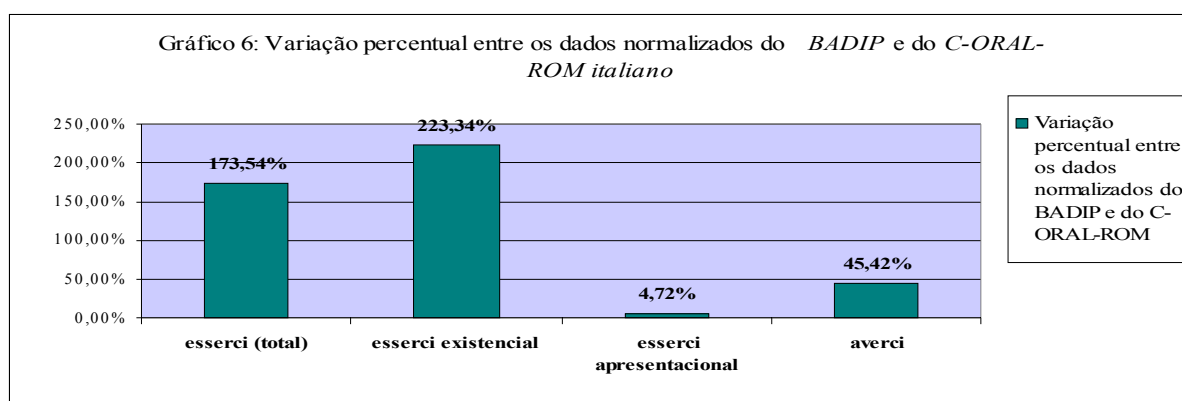
Na tentativa de entender melhor por que estes dados divergem, observamos as ocorrências normalizadas dos dois *corpora* de comparação: o *BADIP* e o *C-ORAL-ROM italiano*. O gráfico abaixo mostra as diferenças entre eles pelo que interessa os clíticos em questão:



Os dados acima são bastante eloquentes sobre a diferença entre os dois *corpora* de comparação utilizados nos dois estudos: no *BADIP* a cada 10.000 palavras encontram-se 31,52 ocorrências da forma *esserci* em total, enquanto no *C-ORAL-ROM italiano* estas ocorrências quase que triplicam, sendo 86,22. Subdividindo o *esserci* em suas duas funções encontramos mais divergências: no *BADIP* a forma existencial comparece 24,33 vezes a cada 10.000 palavras à medida que no *C-ORAL-ROM italiano* as ocorrências também neste caso triplicam, chegando a 78,67. Somente a forma do *esserci* com função apresentacional quase

não apresenta variação: no *BADIP* foram calculadas 7,19 ocorrências e no *C-ORAL-ROM italiano* 7,53, mais uma vez confirmando a estabilidade do *esserci* nesta função. Finalmente, a forma *averci* também patenteou diferenças nos dois *corpora*: no *BADIP* o número de ocorrências normalizadas foi de 27,65 ao passo que no *C-ORAL-ROM italiano* este número aumentou para 40,21.

Um gráfico com as variações percentuais destes dados pode mostrar de forma mais evidente a questão:



O que explica variações tão relevantes entre *corpora* de monolíngues? O que vemos é que em relação ao *BADIP* houve no *C-ORAL-ROM italiano* um aumento de 173,54% nas formas de *esserci* em total, de 223,34% nas ocorrências de *esserci* com valor existencial, uma insignificante variação de + 4,72% nas formas de *esserci* com valor apresentacional e um aumento de 45,42% nas ocorrências de *averci*. Na verdade a resposta reside na arquitetura dos dois *corpora*: o *BADIP* é um *corpus* que além de ser mais antigo, é muito mais formal, mesmo nos textos classificados como informais, e suas interações são muito menos espontâneas que aquelas encontradas no *C-ORAL-ROM italiano*, amplamente coloquial. Isto explicaria porque as formas de *esserci* e *averci* aumentam tanto em relação ao *BADIP*, enquanto a função apresentacional de *esserci* mantém-se estável, por ter um valor informacional que não depende do registro.

Achamos interessante, a este ponto, visto que as ocorrências normalizadas dos dois *corpora* de informantes sujeitos a erosão apresentavam resultados parecidos, e divergiam nas comparações com os *corpora* de monolíngues, operar um confronto cruzado, em que o *corpus* Raso-Vale fosse comparado com o *C-ORAL-ROM italiano* e o *corpus* Raso-Ferrari fosse comparado com o *BADIP*. Vejamos o que resultou nas tabelas abaixo, para se ter uma

referência mais clara, foram postos ao lado os resultados das comparações anteriores:

Tabela 15: Comparação entre a ocorrência normalizada de *esserci* existencial e apresentacional do *corpus* Raso-Vale em relação ao C-ORAL-ROM Italiano e ao BADIP.

Clíticos	<i>Corpus</i> Raso-Vale	C-ORAL-ROM Italiano	<i>Corpus</i> Raso-Vale e C-ORAL-ROM Italiano	<i>Corpus</i> Raso-Vale e BADIP
			Varição percentual	Varição percentual
<i>Esserci</i> existencial	49,77	78,67	(-36,73)	104,54
<i>Esserci</i> apresentacional	9,04	7,53	20,05	25,73
Total	58,81	86,22	(-31,79)	87,71

Pode-se observar que os resultados são opostos àqueles anteriores, quando o *corpus* de informantes sujeitos à erosão de Raso e Vale (2009) foi comparado àquele de monolíngues BADIP. No primeiro, o *esserci* existencial apresentava um aumento de mais de cem por cento, enquanto na comparação com o *C-ORAL-ROM italiano* é presente uma diminuição de ocorrências de 36,73%; a forma *esserci* na função apresentacional manteve-se quase estável na comparação com os dois *corpora* diferentes: em relação ao *C-ORAL-ROM italiano* houve um aumento de 20,05%, enquanto em relação ao BADIP o aumento tinha sido de 25,73%. Finalmente, no que se refere ao *esserci* em total, as duas comparações também evidenciaram resultados opostos: em relação ao BADIP havia sido registrado um aumento de 87,71% no número de ocorrências ao passo que em relação ao *C-ORAL-ROM italiano* há uma diminuição de 31,79%.

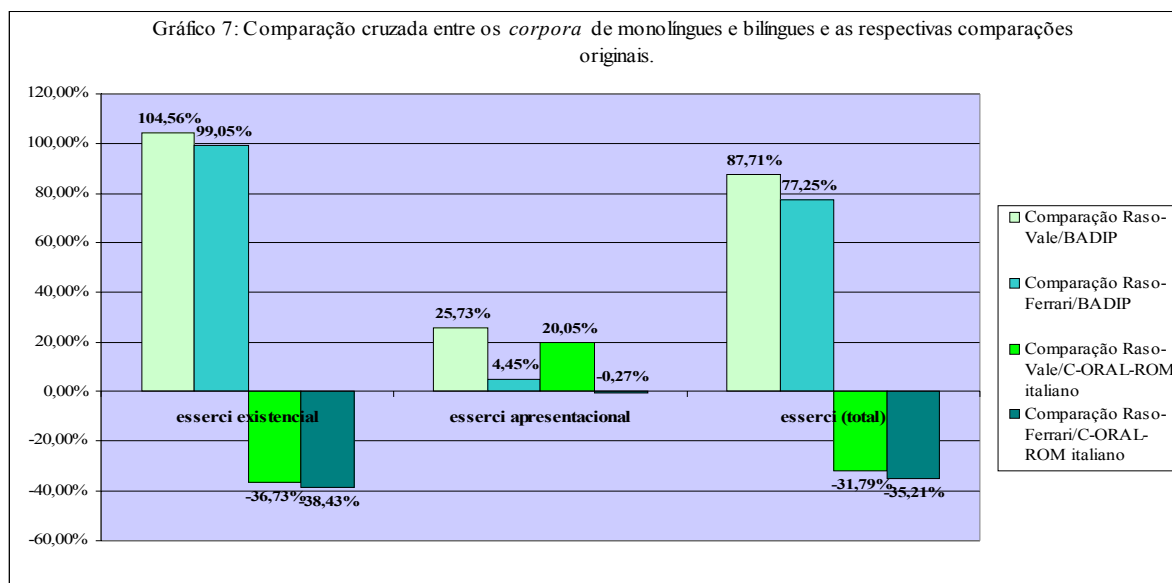
Vejamos agora quais os dados obtidos no mesmo tipo de comparação cruzada a partir do *corpus* Raso-Ferrari:

Tabela 16 - Comparação entre a ocorrência normalizada de *esserci* existencial e apresentacional do *corpus* Raso-Ferrari em relação ao BADIP e ao C-ORAL-ROM italiano.

Clíticos	<i>Corpus</i> Raso-Ferrari	BADIP	<i>Corpus</i> Raso-Ferrari e BADIP	<i>Corpus</i> Raso-Ferrari e C-ORAL-ROM italiano
			Varição percentual	Varição percentual
<i>Esserci</i> existencial	48,43	24,33	99,05	(-38,43)
<i>Esserci</i> apresentacional	7,51	7,19	4,45	(-0,27)
Total	55,87	31,52	77,25	(-35,21)

Também os resultados desta comparação cruzada mostram-se opostos àqueles obtidos anteriormente. Enquanto na comparação do *corpus* Raso-Ferrari com o *C-ORAL-ROM italiano* o *esserci* em função existencial evidenciava uma diminuição de 38,43%, comparando-o com o *BADIP* ocorreu um aumento de 99,05% no número de ocorrências. No *esserci* na forma apresentacional não houve grandes divergências, manifestando-se na comparação com o *BADIP* um aumento de 4,45% e naquela anterior com o *C-ORAL-ROM italiano* a já vista não significativa diminuição de 0,27%: mais uma vez foi confirmada a estabilidade desta forma. Por último, no caso do *esserci* total, na comparação anterior com o *C-ORAL-ROM italiano* havia sido detectada uma perda de 35,21%; na comparação com o *BADIP* a tendência foi novamente oposta, com um aumento do número de ocorrências da ordem de 77,25%.

Uma visualização através de um gráfico pode ser mais eficaz para mostrar as tendências opostas e ajudar a explicar o que está ocorrendo:



Como é possível observar pelo gráfico, quando os *corpora* de bilingües são comparados com o *BADIP*, as tendências são as mesmas: seja no caso do *esserci* existencial que do *esserci* total há um forte aumento das ocorrências seja no *corpus* Raso-Vale que do Raso-Ferrari. No caso do *esserci* apresentacional, observamos que a tendência é um pouco menos estável e mais consistente com os resultados das comparações anteriores, ou seja, com os *corpora* originalmente comparados. É importante ressaltar que na forma do *esserci* apresentacional, os dados normalizados do *BADIP* e do *C-ORAL-ROM italiano* praticamente

coincidem, por isso acreditamos não haver diferenças consistentes nas comparações cruzadas.

Quando os *corpora* são comparados com o *C-ORAL-ROM italiano*, notamos que a tendência é oposta mas consistente com o que foi dito: excluindo os casos do *esserci* apresentacional, os outros casos todos apresentam diminuições bastante similares.

Estas observações, a nosso ver, corroboram a tese da erosão dos dois *corpora* de bilíngues. Em outras palavras, nosso estudo confirmou a erosão linguística da forma *esserci*, tanto que os números absolutos dos *corpora* Raso-Vale e Raso-Ferrari são bastante similares entre si e as variações acontecem somente quando são postos em comparação com *corpora* de monolíngues diferentes, e mesmo assim as variações mantêm o mesmo padrão. Isto é devido à grande diferença entre a arquitetura do *BADIP* e aquela do *C-ORAL-ROM italiano*, como já foi dito acima.

Efetuiremos agora uma comparação cruzada para a forma *averci*: os dados obtidos no *corpus* Raso-Vale serão comparados com aqueles do *C-ORAL-ROM italiano* e aqueles do *corpus* Raso-Ferrari serão comparados com aqueles do *BADIP*. Para uma melhor visualização exporemos também as variações percentuais das comparações originais. A tabela 17 mostra o cruzamento de dados a partir do *corpus* Raso-Vale:

Tabela 17: Comparação entre a ocorrência normalizada de *averci* do *corpus* Raso-Vale em relação ao *C-ORAL-ROM italiano* e ao *corpus* *BADIP*.

Clíticos	<i>Corpus</i> Raso-Vale	<i>C-ORAL-ROM italiano</i>	<i>Corpus</i> Raso-Vale e <i>C-ORAL-ROM italiano</i> Variação percentual	<i>Corpus</i> Raso-Vale e <i>BADIP</i> Variação percentual
<i>Averci</i>	5,53	40,21	(-86,24)	(-80,00)

Observemos agora o cruzamento de dados a partir do *corpus* Raso-Ferrari:

Tabela 18 - Comparação entre a ocorrência normalizada de *averci* do *corpus* Raso-Ferrari em relação ao *BADIP* e ao *C-ORAL-ROM Italiano*.

Clíticos	<i>Corpus</i> Raso-Ferrari	<i>BADIP</i>	<i>Corpus</i> Raso-Ferrari e <i>BADIP</i> Variação percentual	<i>Corpus</i> Raso-Ferrari e <i>C-ORAL-ROM italiano</i> Variação percentual
<i>Averci</i>	7,51	27,65	(-72,83)	(-81,33)

O cruzamento de dados permitiu observar que, independentemente do *corpus* de

comparação, em todos os casos há diminuição nas ocorrências de *averci*. O *corpus* Raso-Vale comparado com seu *corpus* de referência original, *BADIP*, apresentava uma queda de 80%; quando o comparamos com o *C-ORAL-ROM italiano* a tendência foi a mesma com um pequeno aumento na diminuição, chegando a -86,24%. Também o *corpus* Raso-Ferrari manteve o mesmo comportamento nas duas comparações: com seu *corpus* de referência original, o *C-ORAL-ROM italiano*, foi registrada uma diminuição de 81,33%, enquanto quando foi comparado com o *BADIP* esta queda diminuiu um pouco, chegando a -72,83%.

Estes resultados não estavam previstos em nossas hipóteses: Raso e Vale (2009) tinham explicado a grande diminuição da forma *averci* motivando-a com a idade mais avançada dos informantes de seu *corpus*, que teriam migrado antes que essa forma se tornasse aceitável como forma padrão na língua falada. Como a erosão linguística da forma *averci* ficou evidenciada também em nosso *corpus*, e mesmo nas comparações cruzadas, é necessário refletir sobre as causas de uma queda tão grande. Uma das hipóteses é que nossos informantes provém quase todos do norte da Itália, onde o *averci* é menos utilizados. Entretanto, acreditamos haver um motivo mais forte para esta perda tão consistente: poderia se tratar de uma forma de hipercorreção em que os informantes, quase todos professores de italiano, evitam uma forma que eles sentem como não padrão, pouco correta ou normativamente inaceitável.

Queremos, para concluir, apresentar mais alguns dados para complementar nossa exposição mostrando as formas lexicalizadas com o clítico *ci* encontradas em nosso *corpus*, como mostra a tabela 19¹³⁸:

Tabela 19 - *Types* lexicalizantes do clítico *ci* e ocorrência de *tokens* nos *corpora* Raso-Ferrari e C-ORAL-ROM Italiano.

<i>Types</i> lexicais	<i>Corpus</i> Raso-Ferrari	<i>C-ORAL-ROM italiano</i>
<i>andarci (mettere)</i>	1	0
<i>Cascarci</i>	0	1
<i>Crederci</i>	0	2
<i>Entrarci</i>	0	2
<i>Mancarci</i>	0	1
<i>Metterci</i>	2	1

138 Lembramos que optamos por excluir aqui as formas *esserci* e *averci* por considerá-las atualizantes e não lexicalizantes. Veja-se a este respeito a explicação na seção 2.3.1.4

<i>Pensarci</i>	1	1
<i>Provarci</i>	1	1
<i>Rimanerci</i>	0	1
<i>Riuscirci</i>	1	1
<i>Starci</i>	3	2
<i>Vederci</i>	0	2
<i>Volerci</i>	1	7
TOTAL TYPES	7	12
TOTAL TOKENS	10	22

Os *types* presentes em nosso *corpus* são sete, em um total de dez *tokens*, em comparação com doze *types* e vinte e dois *tokens* do *C-ORAL-ROM italiano* que contém um número de palavras um pouco maior. A erosão das formas lexicalizadas é de qualquer jeito evidente, mesmo sem a necessidade de normalizar dados tão diminutos. Lembramos contudo que o *corpus* Raso-Vale tinha encontrado três *types* e três *tokens* de *ci* lexicalizante, contra seis *types* e dez *tokens* do *BADIP*.

Podemos concluir que, mesmo havendo uma erosão linguística em curso em ambos os *corpora* de bilíngues, em relação ao *ci* lexicalizante, o nosso *corpus* apresenta sinais menores de erosão, mas evidentes, apesar dos números pequenos, o que poderia depor a favor da tese de que a erosão é um processo que não se estabiliza, mas continua no tempo, como será discutido adiante. Outro fator relevante é, mais uma vez, a diferença entre os dois *corpora* de monolíngues: novamente o *C-ORAL-ROM italiano* se mostra mais coloquial, por um uso maior de formas lexicalizadas e portanto mais utilizadas na oralidade coloquial.

4.2.4 Os acusativos de terceira pessoa

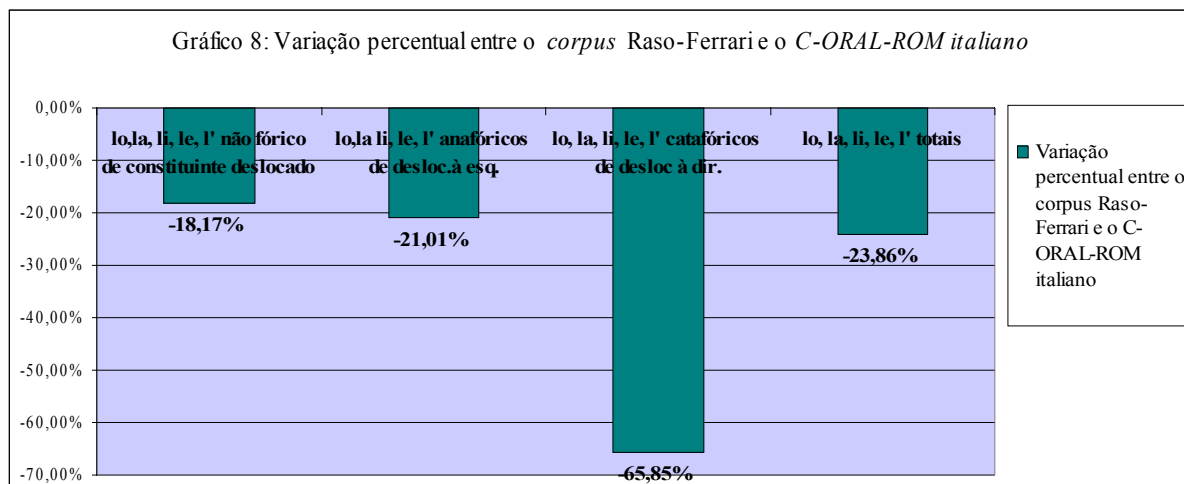
Serão agora apresentados os dados referentes aos pronomes acusativos de terceira pessoa, *lo, la, li, le, l'*, em seus números totais, no uso não fórico de constituinte deslocado e nas posições de deslocamento à direita e à esquerda como foi discutido no caso das análises de Raso e Vale, em comparação com o *corpus* extraído do *C-ORAL-ROM italiano*. Lembramos mais uma vez que serão apresentados os números absolutos e os dados normalizados de ambos os *corpora* apresentados, e a variação percentual entre eles.

Tabela 20: Ocorrência dos clíticos nos *corpora* de italianos bilíngues e monolíngues.

Clíticos	Corpus Raso-Ferrari		C-ORAL-ROM It		Variação percentual
	Absoluta	Normalizada	Absoluta	Normalizada	
<i>lo, la, li, le, l'</i> não fônicos de constituinte deslocado	202	94,84	246	115,9	-18,17
<i>lo, la, li, le, l'</i> anafônicos de elem. desl. à esquerda	23	10,79	29	13,66	-21,01
<i>lo, la, li, le, l'</i> catafônicos de elem. desl. à direita	12	5,63	35	16,49	-65,85
TOTAL	237	111,21	310	146,06	-23,86

É possível observar, já pelos números absolutos, que há uma perda em relação ao total de pronomes acusativos, havendo 237 ocorrências em total no *corpus* Raso-Ferrari e 310 no *corpus* de comparação. No caso dos pronomes usados em função não fônica de constituinte deslocado foram contadas 202 ocorrências no *corpus* Raso-Ferrari e 246 naquele de comparação, e no caso de deslocamentos foram encontradas 23 ocorrências de retomada do clítico de elementos deslocados à esquerda no *corpus* de bilíngues, contra 29 no de monolíngues; e 12 casos de catáfora do clítico em elementos deslocados à direita no *corpus* Raso-Ferrari contra 35 naquele de referência. Os dados normalizados fornecem uma visão mais clara: o total dos acusativos soma 111,21 ocorrências a cada 10.000 palavras no caso do *corpus* Raso-Ferrari, enquanto no *C-ORAL-ROM italiano* este número é mais alto: 146,06 ocorrências. Os pronomes utilizados em função não fônica de constituinte deslocado comparecem 94,84 vezes a cada 10.000 palavras no *corpus* Raso-Ferrari ao passo que no *C-ORAL-ROM italiano* esse número passa para 115,9. Os pronomes fônicos de deslocamentos à esquerda evidenciam um número de ocorrências de 10,79 casos no primeiro *corpus* e 13,66 no segundo. Já nos deslocamentos à direita foram encontradas 5,63 ocorrências de catáfora pronominal no *corpus* de bilíngues contra 16,46 no *corpus* de monolíngues.

O gráfico abaixo permite ter uma visualização mais clara da variação percentual entre os vários casos:



A erosão linguística dos pronomes acusativos de terceira pessoa é evidente em todos os casos, com uma diminuição total de 23,86% na produção dos indivíduos sujeitos a erosão em relação à produção dos monolíngues. Quando os pronomes não são fóricos de constituinte deslocado a diminuição é um pouco menor em relação à média total, sendo de 18,17%. No caso de retomada de elementos deslocados à esquerda a porcentagem de perda é quase análoga àquela total, sendo de -21,02%. Já os clíticos catafóricos de elementos deslocados à direita apresentaram uma diminuição três vezes superior ao total, de -65,88%.

No próximo parágrafo será feita a comparação destes dados com aqueles dos estudos anteriores de forma a obter um quadro mais definido da situação e formular novas hipóteses.

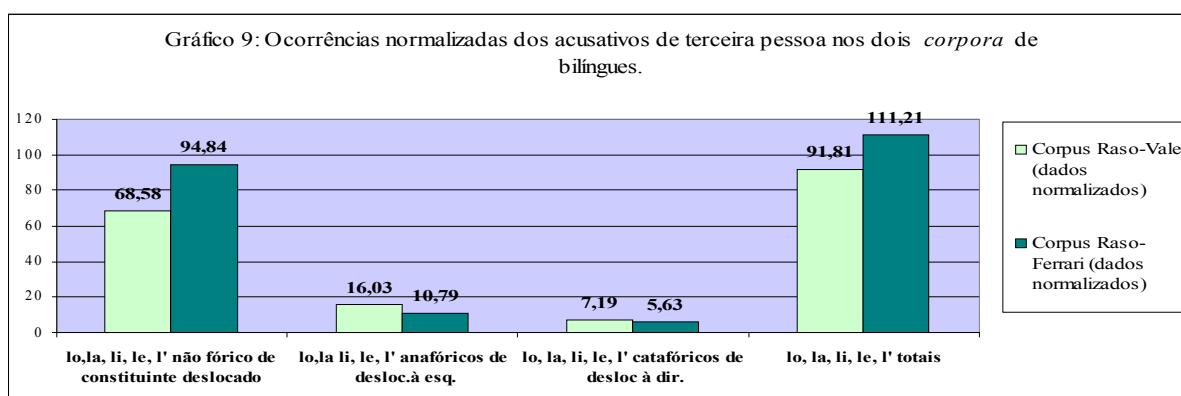
4.2.4.1 Análise contrastiva dos resultados

Os dados apresentados acima serão agora comparados com aqueles apresentados por Raso e Vale. Para facilitar a visualização serão expostos somente os dados normalizados:

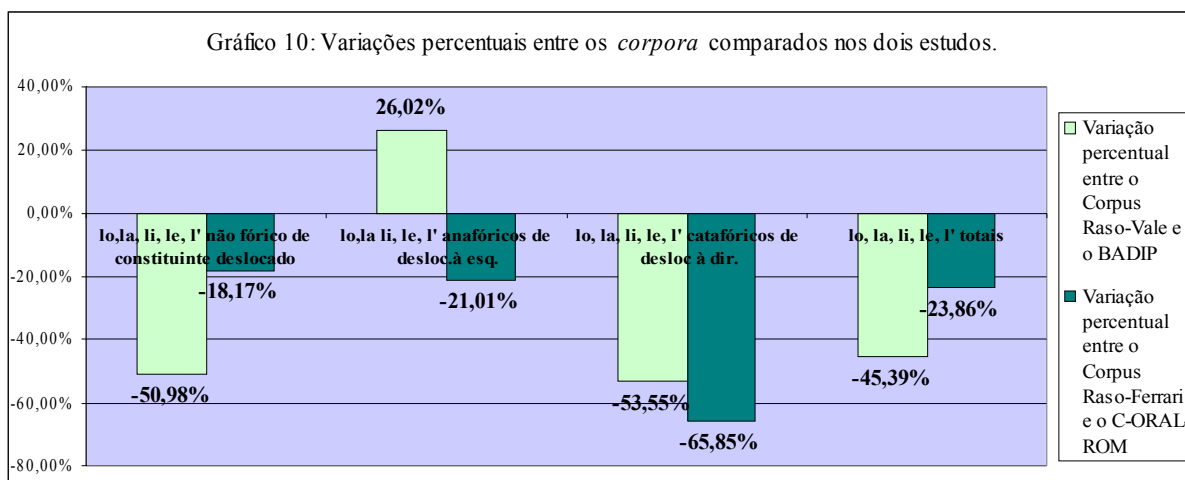
Tabela 21: Comparação entre as ocorrências normalizadas de acusativos de terceira pessoa do *corpus* Raso-Vale em relação ao *BADIP* e do *corpus* Raso-Ferrari em relação ao *C-ORAL-ROM Italiano*.

Clíticos	<i>Corpus</i> Raso-Vale	<i>Corpus</i> Raso- Ferrari	<i>BADIP</i>	C-ORAL- ROM Italiano	<i>Corpus</i> Raso- Vale e <i>BADIP</i> Variação percentual	<i>Corpus</i> Raso- Vale e C- ORAL-ROM Italiano Variação percentual
<i>lo, la, li, le, l'</i> não fórico de constituinte deslocado	68,58	94,84	139,93	115,9	-50,98	-18,17
<i>lo, la, li, le, l'</i> anafóricos de elem. desl. à esquerda	16,03	10,79	12,72	13,66	26,02	-21,01
<i>lo, la, li, le, l'</i> catafóricos de elem. desl. à direita	7,19	5,63	15,48	16,49	-53,55	-65,85
<i>lo, la, li, le, l'</i> TOTAIS	91,81	111,21	168,14	146,06	-45,39	-23,86

É possível notar, pelas ocorrências normalizadas que em geral a erosão linguística detectada no *corpus* Raso-Ferrari é menor em relação àquela detectada no *corpus* Raso-Vale. De fato, se formos observar somente as ocorrências dos pronomes não fóricos de constituinte deslocado poderemos ver que no *corpus* Raso-Vale somam-se 68,58 ocorrências a cada 10.000 palavras enquanto no *corpus* Raso-Ferrari este índice sobe para 94,84 ocorrências, sinal de que há menor erosão. O mesmo pode ser afirmado em relação à totalidade de ocorrências dos pronomes acusativos de terceira pessoa: há 91,81 ocorrências a cada 10.000 palavras no *corpus* Raso-Vale enquanto este número sobe para 111,21 no *corpus* Raso-Ferrari, confirmando o que dissemos acima. Os dados que divergem são aqueles relativos aos deslocamentos à esquerda onde há 16,03 ocorrências de pronomes de retomada anafórica no *corpus* Raso-Vale enquanto no *corpus* Raso-Ferrari temos somente 10,79 ocorrências. Nos casos dos deslocamentos à direita, não há grandes diferenças, passando de 7,19 ocorrências de catáfora pronominal no primeiro *corpus* contra 5,63 no segundo. O gráfico abaixo permite uma visualização mais clara do que foi dito acima:



É necessário contudo que os dados sejam vistos também em relação aos respectivos *corpora* de comparação para se ter um quadro contextualizado. O gráfico abaixo evidencia as porcentagens de variação dos dois estudos:



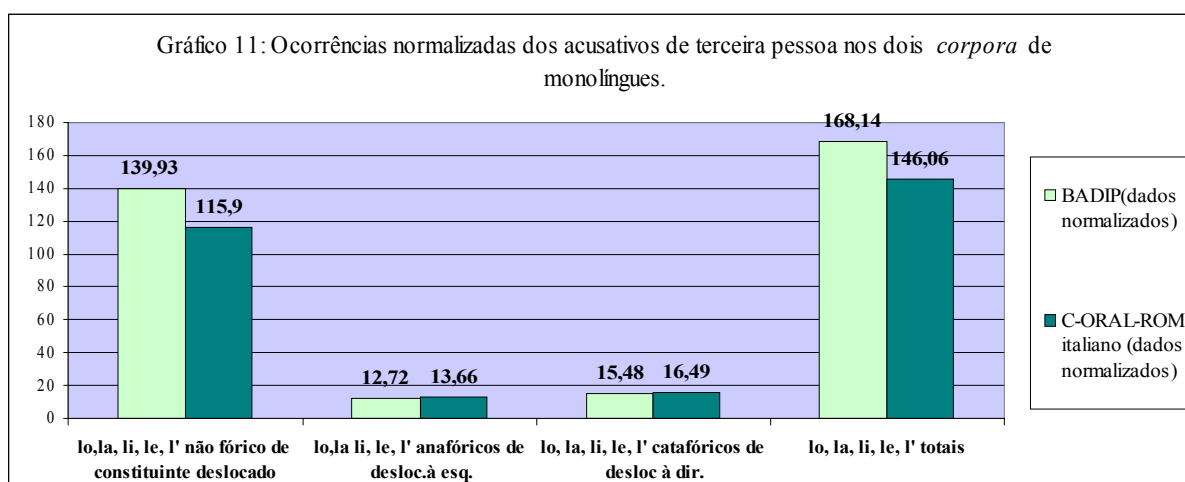
O gráfico mostra que os dois estudos apresentaram tendências parecidas, apesar de, como ter sido dito acima, o *corpus* Raso-Ferrari parecer estar menos sujeito à erosão. Lembramos que cada um dos *corpora* de bilíngues foi comparado com um *corpus* de referência diferente, e isto pode explicar em parte os resultados. Os acusativos de terceira pessoa em total patentearam perdas: de 45,39% no *corpus* Raso-Vale em relação ao *BADIP* e de 23,86 no *corpus* Raso-Ferrari em relação ao *C-ORAL-ROM italiano*. Houve perdas significativas mas diferentes também no caso dos acusativo em sua função não forica de constituinte deslocado, com uma diminuição de 50,98% no *corpus* Raso-Vale, ou seja ainda maior do que nos pronomes totais, e de 18,17% no *corpus* Raso-Ferrari, menor do que os pronomes totais. Nos casos de deslocamentos: naqueles à direita o *corpus* Raso-Ferrari apresentou uma perda maior, com -65,85% contra os -53,55% do *corpus* Raso-Vale. Já no caso dos deslocamentos à esquerda as tendências dos dois estudos são opostas: enquanto no *corpus* Raso-Vale houve um aumento de 26%, naquele de Raso-Ferrari foi registrada uma queda de 21,01%.

A hipótese aqui apresentada é que estas diferenças percentuais reflitam as desigualdades entre os *corpora* de italianos sujeitos à erosão coletados. O *corpus* Raso-Vale compõe-se em sua maioria por entrevistas: são interação não somente menos espontâneas mas

também prevalentemente monológicas, em que o informante deve continuamente informar ao interlocutor, instaurando cognitivamente tópicos que fazem referência a situações e lugares conhecidos por quem fala mas não por quem escuta, portanto novos ou diferentes dos anteriores para o interlocutor. As interações gravadas para o *corpus* Raso-Ferrari são mais espontâneas e coloquiais, fortemente baseadas na referência coloquial: em boa parte dos casos os interlocutores se referem a fatos e situações situacionalmente dados, e portanto cognitivamente a referência é dada pela situação e não requer ser posta como tema da ilocução; assim sendo não é induzida a retomada pronominal das topicalizações. Isto se soma ao fato de que os informantes de Raso e Vale estavam no Brasil no momento da entrevista há mais de trinta anos, em alguns casos há quarenta anos. Já os informantes contatados para o *corpus* Raso-Ferrari vivem no Brasil há menos tempo, em média quinze a vinte anos, mas alguns há oito-dez anos. Pesa portanto também o fator tempo de permanência na diferença entre os *corpora*, já que, como se viu, a estruturação dos enunciados dos brasileiros tende a topicalizar mais.

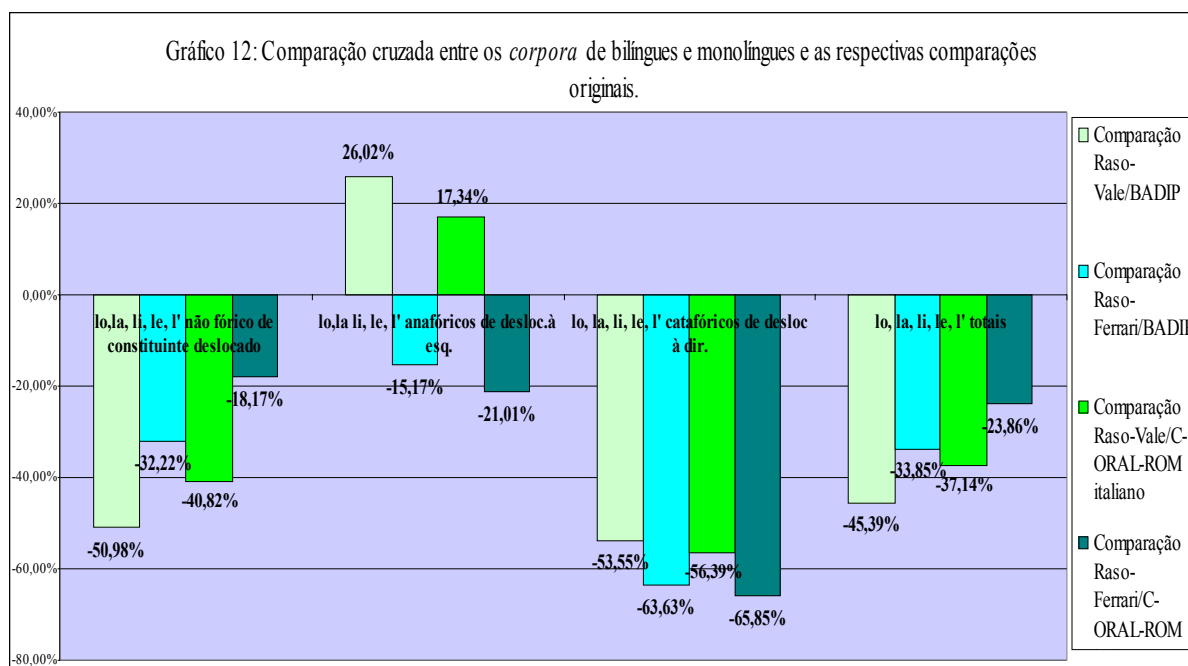
O que sugerimos é que o *corpus* Raso-Ferrari mostre sinais de erosão linguística, confirmando os dados de Raso e Vale, apesar das diferenças, que são devidas aos diferentes *corpora* utilizados. Caso não tivéssemos utilizado dados diferentes, seria impossível detectar as divergências que somente desta forma puderam ser notadas. Mas essa hipótese deve ser testada em um *corpus* maior e com textos variados tipologicamente.

Para concluir nossas comparações e reflexões sobre os acusativos de terceira pessoa, achamos interessante rever também os dados dos *corpora* de comparação dos monolíngues, para observar as diferenças que já foram notadas no caso do clítico *ci* atualizante. Observemos os dados no gráfico abaixo:



Nota-se à primeira vista que o *BADIP* apresenta um maior número de ocorrências de pronomes acusativos de terceira pessoa em praticamente todas as categorias, seja no total, com 168,14 ocorrências a cada 10.000 palavras contra 146,06 do *C-ORAL-ROM italiano*, seja em sua função não forica de constituinte deslocado, com 139,93 ocorrências no *BADIP* e 115,9 no *C-ORAL-ROM italiano*. No caso dos deslocamentos, seja à direita seja à esquerda, as divergências não são significativas, pois contam 15,48 ocorrências no *BADIP* e 16,49 no *C-ORAL-ROM italiano* no caso dos deslocamentos à direita, e foram encontradas 12,72 ocorrências a cada 10.000 palavras no *BADIP* e 13,66 ocorrências *C-ORAL-ROM italiano* no caso dos deslocamentos à esquerda.

Decidiu-se a este ponto operar uma comparação cruzada, comparando os dados do *corpus* Raso-Vale com aqueles do *C-ORAL-ROM italiano* e aqueles do *corpus* Raso-Ferrari com aqueles do *BADIP* para verificar se as divergências permaneciam ou se eram os *corpora* de monolíngues que estavam distorcendo nossos dados.



O gráfico explicita de forma clara as tendências: em todos os casos há erosão linguística, apesar de aparentemente haver um andamento oposto nos deslocamentos à esquerda quando a comparação é feita com o *corpus* Raso-Vale. Vejamos as porcentagens nos detalhes: nos casos em que os acusativos de terceira pessoa são não fóricos de constituinte deslocado o confronto de *corpora* Raso-Vale comparado com o *BADIP* evidencia uma diminuição de 50,98%, aquele

Raso-Ferrari, sempre comparado com o *BADIP* mostra sempre uma diminuição menor de 32,22%; as comparações com o *C-ORAL-ROM italiano* evidenciam reduções bem menores: o confronto Raso-Vale com o *C-ORAL-ROM italiano* mostra uma queda de 40,82% nas ocorrências enquanto aquele Raso-Ferrari comparado com o *C-ORAL-ROM italiano* de -18,17. As comparações com os pronomes anafóricos de constituintes deslocados à esquerda apresentaram no *corpus* Raso-Vale confrontado com o *BADIP* um aumento de 26%, naquele Raso-Ferrari comparado com o *BADIP* é registrada uma diminuição de 15,17%; no caso de Raso-Vale no confronto com o *C-ORAL-ROM italiano* foi novamente evidenciado um aumento, de 17,34%, enquanto no Raso-Ferrari comparado com o *C-ORAL-ROM italiano* novamente uma diminuição, de 21,01%. Os acusativos catafóricos de constituintes deslocados à direita evidenciaram em todas as comparações perdas bem parecidas: no caso de Raso-Vale comparado com o *BADIP* de -53,55%, no Raso-Ferrari confrontado com o *BADIP* de -63,63%, no Raso-Vale na comparação com o *C-ORAL-ROM italiano* de -53,39%, no Raso-Ferrari no confronto com o *C-ORAL-ROM italiano* de -65,85%. Finalmente, os pronomes acusativos de terceira pessoa em seus totais mostraram também perdas em todos os casos comparados: no caso de Raso-Vale confrontado com o *BADIP* de -45,39%, no Raso-Ferrari comparado com o *BADIP* de -33,85%, no Raso-Vale comparado com o *C-ORAL-ROM italiano* de -37,14%, no Raso-Ferrari comparado com o *C-ORAL-ROM italiano* de -23,86%.

Os números e gráficos expostos acima evidenciam duas questões intimamente ligadas entre si. Os dados que à primeira vista se mostram divergentes são aqueles relativos ao *corpus* Raso-Vale em comparação com os dois *corpora* de bilíngues. E os dados de Raso-Ferrari parecem não confirmar as explicações de Raso e Vale (2009) e Raso (2009) a respeito do aumento dos pronomes anafóricos de deslocamentos de elementos à esquerda. A questão que é posta é: seriam mais confiáveis os dados do *corpus* de Raso-Vale ou aqueles de Raso-Ferrari? Ou existe uma explicação alternativa? Os dados de Raso e Vale, quanto aos pronomes anafóricos de deslocamentos à esquerda, são confirmados também pelos dados relativos às frequências de topicalizações em PB, nos bilíngues e em italiano¹³⁹. Como foi visto, a pesquisa dos estudiosos confirmou que o PB utiliza muito mais tematizações, os bilíngues se colocam em uma posição intermediária e os italianos monolíngues utilizam menos tematizações e geralmente tematizam somente através do deslocamento à esquerda. Ou seja, o

139Veja-se seção 4.1.4.1.1 e 4.1.4.1.2

aumento de tematizações é por si um sinal de erosão, o que gera mais pronomes anafóricos de constituintes deslocados à esquerda, já que a retomada pronominal é obrigatória no italiano.

Talvez a explicação para as divergências nesta pesquisa seja a diferente arquitetura dos dois *corpora* de bilíngues. Os textos do *corpus* Raso-Vale, como dissemos várias vezes, são entrevistas, portanto a tendência é que sejam mais monológicos: o informante está narrando uma situação e deve em continuação operar uma referência cognitiva ao tema sobre o qual realiza a locução. Ao contrário, o *corpus* Raso-Ferrari contém situações muito mais interativas, os informantes estão agindo em um ambiente enquanto conversam e assim sendo os elementos cognitivos que constituem o âmbito semântico de um ato são dados justamente pela divisão do espaço dos participantes e do tempo situacional. Consequentemente não é necessário tematizar estes elementos e não haverá a retomada anafórica pronominal. Isto explicaria a diminuição dos pronomes anafóricos de elementos deslocados à esquerda no *corpus* Raso-Ferrari. Não é portanto que não haja erosão linguística neste traço, mas é a tipologia dos textos do *corpus* que determina uma menor necessidade de uso deste tipo de pronome.

Isso seria confirmado pela proximidade dos dados quanto aos deslocamentos das catáforas pronominais de elementos deslocados à direita. Os deslocamentos à direita são usados frequentemente também para retomar elementos dados. Se pensamos em instaurar um âmbito de aplicação de uma locução, a estratégia normal é topicalizá-lo, para fornecer a referência cognitiva da locução. Mas se a locução é feita sobre algo que é dado situacionalmente, podemos retomar a referência através de um deslocamento à direita.

O fato de ter utilizado um *corpus* diferente permitiu que estas características fossem evidenciadas e confirmou outros dados anteriores sobre a erosão pronominal. Dessa forma foi possível explicar, acreditamos, os dados vistos acima. Naturalmente a confirmação desta hipótese poderá ser feita somente com *corpora* maiores e que sejam comparáveis entre si.

5 CONSIDERAÇÕES FINAIS

O objetivo desta pesquisa foi verificar com um *corpus* novo e um diferente *corpus* de referência os resultados obtidos no estudo de Raso e Vale (2009) sobre a erosão linguística de italianos cultos em contato prolongado com o PB, e a metodologia adotada.

Os dados obtidos confirmaram a presença de erosão em todos os níveis, mas com índices diferentes; os pronomes investigados, a dizer: o pronome *ci* em seus valores atualizante, lexicalizante e locativo, os pronomes *ne* em suas funções partitiva, argumental e locativa e os pronomes acusativos de terceira pessoa, a dizer: *lo, la li, le, l'*, evidenciaram todos perdas significativas mas menos acentuadas daquelas vistas nos estudos de Raso e Vale.

É importante salientar que as diferentes metodologias adotadas neste estudo permitiram perceber situações que de outra forma não teriam sido enxergadas. Estamos cientes dos limites do *corpus* por nós montado que, apesar de seguir critérios mais rigorosos de coleta e transcrição, obtendo dados de textos mais diversificados situacionalmente em que a presença do pesquisador fosse quase imperceptível, peca por seu tamanho reduzido e por sua falta de balanceamento, pois mais de um terço das palavras pertencem a um único informante, o que pode ter comprometido os resultados finais. Contudo, mesmo com estes limites, os dados normalizados não evidenciaram grandes diferenças entre as ocorrências dos *corpora* Raso-Vale e daquele Raso-Ferrari, a não ser em alguns pontos específicos, o que também foi importante.

As diferenças que emergiram foram quase todas devidas à escolha de comparar os *corpora* de italianos bilíngues com *corpora* de monolíngues diferentes. Esta escolha não foi casual pois acreditamos que o *C-ORAL-ROM italiano* seja um *corpus* mais completo e mais coloquial em relação ao *BADIP*, permitindo um tipo de análise mais conforme com a realidade diária da fala. Sendo assim, gostaríamos de fazer algumas reflexões justamente sobre os pontos em que emergiram de tais divergências.

O *ci* atualizante com o verbo *essere*, que não tinha apresentado praticamente nenhuma alteração no estudo de Raso e Vale, nesta nossa pesquisa patenteou uma queda consistente. Acreditamos que isto seja devido justamente à diferença dos *corpora* de comparação: sendo o *C-ORAL-ROM italiano* muito mais variado e natural é óbvio que o uso desta forma seja maior enquanto os informantes sujeitos a erosão tendem a evitá-la. Este fato é ainda mais evidente no caso do *averci*: Raso e Vale tinham explicado a forte queda desta forma como devida a uma questão etária dos informantes, que migraram antes que esta forma fosse considerada

padrão na Itália. Dada a forte queda apresentada também em nosso estudo, esta argumentação não mais se sustenta. Cremos que, juntamente com as considerações feitas em Raso e Vale, haja uma explicação mais sutil, de tipo sócio-psicológico: os informantes são quase todos professores de italiano, portanto sentem um certo compromisso, mesmo em sua fala mais coloquial, em manter algum rigor normativo. Estamos dizendo no fundo que mesmo sua fala mais informal tende a ser controlada, pois a atitude de quem perdeu o contato com a pátria, na qual por outro lado a realidade linguística está se modificando em continuação, e deve ser um exemplo para os outros (ou pelo menos pensa que deva ser um exemplo), é aquela de hipercorrigir formas que são sentidas como não completamente padrão, evitando-as. Desta forma, psicologicamente, retomam um traço identitário com sua língua nativa, em uma maneira de agir que é bastante compreensível e típica de quem está fora de seu país e não recebe no ambiente natural a confirmação constante dos traços de mudança mais recentes.

Um outro dado que emergiu foi a forte diminuição do clítico *ci* locativo: Raso e Vale haviam suposto que a diminuição relativamente pequena por eles encontrada seria devida a uma tipologia de interação em que um entrevistador incentivava o informante ao uso deste pronome com perguntas sobre as origens e as andanças do migrante. De fato nosso *corpus* apresentou uma perda muito maior do *ci* locativo, confirmando as previsões de Raso-Vale; é importante de fato lembrar quão grande seja o peso das características próprias de cada texto. Nosso *corpus* não previa um assunto específico que induzisse a emergência de um traço como no *corpus* Raso-Vale.

O pronome *ci* com função existencial apresentou em todos os casos uma perda, confirmando a hipótese da erosão lexical. O que ao contrário permaneceu estável em todas as comparações feitas, foi o *ci* apresentacional. Acreditamos haver uma dupla explicação para este fato. Por um lado a função informacional que o *ci* apresentacional carrega tem um correspondente estrutural análogo no PB¹⁴⁰, portanto não haveria interferência negativa. Por outro lado, justamente por possuir uma função informacional tão forte, pois apresenta algo novo de maneira cognitivamente menos pesada, acaba por ser uma forma usada em qualquer tipo de interação, e portanto é menos sensível à tipologia do *corpus* que está-se pesquisando, tendendo a se manter estável.

Quanto ao caso do *ci* lexicalizante, pouco discutido pela exiguidade de suas ocorrências, é importante lembrar que foi confirmada sua erosão (em pleno acordo com a

140 Como no exemplo: "Tem um senhor que quer falar com você".

erosão lexical): o *C-ORAL-ROM italiano* foi aquele que apresentou o maior número de ocorrências, por ser mais recente e mais coloquial, o *corpus* Raso-Ferrari, apesar de uma certa erosão, apresentou alguns casos de verbos lexicalizados com *ci*, devido ao fato dos informantes residirem no Brasil há mais tempo. E o *corpus* Raso-Vale foi o que apresentou o menor número de ocorrências, sendo composto pela fala de informantes que moram fora da Itália há muito mais tempo.

Finalmente passamos a discutir o único caso em que as divergências não foram devidas aos *corpora* de comparação mas àqueles dos bilíngues, ou seja, os pronomes acusativos de terceira pessoa. Como foi discutido anteriormente, ter utilizado *corpora* diferentes no confronto com *corpora* diversos, permitiu descobrir duas tendências opostas no caso dos acusativos anafóricos de constituintes deslocados à esquerda e anafóricos de constituintes deslocados à direita: quando as comparações são feitas com o *corpus* Raso-Vale, há sempre um aumento das ocorrências; com o *corpus* Raso-Ferrari, há sempre diminuição no número de ocorrências. Discutiremos primeiro os deslocamentos à esquerda. Explicamos esta divergência com base nos diferentes tipos de texto dos dois *corpora* de bilíngues: o nosso *corpus* não precisa cognitivamente de instauração de muitos tópicos, por serem eles situacionalmente dados, o que acaba por diminuir a necessidade de delimitar semanticamente o âmbito da ilocução, portanto ocorre uma diminuição dos deslocamentos e consequentemente das retomadas de constituinte deslocado.

Entretanto, mais uma vez é bom esclarecer que esta hipótese deveria ser verificada em *corpora* maiores e que possuam situações comunicativas que possam ser comparadas.

Por outro lado a forte diminuição das catáforas pronominais de constituintes deslocados à direita é a confirmação da variação das anáforas pronominais de constituintes deslocados à esquerda. Como o âmbito da ilocução já é delimitado situacionalmente por todos os interlocutores, não há a necessidade de se fazer referência constante a este, portanto diminuem também, e de forma maior, os deslocamentos à direita, e consequentemente as catáforas pronominais.

Para concluir queremos evidenciar mais uma vez que a erosão linguística percebida nos estudos anteriores foi confirmada mas que é necessário ampliar o *corpus* de bilíngues e aquele de referência para obter dados mais sólidos e confirmar as hipóteses feitas com relação à presença pronominal nos deslocamentos. A Hipótese da Regressão parece ser confirmada, mas como mecanismo base integrado com outras causas (de acordo com as conclusões de

Keijzer 2007).

O que ao contrário ficou evidente é que a erosão pronominal afeta não somente o sistema morfológico, mas também aquele sintático e pragmático, o que mais uma vez nos faz refletir sobre a necessidade de dados mais numerosos e mais diversificados que possam confirmar esta afirmação.

Finalmente, achamos que os dados obtidos neste estudo depõem a favor de uma erosão linguística gradual no tempo. Muitos estudiosos¹⁴¹ afirmam que após cerca de dez anos o grau de erosão tende a estabilizar-se. Este estudo, feito com informantes que possuem um período médio de residência no Brasil inferior àqueles estudados por Raso e Vale, mostra em geral uma intensidade menor da erosão, confirmando a visão de quem diz que o processo de perda linguística continua no tempo¹⁴². Na realidade, Hutz (2004) afirma que os resultados sobre o peso do tempo de permanência no exterior mostram-se até agora inclusivos e que muitos outros fatores, como a frequência do contato com a terra nativa, entram em jogo. Mais uma vez é clara a necessidade de novas e mais acuradas pesquisas que verifiquem este e outros fatores.

Consideramos que este estudo se insere perfeitamente dentro da nova fase desejada por Köpke e Schmid (2004) que busca um esforço metodológico para consolidar conclusões em um âmbito de pesquisa complexo e condicionado por muitos fatores.

141Bot & Clyne (1994); Köpke & Schmid (2004).

142Soesman (1997); Waas (1996).

REFERÊNCIAS BIBLIOGRÁFICAS

ANDERSEN, R. W. "Determining the linguistic attributes of language attrition". In LAMBERT R. D.& FREED B. F. (Eds.), *The Loss of Language Skills*. Rowley, MA: Newbury House, 1982, pp. 83-119

APPEL R. PUYSKEN P. *Language contact and bilingualism*. London, New York, Sydney, Auckland: Arnold, 1987

ATKINS S., CLEAR J., OSTLER N.. 1992. "Corpus Design Criteria". in *Literary and Linguistic Computing*, vol 7, n. 1, 1992. Oxford University Press

AVRUTIN, S.; HAVERKORT M.; & HOUT A. VAN . "Introduction: Language acquisition and language breakdown". In *Brain and Language*, 77, 2001, p. 269-273

BAGNO, M. *Dramática da língua portuguesa: tradição gramatical, mídia & exclusão social*. 2. ed. São Paulo: Loyola, 2001.

_____. (Org.). *Linguística da norma*. São Paulo: Loyola, 2002.

BASTIAANSE, R. & BOL, G. "Verb inflection and verb diversity in three populations: Agrammatic speakers, normally developing children, and children with Specific Language Impairments (SLI)". In *Brain and Language*, 77, 2001, pp. 274-282.

BENINCÀ, Paola; SALVI, Giampaolo; FRISON, Lorenza. *L'ordine degli elementi della frase e le costruzioni marcate*. In: RENZI, Lorenzo; SALVI, Giampaolo; CARDINALETTI, Anna (Orgs.), 2001. v. 1, p. 129-239.

BENINCÀ, P. *Sintassi*, in SOBRERO 1993, vol. I, pp. 247-290

BERRETTA M. *Morfologia*. In: SOBRERO A. (Org.) *Introduzione all'italiano contemporaneo: le strutture*. Roma: Laterza, v. 1, 1991, p. 193-290

_____. *I pronomi clitici nell'italiano parlato*. In HOLTUS; RADKE (Orgs.) *Gesprochenes Italienisch in Geschichte und Gegenwart*, Tübingen: Narr, 1985, pp. 185-504

_____. *Il parlato italiano contemporaneo*. SERIANNI, L.; TRIFONE, P. (a cura di) *Storia della lingua italiana*. vol II: Scritto e parlato, Torino, Einaudi, 1994, p. 239-270

_____ *Morfologia*, In SOBRERO, VOL. I, 1993, p. 193-245

_____ *Per uno studio dell'apprendimento dell'italiano in contesto naturale: il caso dei pronomi personali atoni*. In: RAMAT, A. G. (Org.). *L'apprendimento spontaneo di una seconda lingua*. Bologna: Il Mulino, 1986. p. 329-352

BERRUTO G., *Dislocazioni a sinistra e grammatica dell'italiano parlato* in DE BELLIS-SAVOIA (a cura di), *Sintassi e morfologia della lingua italiana d'uso. Teorie e applicazioni descrittive. Atti del XVII Congresso della 266 Società di Linguistica italiana* (Urbino, 11-13 settembre 1983), Roma, Bulzoni 1985 pagg 59-82

_____, *Le dislocazioni a destra in italiano*, in STAMMERJOHANN (a cura di), *Tema-Rema in italiano*, Tübingen, Narr, 1986, pp. 55-69

_____ "Note sul repertorio linguistico degli emigrati italiani in Svizzera tedesca". In *Linguistica*, 31, 1991, pp. 61-79

_____ *Fondamenti di sociolinguistica*. Bari: Laterza, 2002

_____ *Italiano e tedesco in contatto nella Svizzera germanofona: interferenze lessicali presso la seconda generazione di immigrati italiani*. In Cordin, P.; Iliescu, M.; Siller-Runggaldier, H. (a cura di) *Parallela 6. Italiano e tedesco in contatto e a confronto. Atti del VII incontro italo-austriaco dei linguistici*. Innsbruck, 17-19 ottobre 1996. Trento: Dipartimento di Scienze Filologiche e Storiche, 1998, pp. 143-159

_____ *Le varietà del repertorio*. In A. Sobrero. *Introduzione all'italiano contemporaneo. La variazione e gli usi*. Laterza, Bari: 1993a, pp. 3-36

_____. *Varietà diamesiche, diastratiche, diafasiche*. In A. Sobrero. *Introduzione all'italiano contemporaneo. La variazione e gli usi*. Laterza, Bari 1993b, pp. 37-92

BETTONI C. *Imparare un'altra lingua*. Roma-Bari: Laterza, 2001

_____ *Italian language attrition in Sydney: the role of dialect*. In M.A.K. HALLIDAY, J. GIBBONS & H. NICHOLAS (Eds.) *Learning, Keeping and Using Language*. Amsterdam: Benjamins, Vol. 2, 1990, pp. 75-89,

_____ "Language shift and morphological attrition among second generation Italo australians". In *Rivista di Linguistica (RdLing)*. Pisa, Italia, 3 (2), 1991, pp. 369-387

BETTONI, C. & RUBINO, A. *Lingua e dialetto a confronto tra gli italiani di Sydney (Australia)*. In ROMANELLO M.T.; TEMPESTA I. (Eds) *Dialetto e lingue nazionali*. Roma, Bulzoni, 1995, pp. 361-395

BETTONI, C. *L'italiano fuori d'Italia*. In SOBRERO A. A. et al. (Eds.) *Introduzione all'italiano contemporaneo. La variazione e gli usi*. Bari: Laterza, 1993, pp. 411-460

BETTONI, C.; RUBINO, A. *Emigrazione e comportamento linguistico. Un'indagine sul trilinguismo dei siciliani e dei veneti in Australia*. Galatina (Lecce): Congedo Editore, 1996

BIBER D. "Representativeness in Corpus Design" in *Literary and Linguistic Computing*, Vol. 8, n. 4, 1993. Oxford University Press

BIBER D., CONRAD S., REPPEN R. *Corpus Linguistics: Investigating Language structure and Use*. Cambridge: CUP, 1998

BICK, E. *The Parsing System PALAVRAS: Automatic Gramatical Analysis of Portugese in a Constraint Grammar Framework*. Aarhus University Press, 2000

BODE S. de. *First Language Attrition: Productive Morphology Disintegration and Neurobiological Support. A case study*. Unpublished MA Thesis, Ponom, CA: California State Polytechnic University, 1996

BOT K. de & CLYNE M. A 16-year longitudinal study of language attrition in Dutch immigrants in Australia. In *Journal of Multilingual and Multicultural Development*, 15 (1), 1994, p. 17-28

BOUBA M., FILIACI F., HEYCOCK C., SORACE A., TSIMPLI I. (2002). *Syntactic attrition in Greek and Italian near-native speakers of English*. In *International Conference on First Language Attrition*. Amsterdam, 22-24 august 2002

BROWN R. *La prima lingua*. Roma: Armando, 1979

BRUNI, F. (Org.) *Storia della Lingua Italiana*. Bologna: Il Mulino, 1992-2000 (10 voll.)

CALVO CAPILLA, M.C. *Espanhol e português em contato: o atrito da L1 de imigrantes espanhóis no Brasil*. Brasília: 2007. 173 f. Dissertação de Mestrado. Universidade de Brasília

_____ "Espanhol e português em contato: o atrito da L1 de imigrantes espanhóis no Brasil". In *Revista Intercâmbio dos Congressos de Humanidades X Congresso Internacional de Humanidades Brasil - Chile*. Universidade de Brasília - UnB, 17, 18 e 19 de Outubro de 2007b. In <http://www.onda.eti.br/revistaintercambio/conteudo/visualiza.php?cod=77> acesso em 2 de julho de 2010

CAPRARA, Loredana. *Revista de Italianística* (5, 1997, pp.29-272)

CARAMAZZA, A. & ZURIF, E. B. *Comprehension of complex sentences in children and aphasics: A test of the regression hypothesis*. In CARAMAZZA A. & ZURIF E. B (Eds.)

Language Acquisition and Language Breakdown: Parallels and divergencies. Baltimore, MA: John Hopkins University, 1978, pp. 145-161

CASTELLANI, A. "Quanti erano gli italofoeni nel 1861?". In *Studi linguistici italiani*, 8, 1982, pp. 3-26

CELATA C.; CANCELILA J. *Erosione sociolinguistica nell'italiano di emigrati. Il caso della lunghezza consonantica*. In COSTAMAGNA L. SCAGLIONE S. (Orgs) *Italiano. Acquisizione e perdita*. Milano: FrancoAngeli, 2008

CHINI M. *Genere grammaticale e acquisizione. Aspetti della morfologia nominale in italiano L2*. Milano: FrancoAngeli, 1995

_____ *Che cos'è la linguistica acquisizionale*. Roma: Carocci, 2005

CIPRIANI, P. et al. *L'acquisizione della morfosintassi in italiano: fasi e processi*. Padova: Unipress, 1993, p. 126-139

COOK V.J. *Linguistics and second language acquisition*. London: MacMillan, 1993

_____ *Universal grammar in second language acquisition*. In Jordens, P.; Lalleman, J. (Eds.) *Investigating second language acquisition*. Berlin: Mouton de Gruyter, 1996, pp. 167-85

CORDIN; CALABRESE, A; *I pronomi personali*. In RENZI, 1988, pp 535-592

CORRÀ, L. *Lingua e identità etnica nelle comunità di origine veneta del Rio Grande do Sul (Brasile)*. In: Bombi, R.; Graffi, G. (a cura di), 1998, pp. 257-265

COSTAMAGNA, L.; SCAGLIONE, S. *Italiano acquisizione e perdita*. Milano: Franco Angeli, 2008

CRESTI, E.; MONEGLIA, M. (a cura di) *C-ORAL-ROM, Integrated Reference Corpora for Spoken Romance Language*, full edition in 9DVDs. Paris ; ELRA, 2005

D'ACHILLE, P. *La lingua dei semicolti*. In Serianni-Trifone. 1994, pp. 41-79.

_____. *L'italiano contemporaneo*. Bologna: Il Mulino, 2003.

DE BOT, KEES & WELTENS, BERT. . "Recapitulation, regression, and language loss". In: *First language attrition*. Eds. Selinger, Herbert W. & Vago, Robert M. Cambridge: University Press, 1991, p. 31-53.

DE MAURO, T. *Storia linguistica dell'Italia unita*. Bari: Laterza 1963

_____. *Grande dizionario italiano dell'uso*. Torino: UTET, 1999-2000

DE MAURO, T. *et alii* . *Lessico di frequenza dell'italiano parlato*. Milano: EtasLibri, 1993

DOUGHTRY C.J., LONG M.H. (Eds.) *The handbook of second language acquisition*. Oxford: Blackwell, 2003

DRESSLER W. U. *The sociolinguistic and patholinguistic attrition of Breton phonology, morphology, and morphonology*. In SELIGER & VAGO (Orgs.) *First Language Attrition*. Cambridge: Cambridge University Press, 1991

DUARTE, M. E. L. *Clítico acusativo, pronome lexical e categoria vazia no português do Brasil*. In: TARALLO, Fernando (Org.). *Fotografias sociolinguísticas*. Campinas: Pontes: Editora da Universidade Estadual de Campinas, 1989. cap. 1, p. 19-34. (Coleção linguagem-crítica).

DULAY H., BURT M., KRASHEN S. *La lingua seconda*. Bologna: Il Mulino, 1985

ELLIS R. *Multilingual Matters*, Clevedon, 1992 *Second language acquisition and language*

padagogy.

_____ *Second language acquisition and language pedagogy*. Clevedon: Multilingual Matters, 1992

_____ "Emergentism, connecionism and language learning". In *Language learning*, 48/4, 1998, pp. 631-64

GALVES, C.; ABAURRE, M. B. M.. *Os clíticos no português brasileiro: elementos para uma abordagem sintático-fonológica*. UNICAMP. Disponível em: <http://www.ime.usp.br/~tycho/participants/abaurre/abaurre_galves.htm>.

GIACALONE RAMAT A. *Italiano di stranieri*. In Sobrero (org.), *Introduzione all'italiano contemporaneo. La variazione e gli usi*. Roma-Bari: Laterza, 1993, pp. 341-410

_____ (Org.) *L'apprendimento spontaneo di una seconda lingua*. Bologna: Il Mulino, (1986)

GRASSI, C. *Italiano e dialetti*. In A. Sobrero. *Introduzione all'italiano contemporaneo. La variazione e gli usi*. Laterza, Basi 1993, pp. 279-310

GRODZINSKY, Y. *Theoretical Perspectives on Language Deficits*. Cambridge, MA: MIT Press, 1990

HALLER H. W. "Tra italiano, dialetto e anglo-americano: riflessi scritti della lingua parlata dagli emigrati italiani negli Stati Uniti nel primo Novecento, in Emanuele Banfi, Laura Gavioli, Cristina Guardino e Massimo Vedovelli (eds.), *Problemi e fenomeni di mediazione linguistica e culturale*. Atti del Quinto Congresso Internazionale dell'Associazione Italiana di Linguistica Applicata. Perugia: Guerra, 2006, p. 345-358.

_____ (2000) "Il plurilinguismo nell'emigrazione: riflessi linguistici nelle lingua parlata degli emigrati e in testi italo-americani." V Convegno Internazionale, Leuven, 22-25

Aprile 1998. *L'italiano oltre frontiera*. Serge Vanvolsem, Dieter Vermandere, Yves D'Hulst, Franco Musarra, eds. Leuven University Press. Florence: Franco Cesati Editore, 2, 281-289.

HALLER H. W. *Una lingua perduta e ritrovata: l'italiano degli italo-americani*. Biblioteca "Italiano & Oltre." Florence-Rome: La Nuova Italia, 1993

_____ *The Other Italy. The Literary Canon in Dialect*. Toronto-Buffalo-London: Toronto University Press, 1999

_____ "I piemontesi nel Far West - usi e atteggiamenti linguistici nella comunità piemontese di San Francisco." Atti XIIe XIII Rëscontr antërnational dë studi an sla lenga e la literatura piemontèisa. Ivrea: Tip. Vittorio Ferraro, 1998, p. 283-286.

HANSEN, L. *Not a total loss: The attrition of Japanese negation over three decades*. In HANSEN L. (Ed.) *Second Language Attrition in Japanese Contexts*. Oxford: Oxford University Press, 1999, pp. 142-153

HARLEY B. & WANG W. *The critical period hypothesis: Where are we now?* In GROOT A.M.B & KROLL J.F. (Eds.) *Tutorials in Bilingualism. Psycholinguistic Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, 1997, pp. 19-52

HENDRIKS H. *The structure of learner varieties*. Berlin: Mouton de Gruyter, 2005

HUTZ M. *Is there a natural process of decay? A longitudinal study of language attrition*. In KÖPKE B.; SCHMID M.S. *First Language Attrition. Interdisciplinary perspectives on methodological issues*. Amsterdam/Philadelphia: John Benjamin Publishing Company, 2004, pp. 189-206

ISTAT, *Lingua italiana e dialetti in Italia*, Statistiche ISTAT, 2002

ISURIN L. *Mechanism of First Language Forgetting*. Dissertation Abstracts International, Section A: The Humanities and Social Sciences (DAIA) 60 (9), 3342, 2000b

_____ Deserted island or a child's first language forgetting. In "*Bilingualism: Language and Cognition*", Cambridge University Press, 3 : 2000, 151-166

JACOBS B. Neurobiological Differentiation of primary and secondary language acquisition. In *Studies in Second Language Acquisition*, 10, 1988, pp. 303-335

JAKOBSON, Roman. *Il farsi e il disfarsi del linguaggio: linguaggio infantile e afasia*. Torino: Einaudi, 1971 (1941).

JASPAERT K.; KROON S. "Social determinants of language shift by Italians in the Netherlands and Flanders". In: *International Journal of the Sociology of Language* 90, 1991, pp. 77-96.

_____ "Social determinants of language loss". In *I.T.L. : Review of applied linguistics*, 83/84, 1989, p. 75-98

JORDENS, P.; DE BOT, K.; VAN OS, C.; SCHUMANS, J. *Regression in German case marking*. In Weltens, B. ; de Bot, K.; van Els T. (Eds.) *Language Attrition in Progress*. Dordrecht: Foris, 1986, pp. 159-174

KARL-FRANZENS-UNIVERSITÄT GRAZ. Geisteswissenschaftliche Fakultät. *Banca dati dell'italiano parlato: banco de dados*. Disponível em: <<http://languageserver.uni-graz.at/badip/badip/home.php>>. Acesso em: 15 de março de 2008.

KAUFMAN D., & ARONOFF, M. *Morphological disintegration and reconstruction in first language attrition*. In H.W. Seliger & R.M. Vago (Eds.), *First Language Attrition* (pp. 175-188). Cambridge: CUP, 1991

KEIJZER, Merel. *Last in first out? An investigation of the regression hypothesis in Dutch emigrants in Anglophone Canada*. Vrije Universiteit, 2007

KLEIN W. *Second Language acquisition*. Cambridge: Cambridge University Press, 1986

Kolk, H. H. J. "Does agrammatic speech constitute a regression to child language? A three-

way comparison between agrammatic, child and normal ellipsis". In *Brain and Language*, 77, 2001, pp. 340-350.

KÖPKE B.; SCHMID M.S. *First Language Attrition. Interdisciplinary perspectives on methodological issues*. Amsterdam/Philadelphia: John Benjamin Publishing Company, 2004

KRASHEN S. *Some issues relating to the Monitor Model*. In H. BROWN, C. YORIO, R. CRYMES (eds) *On Tesol '77*, TESOL, Washington, 1977, pp. 144-8

_____ *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press, 1981

LENNEBERG E. *I fondamenti biologici del linguaggio*. Torino: Boringhieri, 1971

LEONE, Fulvio. *I pronomi personali di terza persona: l'evoluzione di un microsistema nell'italiano di fine millennio*. Roma: Carocci, 2003.

LIGHTBOWN, P.M., SPADA, N. *How languages are learned*. Oxford: Oxford University Press, 1999 (ed. orig. 1993)

LUDI, G. *Forms and functions of bilingual speech in pluricultural migrant communities in Switzerland*. In: Fishman, J.; Toubret-Keller, A.; Clyne, M.; Krishnamurti, Bh.; Abdulaziz M. (Eds.) 1986, II: 217-236

LUZZATO, D. L. *TALIAN (Vêneto Brasileiro). Noções de Gramática, História e Cultura*. Porto Alegre: Sagra-Dc Luzatto, 1994

MACWHINNEY, B. *The CHILDES project: tools for analysing talk*. Hillsdale: Lawrence Erlbaum, 1994

_____ *The CHILDES Project: Tools for Analyzing Talk. Volume 1: Transcription format and programs. Volume 2: The Database*. Mahwah, NJ: Lawrence Erlbaum Associates, 2000

MAJOR R.C. "Interlanguage phonetics and phonology". In *Studies in Second Language Acquisition*, 20, 1998, p. 131-137

MARAZZINI, C. *La storia della lingua italiana attraverso i testi*. Bologna: Il Mulino, 2006

MARTIN, P. *WinPitch Corpus: A Software Tool for Alignment and Analysis of Large Corpora*. Disponível em: <<http://emeld.org/workshop/2003/martin-paper.pdf>>. Acesso em: 6 set. 2010.

McENERY T., XIAO R., TONO Y. *Corpus-based language studies: an advances resource book*. London; New York: Routledge, 2006

McENERY T., WILSON A. *Corpus Linguistics*. Edinburgh: Edinburgh University Press, 1996

MEYER C.F. *English Corpus Linguistics. An introduction*. Cambridge: Cambridge University Press, 2004

MEO ZILIO, G. "Sull'elemento italiano nello spagnolo rioplatense". In *Lingua Nostra*. 21, 1960, pp. 97-103

MEO ZILIO, G. "Una serie di morfemi italiani con funzione stilistica nello spagnolo dell'Uruguay". In *Lingua Nostra*, 20, 1959, pp. 49-54

MITCHELL, R.; MYLES, F. *Second language learnig theorie.*, London: Arnold, 1998

MONTRUL, S. *Convergent outcomes in L2 acquisition and L1 loss*. In: KOPKE B.; SCHMID M.S. *First Language Attrition. Interdisciplinary perspectives on methodological issues*. Amsterdam/Philadelphia: John Benjamin Publishing Company, 2004

_____ Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. In: *Bilingualism: Language and Cognition*, Cambridge

University Press, 2004, 7 (2), pp. 125–142

MYERS-SCOTTON, C. *Social Motivations for Codeswitching: Evidence from Africa*. Oxford: Clarendon, 1993

NEVES, Maria Helena de Moura. *Gramática de usos do português*. São Paulo: Ed. UNESP, 2000.

NICOLADIS E. & GRABOIS H. "Learning English and losing Chinese: a case study of a child adopted from China". In *International Journal of Bilingualism*, 6 (4), 2002, pp. 441-454

OBLER K.L.; MAHECHA N. R. *First language loss in bilingual and polyglot aphasic*. In SELIGER & VAGO (Orgs.) *First Language Attrition*. Cambridge: Cambridge University Press, 1991

PANUNZI A. *La variazione semantica del verbo essere nell'italiano parlato: uno studio su corpus*. Firenze: Firenze University Press, 2010

PERINI, Mário A. 2002. *Modern Portuguese: a reference grammar*. New Haven & London: Yale University Press, 2002.

PY, B.; GROSJEAN, F. *Variantes de contact, restructuration et compétence bilingue: approche expérimentale. Notions en questions. Rencontres en didactique des langues*, n. 6, set. 2002, pp. 19-27

PONTES, Eunice. *O tópico no Português do Brasil*. Campinas: Pontes, 1987

RASO T. "L'italiano parlato a San Paolo da madrelingua colti: primi sondaggi e ipotesi di lavoro". In *Revista de Italianística*, São Paulo, n. 8, p. 9-49, 2003

_____ *L'erosione linguistica degli italiani colti in contatto prolungato col PB*. In: X Congresso da ABPI, 2005, Foz de Iguaçu. Anais do X Congresso da Associação Brasileira dos

Professores de Italiano, 2005.

_____. *Erosione dei clitici e strutture tematizzanti in italiani colti in contatto prolungato col portoghese brasiliano*. In: X Congresso della Società Internazionale di Linguistica e Filologia Italiana. Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione, giustapposizione, 2009, Basileia. Atti del X Congresso della Società Internazionale di Linguistica e Filologia Italiana. Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione, giustapposizione, 2009. p. 384-399.

_____. "L'erosione linguistica dei madrelingua italiani colti in contatto prolungato col portoghese brasiliano: un progetto in corso". ABPI em revista, v. 2, p. 57-70, 2009.

RASO, T. ; VALE, H. P. "A erosão lingüística em italianos cultos em contato prolongado com o português do Brasil: os clíticos e alguns efeitos na estrutura do enunciado". In *Revista de Italianística*, v. 16, 2009, p. 1-22

REVISTA DE ITALIANÍSTICA. São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas da USP, n. 5, ano V. 1997. 291 p.

RIBEIRO, A. S. *Da perda dos clíticos no falar coloquial do Rio de Janeiro*. UFRJ. Disponível em: <<http://www.filologia.org.br/viiiicnlf/anais/caderno14-06.html>>.

RUSSI C. *Italian Clitics. An empirical Study*. Berlin New York: Mouton de Gruyter, 2008

SABATINI, F. *L'italiano dell'uso medio: una realtà tra le varietà linguistiche dell'italiano*, HOLTUS; RADKE (Orgs.) *Gesprochenes Italienisch in Geschichte und Gegenwart*, Tübingen: Narr, 1985, p. 154-184

SALVI, G.; VANELLI, L.. *Nuova grammatica italiana*. Bologna: Il Mulino, 2004.

SARDINHA T.B. *Lingüística de corpus*. Baruerí, SP: Manole, 2004

SCAGLIONE, S. *Attrition. Mutamenti sociolinguistici nel lucchese di San Francisco*. Milano: Franco Angeli, 2000.

SCHMID M.S. "A new blueprint for language attrition research". In KÖPKE B. e SCHMID M.S., *First Language Attrition. Interdisciplinary perspectives on methodological issues*. Amsterdam/Philadelphia: John Benjamin Publishing Company, 2004

_____ "First language attrition: The methodology revised". In *The International Journal of Bilingualism*, 8 (3), 2004 , pp. 239-255.

_____ *First Language Attrition, Use and Maintenance: The case of German Jews in Anglophone countries*. Amsterdam: John Benjamins, 2002

SCHMID, M. S.; DE BOT K. *Language Attrition*. In: DAVIES, A.; ELDER, C. (Eds.). *The Handbook of Applied Linguistics*. Oxford: Blackwell Publishing, 2004, p. 210-34,

SCHMITT E. *No more reductions! To the problem of evaluation of language*. In SCHMID & KOPKE. *First Language Attrition*. Amsterdam/Philadelphia: John Benjamin Publishing, 2004

SCHOENMAKERS-KLEIN GUNNEWIEK, M. *Structural aspects of the loss of Portuguese amongigrants: A Research Outline*. *ITL: Review of Applied Linguistics*, v. 83-84,1989, p. 99-123

SELIGER H.W. *Deterioration and creativity in childhood bilingualism*. In K. Hyltenstam & L.K. Opler (Eds.), *Bilingualism Across the Lifespan* (pp. 173-184).Cambridge: CUP, 1989

_____ Language attrition, reduced redundancy, and creativity. In H.W. Seliger & R.M. Vago (Eds.), *First Language Attrition* (pp. 227-240). Cambridge: CUP, 1991

SELINKER, L. *Interlanguage*. In *IRAL - International Review of Applied Linguistics in Language Teaching*. Volume 10, Issue 1-4, 1972, p. 209–232

SERIANNI, L.; CASTELVECCHI, A. (Colab.). *Grammatica italiana: italiano comune e lingua letteraria*. 2. ed. Torino: Utet, 1991.

SERIANNI, L.; TRIFONE, P. (a cura di) *Storia della lingua italiana*. vol II: Scritto e parlato, Torino, Einaudi, 1994

SERIANNI, L.; TRIFONE, P. (Orgs.) *Storia della lingua italiana*. 1992-94 (3 voll.)

SHARWOOD SMITH M.A. *On explaining language loss*. In FELIX S. & WODE H. (Eds.) *Language Development at the crossroads*. Tübingen: Gunter Narr, 1983a, pp. 49-69

_____ *On first language loss in the second language acquirer: Problems of transfer*. In GASS S. & SELINKER L. (Eds.) *First Language Attrition*. Cambridge: CUP, 1983b, pp. 17-30)

_____. *Crosslinguistic influence in language loss*. In: HYLSTENSTAM, K.; OBLER, L. K. *Bilingualism across the lifespan. Aspects of acquisition, maturity, and loss*. Cambridge: Cambridge University Press, 1989, p. 185-201.

SILVA, R. V. M. *Variação, mudança e norma: movimentos no interior do português brasileiro*. In: Bagno, 2002.

SILVA-CORVALÁN, C. *Language Contact and Change. Spanish in Los Angeles*. Oxford: Clarendon Press, 1994

SIMONE, Raffaele. *Stabilità e instabilità nei caratteri originali dell'italiano*. In: Sobrero, 1991. p. 41-100.

SINGLETON D. *Language acquisition: the age factor*. Multilingual Matters, Clevedon, 1989

SOESMAN, A. "An experimental study on native language attrition in Dutch adult

immigration in Israel". In J. Klatter-Folmer & S. Kroon (Eds.), *Dutch overseas: Studies in maintenance and loss of Dutch as an immigrant language* (pp. 181-194). Tilburg: Tilburg University Press, 1997

SOBRERO, Alberto A. (Org.). *Introduzione all'italiano contemporaneo: le strutture*. Roma: Laterza, 1991. v. 1.

SORACE, A. *Native language attrition and developmental instability at the syntax-discourse interface: data, interpretations and methods*. In *Bilingualism: Language and Cognition* 7: 2004, pp.143-145

TAMBURINI F. Annotazione grammaticale e lemmatizzazione di *corpora* in italiano. (p. 57-74, in Rossini Favretti (2000)

TELMON, T. *Varietà regionali*. in A. Sobrero. *Introduzione all'italiano contemporaneo. La variazione e gli usi*. Laterza, Bari 1993, pp. 93-150

TOMIYAMA M. *The first stage of second language attrition: a case study of a Japanese returnee*. In HANSEN L. (Ed.), *Second Language Attrition in Japanese Contexts*. Oxford: Oxford University Press, 1999, pp. 59-79

TSIMPLI, I.; SORACE, A.; HEYCOCK, C.; FILIACI, F. "First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English". In *International Journal of Bilingualism*, 8 (3), 2004, pp. 257-277

TURCHETTA B. *Il mondo in italiano. Varietà e usi internazionali della lingua*. Roma-Bari: Laterza, 2005

TURIAN D. & ALTENBERG E.P. *Compensatory strategies of child first language attrition*. In H.W. Seliger & R.M. Vago (Eds.) *First Language Attrition* (pp. 207-226). Cambridge: CUP, 1991

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO. Faculdade de Letras. *Análise*

contrastiva de variedades do português: banco de dados. Disponível em: <<http://www.lettras.ufrj.br/varport/corpora.htm>>. Acesso em: 21 mar. 2007.

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO. Faculdade de Letras. *Projeto norma lingüística urbana culta - RJ*: banco de dados. Disponível em: <<http://www.lettras.ufrj.br/nurc-rj/>>. Acesso em: 21 mar. 2007.

VAGO R. M. *Paradigmatic regularity in first language attrition*. In H.W. Seliger & R.M. Vago (Eds.) *First Language Attrition* (pp. 2241-252). Cambridge: CUP, 1991

VALE, H. P. A erosão lingüística dos italianos cultos em contato prolongado com o português do Brasil: os clíticos. Belo Horizonte: Monografia apresentada no Curso de Graduação em Letras da Faculdade de Letras da Universidade Federal de Minas, 2007

VALE, H. P. *A erosão dos clíticos verificada em um novo corpus: ESSERCI, AVERCI e CI lexicalizante*. In *XIII Congresso da ABPI Associação Brasileira de Professores de Italiano*, Universidade Federal do Pará, Instituto de Letras e Comunicação, Belém, de 14 a 19 de setembro de 2009 (Apresentação de Trabalho/Comunicação).

VAN ELS, T. "An overview of European research on language attrition". In: *Language Attrition in Progress*. Eds. Weltens, Bert, de Bot, Kees & van Els, Theo. Dordrecht: Foris Publications, 1986, p. 3-18.

VENTUREYRA V.A.G. E PALLIER C. *The case of adopted Koreans in France*. In KOPKE B.; SCHMID M.S. *First Language Attrition. Interdisciplinary perspectives on methodological issues*. Amsterdam/Philadelphia: John Benjamin Publishing Company, 2004, pp. 207-221

WAAS, M. *Language Attrition Downunder: German Speakers in Australia*. Frankfurt: Peter Lang, 1996

WILSON, A.; ARCHER, D.; RAYSON, P. (Edited by). *Corpus linguistics around the world*. Amsterdam-New York, NY: Editions Rodopi B.V. 2006

WYNNE M. *Developing Linguistic Corpora: a Guide to Good Practice*. Em <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/index.htm> (acesso em setembro de 2008)

YAĞMUR K. *First Language Attrition Among Turkish Speakers in Sidney*. Tilburg: Tilburg University Press, 1997

ANEXO A: Termo de Consentimento Livre e Esclarecido

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

N.º Registro COEP: **0075.0.203.000-09**

Título do Projeto: **O italiano em contato prolongado com o português do Brasil**

Prezado Senhor (a),

Este Termo de Consentimento pode conter palavras que você não entenda. Peça ao pesquisador que explique as palavras ou informações não compreendidas completamente.

*1) Introdução

Você está sendo convidado(a) a participar de uma pesquisa com o objetivo de estudar os efeitos sobre os falantes de língua materna italiana do contato linguístico prolongado (pelo menos 8 anos) com o português do Brasil. Você foi selecionado porque enquadra-se como italiano culto (terceiro grau, sendo que pelo menos até o segundo conseguido na Itália) e residente no Brasil há pelo menos oito anos. A pesquisa visa estudar os efeitos da língua portuguesa falada no Brasil sobre a língua italiana dos imigrantes e tem como um de seus objetivos finais a elaboração de materiais didáticos que sirvam de suporte aos professores nativos de italiano que possam estar sujeitos a atrito linguístico. Qualquer outra informação pode ser conseguida entrando em contato com o pesquisador responsável ou conversando diretamente com o pesquisador que entrou diretamente em contato com você para efetuar a gravação. As amostras escolhidas para o estudo constarão em um CD ou outro suporte de acesso público. Você vai poder, portanto, acessar as gravações que serão consideradas idôneas para formar o corpus.

A pesquisa visa analisar os efeitos do *input* prolongado do português brasileiro sobre o italiano dos nativos cultos, assim como do *input* de outros nativos na mesma situação, para que se possam descrever os traços que mais são sujeitos a erosão. Após vários anos de residência em um país estrangeiro e tendo um contato menor com a língua nativa, ou mesmo um contato com conterrâneos que ali residam há muito tempo, há uma tendência muito grande à perda da fluência na língua nativa e até mesmo de certas estruturas próprias de dita língua, fenômeno este conhecido como *erosão lingüística*. Este estudo visa justamente detectar quais aspectos são mais ou menos sujeitos à erosão.

Dispor de um *corpus* com o maior tipo possível de situações comunicativas (isto é, além de entrevistas, diálogos, falas espontâneas, monólogos, diálogos telefônicos) oferece a chance de estudar em detalhes como realmente a língua dos nativos italianos se modifica após vários anos de contato com o português do Brasil.

Do ponto de vista prático, este tipo de estudo pode ser a base de futuros estudos de matriz mais propriamente didática pois poderiam servir para implementar estratégias didáticas que sirvam como 'antídoto' à erosão nos professores de língua materna.

*2) Procedimentos do Estudo

Para participar deste estudo, solicito a sua especial colaboração em permitir que seja gravada sua fala espontânea durante um intervalo de tempo de não mais de duas horas. A sua fala pode ser gravada ou através de um microfone de ambiente ou através de microfone de lapela, dependendo da situação.

No caso de a sua fala já ter sido gravada, porque você entrou em uma situação em que estava já acontecendo a gravação, solicito a sua especial colaboração em permitir que essa gravação seja utilizada para os fins de pesquisa indicados. Você tem o direito de exigir que a gravação seja destruída. Você tem o direito de escutar a gravação antes de decidir sobre o destino dela.

*3) Riscos e desconfortos

Consideramos que a metodologia utilizada para coleta de dados não oferece riscos ou desconforto, além daquele inevitável devido à necessidade de gravar a fala do informante (em caso de diálogo entre duas pessoas e em alguns monólogos, o desconforto consiste em aplicar ao informante um microfone de lapela, sempre sem fio, ou seja, sempre sem nenhuma limitação de movimento).

*4) Benefícios

Consideramos que essa pesquisa não trará benefícios diretos para você.

*5) Custos/Reembolso

Você não terá nenhum gasto com a sua participação no estudo e também não receberá pagamento pelo mesmo.

*6) Caráter Confidencial dos Registros

A sua identidade será mantida em sigilo. Os resultados do estudo serão sempre apresentados como o retrato de um grupo e não de uma pessoa. Dessa forma, você não será identificado quando o material de seu registro for utilizado, seja para propósitos de publicação científica ou educativa.

Na parte da gravação que será publicada, o falante será indicado com uma sigla que não permitirá de maneira alguma a sua identificação. Relacionados à sigla serão disponibilizados somente os seguintes dados: faixa etária (dentro de um leque de pelo menos 10 anos); nível de estudo (dividido entre nenhum, primeiro grau completo, segundo grau completo, terceiro grau completo); sexo; ocupação. O registro da correspondência de cada sigla ao informante será mantido sigiloso e nunca será divulgado.

*7) Participação

Sua participação neste estudo é muito importante porém é voluntária. Você tem o direito de não querer participar ou de sair deste estudo a qualquer momento. Em caso de você decidir retirar-se do estudo, favor informar o pesquisador e/ou a pessoa de sua equipe que esteja atendendo-o.

*8) Informações

Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais, que poderá ser contatado para esclarecimentos pelo telefone 3499-4592, por email coop@prpq.ufmg.br ou no seguinte endereço: Av. Antonio Carlos, 6627 – Unidade Administrativa II, sala 2005. CEP 31270-901 - Belo Horizonte, MG.

Os pesquisadores responsáveis poderão fornecer qualquer esclarecimento sobre essa pesquisa, assim como tirar dúvidas, bastando contatá-los no seguinte endereço e/ou telefone:

Nome do pesquisador: Lúcia de Almeida Ferrari
Endereço: rua Benjamin Jacob, 196/701 Belo Horizonte
Telefone: (31) 32939202
Email: ferrari.lu@gmail.com

***9) Declaração de Consentimento**

Li ou alguém leu para mim as informações contidas neste documento antes de assinar este termo de consentimento. Declaro que toda a linguagem técnica utilizada na descrição deste estudo de pesquisa foi satisfatoriamente explicada e que recebi respostas para todas as minhas dúvidas. Confirmando também que recebi uma cópia deste Termo de Consentimento Livre e Esclarecido. Compreendo que sou livre para me retirar do estudo em qualquer momento, sem perda de benefícios ou qualquer outra penalidade.

Dou meu consentimento de livre e espontânea vontade para participar deste estudo.

Nome do participante (em letra de forma)

Assinatura do participante ou representante legal

Data

Obrigado pela sua colaboração e por merecermos sua confiança.

Nome (em letra de forma) e Assinatura do pesquisador

Data

ANEXO B: Parecer positivo do Comitê de Ética em Pesquisa da UFMG

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
COMITÊ DE ÉTICA EM PESQUISA - COEP**

Parecer nº. ETIC 075/09

**Interessado(a): Prof. Tommaso Raso
Câmara de Pesquisa
Faculdade de Letras - UFMG**

DECISÃO

O Comitê de Ética em Pesquisa da UFMG – COEP aprovou, no dia 06 de maio de 2009, após atendidas as solicitações de diligência, o projeto de pesquisa intitulado **"O italiano em contato prolongado com o português do Brasil"** bem como o Termo de Consentimento Livre e Esclarecido.

O relatório final ou parcial deverá ser encaminhado ao COEP um ano após o início do projeto.

**Profa. Maria Teresa Marques Amaral
Coordenadora do COEP-UFMG**