

# L'EROSIONE LINGUISTICA DI ITALIANI COLTI IN CONTATTO CON IL PORTOGHESE BRASILIANO: ASPETTI DEL SISTEMA PRONOMINALE

LÚCIA DE A. FERRARI\*  
TOMMASO RASO\*\*  
HELOISA PEREIRA VALE\*\*\*

ABSTRACT: L'articolo espone i risultati delle ricerche effettuate sull'erosione linguistica (o attrito linguistico) dell'italiano a contatto prolungato con il portoghese brasiliano (o PB). Il presente lavoro ha ampliato i dati degli studi compiuti da Raso e Vale (2007 e 2009) con la creazione di un nuovo *corpus* che seguisse criteri più rigorosi onde poter verificare i risultati ottenuti dagli studiosi. Oggetto di analisi sono gli stessi clitici studiati anteriormente: i pronomi accusativi di terza persona, i *ci* locativi, i *ci* attualizzanti e lessicalizzanti e i *ne*, nei valori partitivo, argomentale e locativo. I risultati hanno confermato nella sostanza quelli precedenti ma il fatto più interessante sono le divergenze emerse in questo studio rispetto a quello anteriore. Tale fatto è dovuto alle differenze presentate dai vari *corpora* analizzati: quello verificato da Raso e Vale è composto in maggioranza da interviste con un argomento definito, mentre quello attuale possiede un'architettura che privilegia interazio-

\*Universidade Federal de Minas Gerais (UFMG)

ferrari.lu@gmail.com

\*\*Universidade Federal de Minas Gerais (UFMG)

tommaso.raso@gmail.com

\*\*\*Universidade Federal de Minas Gerais (UFMG)

helopv@terra.com.br



ni di vario tipo. I due *corpora* quindi presentano strutture assai diverse, il che ha favorito la maggiore o minore produzione non solo di certi tipi di enunciati, ma anche di pronomi clitici. Sono state rinvenute diverse divergenze anche nei *corpora* di paragone: il C-ORAL-ROM italiano e il BADIP. Questo ancora una volta ha confermato quanto le tipologie testuali pesino nella comparsa o meno di determinate strutture. I dati dei quattro *corpora* sono stati analizzati e confrontati incrociandoli al fine di ottenere un panorama generale più completo dei fenomeni studiati. Il lavoro ha mostrato quanta attenzione debba essere posta nelle tipologie testuali presenti nei vari testi.

**Parole chiave:** erosione linguistica; contatto linguistico; *corpus*; italiano; portoghese brasiliano.

**RESUMO:** *O artigo expõe os resultados das pesquisas sobre a erosão linguística (ou atrito linguístico) do italiano em contato prolongado com o português brasileiro (ou PB). Este trabalho ampliou os dados resultantes dos estudos de Raso e Vale (2007 e 2009) através da criação de um novo corpus que seguisse critérios mais rigorosos com o intuito de verificar os resultados obtidos pelos estudiosos. Objeto de análise são os mesmos clíticos antes estudados: os pronomes acusativos de terceira pessoa, os clíticos locativos, os clíticos atualizantes e lexicalizantes e os ne, com valores partitivo, argumental e locativo. Os resultados substancialmente confirmaram as anteriores; todavia, o fato mais interessante consiste nas divergências apresentadas pelos vários corpora analisados: aquele conferido por Raso e Vale é composto, em sua maior parte, por entrevistas com um assunto já definido, enquanto o atual possui uma arquitetura que privilegia interações diferenciadas. Os dois corpora apresentam, portanto, estruturas muito diferentes, o que favoreceu a maior ou menor produção não somente de certos tipos de enunciado, mas também de pronomes clíticos. Foram encontradas várias divergências também nos corpora de comparação: o C-ORAL-ROM italiano e o BADIP. Este fato, mais uma vez, confirmou que as tipologias textuais têm um peso determinante no aparecimento*

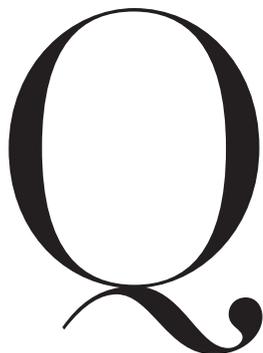
*ou na ausência de certas estruturas. Os dados dos quatro corpora foram analisados e comparados de forma cruzada para se obter uma visão geral mais completa dos fenômenos estudados. O trabalho aqui apresentado mostrou quanta atenção mereçam as tipologias textuais presentes nos vários textos.*

*PALAVRAS-CHAVE: erosão linguística; contato linguístico; corpus; italiano; português brasileiro.*

*ABSTRACT: This article shows the results of researches about attrition of Italian in contact with Brazilian Portuguese. This work compares the results of Raso's and Vale's studies (2007 and 2009) with the collection of a new corpus following more rigorous criteria. Object of the analysis is the same group of clitics previously studied: third person accusative pronouns, locative *ci*, "attualizzanti" and "lessicalizzanti" *ci*, partitive, argumental and locative *ne*. The results confirmed a great part of the previous ones some differences emerge. This is due to differences between the corpora: that used by Raso and Vale was mainly built with interviews about the same topic, while the new one has an architecture that collates a much wider variety of interactions. The two corpora present then very different structures, which induced different percentages of some phenomena. This confirmed how important it is to have different textual varieties for the elicitation of particular structures. Data from the four corpora were analyzed and compared in order to obtain a broader picture of the studied phenomena.*

*KEY-WORDS: attrition; linguistic contact; corpus; Italian; Brazilian Portuguese.*

## 1. Obiettivo del lavoro e contestualizzazione



Questo studio intende verificare e continuare i lavori svolti da Vale (2007), Raso e Vale (2009) e da Raso (2009) sull'erosione linguistica del sistema pronominale clitico di italiani colti in contatto prolungato con il portoghese brasiliano (PB). Le ricerche precedenti si sono concentrate sull'analisi di uno specifico gruppo di clitici: i pronomi accusativi di terza persona, i *ci* locativi, i *ci* attualizzanti e lessicalizzanti e i *ne*, nei valori partitivo, argomentale e locativo di origine.

Il lavoro qui presentato riprende l'analisi di questo stesso gruppo di pronomi utilizzando un diverso *corpus* di bilingui, compilato con l'obiettivo di integrare quello precedente; al BADIP, come *corpus* di comparazione, è stato aggiunto anche il C-ORAL-ROM italiano<sup>1</sup>.

I risultati confermano l'erosione linguistica in corso ma mostrano alcune differenze interessanti rispetto alle conclusioni degli studi precedenti sui dati relativi al sistema pronominale, e mettono in evidenza come, contrariamente a ciò che la maggior parte degli studi sostiene, il processo di erosione non si compie fondamentalmente nel primo decennio di contatto ma continua ben oltre questo periodo.

Nella ricerca sono stati adottati gli stessi criteri di Raso e Vale (2009) e di Raso (2009): l'erosione linguistica è intesa, sulla base di Kopke e Schmid (2004, 5), come una perdita non patologica nella prestazione nella propria lingua materna, cioè una perdita intergenerazionale; gli informatori sono tutti italiani colti in con-

<sup>1</sup> CRESTI, MONEGLIA (2005).

tatto prolungato con il PB: devono cioè aver completato l'intero ciclo di studi, fino alle superiori, in Italia, e possedere un titolo universitario, conseguito in Italia o altrove, il che garantisce una completa acquisizione della L1 oltre ad un buon grado di riflessione metalinguistica e sorvegliatezza; inoltre devono vivere in Brasile da almeno otto anni, periodo considerato sufficiente per un'acquisizione quasi nativa della L2.

Le caratteristiche degli informatori consentono di isolare gli effetti dell'erosione linguistica da quelli di una acquisizione parziale (come nel caso di parlanti con limitato grado di istruzione), riducendo le variabili nel tentativo di distinguere ciò che nel processo di erosione è dovuto al contatto con una L2 e ciò che invece può essere dovuto a un processo di ristrutturazione dovuto alla mancanza di *input* nella L1, noto come l'ipotesi dell'Isola Deserta<sup>2</sup>, secondo cui un parlante abbandonato ipoteticamente in un'isola deserta per lungo tempo sarebbe sottoposto a un processo di erosione pur senza l'intervento di nessuna L2 di contatto. Nel caso dei nostri informatori non sarebbe esatto dire che non ricevano *input* nella L1; tuttavia si tratta quasi esclusivamente di *input* di parlanti sottoposti allo stesso processo di erosione.

## 2. I risultati degli studi precedenti

Gli studi di Raso e Vale (2009) e di Raso (2009) hanno analizzato un *corpus* di 18.080 parole, che chiameremo *corpus* Raso-Vale, che è un sottocorpus di un *corpus* compilato presso la USP e pubblicato sulla *Revista de Italianística* (1997)<sup>3</sup>, e di un *corpus* raccolto presso la UFMG tra il 2004 e il 2006. Il *corpus* che ne è risultato è stato poi confrontato con stralci del BADIP<sup>4</sup>, dovutamente selezionati per somiglianza di tipologia testuale, per un totale di 18.080 parole.

Gli studi hanno mostrato una riduzione generale dei clitici di circa il 34%, ma in misura variabile sia per il tipo di clitici sia per la funzione che essi svolgono nella struttura informativa dell'enunciato<sup>5</sup>.

Per quanto riguarda la *ci* locativo, la riduzione è stata del 38,4%, al di sotto dell'attesa, considerata la posizione che il *ci* locativo occupa nella scala di acquisizione dell'italiano sia come L1 sia come L2<sup>6</sup>. Il rapporto con la sequenza acqui-

2 SHARWOOD SMITH; VAN BUREN (1991); ISURIN (2000).

3 Si veda in proposito anche Raso (2003).

4 <http://badip.uni-graz.at/>. Si veda De Mauro et al. (1993).

5 Sui clitici italiani si vedano Berretta (1985 e 1986); Cordin-Calabrese (1988); Leone (2003); Russi (2008).

6 Si vedano Berretta (1986) e Giacalone Ramat (1993).

sizionale è importante alla luce della cosiddetta “ipotesi della regressione” di Jakobson (1971), secondo cui si perderebbe prima ciò che si acquisisce dopo. Questa ipotesi è stata testata da Keijzer (2007), che l’ha sostanzialmente confermata ma ridimensionandone molto il peso, sostenendo che altri fattori di ordine psicologico e la tipologia della L2 di contatto possono pesare anche di più di questa tendenza, per così dire, naturale. Tornando alla relativamente alta presenza di *ci* locativi nel *corpus* Raso-Vale, ciò è stato giustificato col fatto che la tipologia dei testi utilizzati nel *corpus* induceva un uso del pronome superiore alla media: essendo quasi tutte interviste semiguideate il cui argomento principale erano i trasferimenti degli informatori, si è supposto che l’argomento inducesse un uso molto maggiore di strutture locative e, di conseguenza, la loro ripresa anaforica con il *ci*.

Il *ne* nelle varie funzioni analizzate ha presentato forti segnali di erosione: in totale è stata segnalata una perdita generale pari al 52%, ma va detto che i numeri di base erano ben più ridotti, dal momento che il *ne* ha una frequenza di occorrenza più bassa di quella di altri pronomi. Si è avuto un calo del 62% per il *ne* partitivo, del 20% per quello argomentale, mentre non è stata incontrata alcuna occorrenza di *ne* locativo di origine; il *corpus* monolingue presentava appena tre occorrenze.

Per quanto riguarda la funzione attualizzante<sup>7</sup>, sono stati vagliati i verbi *esserci* e *averci*: la prima forma ha fatto riscontrare un aumento dell’87%, mentre la seconda una diminuzione dell’80%. Raso e Vale (2009) e Raso (2009) spiegano questa distribuzione così diversa come segue. Nel caso dell’enorme aumento di *esserci* si tratterebbe di una conseguenza dell’erosione lessicale: forme verbali equivalenti verrebbero tutte ridotte a quella più generica di *esserci*. Per quanto riguarda invece il grande calo di *averci*, gli autori lo attribuiscono al diverso statuto che la forma presenta nell’italiano parlato contemporaneo in relazione a quello dell’epoca di trasferimento dei bilingui da loro studiati. Difatti questi ultimi, nella maggior parte, sono emigrati in Brasile negli anni Cinquanta-Sessanta, quando l’espressione in questione era ancora considerata *substandard*. Il *corpus* BADIP è stato raccolto negli anni Novanta, per cui è comprensibilmente ovvio che la forma *averci* vi sia ben rappresentata.

Infine, per quanto riguarda il clitico *ci* con valore lessicalizzante<sup>8</sup> sono stati riscontrati solamente tre *types* lessicali e tre *tokens* nel *corpus* di bilingui, a confronto di sei *types* e dieci *tokens* nel *corpus* di monolingui. Ancora una volta, nonostante l’e-

7 Si vedano Sabatini (1985) e, per “esserci”, Panunzi (2010).

8 Si vedano Sabatini (1988) e soprattutto Russi (2010).

siguità dei numeri e la necessità di verifiche su *corpora* maggiori, viene evidenziata l'erosione lessicale da parte del gruppo studiato dai ricercatori.

I dati più interessanti riguardano gli accusativi di terza persona. La perdita totale presenta un calo del 45% circa rispetto al *corpus* dei monolingui, in linea con le attese, basate sull'ordine acquisizionale. Tuttavia Raso e Vale (2009) e Raso (2009) hanno voluto operare una analisi più approfondita, distinguendo tali pronomi nelle funzioni e posizioni che occupano nell'enunciato, ossia disaggregando i dati per vedere il comportamento dei pronomi anaforici di elementi dislocati a sinistra e di quelli cataforici di elementi dislocati a destra. Ciò che è emerso è osservabile nella tabella sotto:

Clitici	Italiani bilingui		Italiani monolingui
	Valore assoluto	Variazione percentuale	Valore assoluto
Clitici accusativi di terza persona (lo, la, li, le, l')	166	(-45,4)	304
In enunciati con dislocazione a sinistra	29	25	23
In enunciati con dislocazione a destra	13	(-54)	28

**Tabella 1:** Occorrenze dei clitici accusativi di terza persona nei *corpora* di italiani bilingui e monolingui totali e nelle diverse posizioni.

Come detto sopra, la perdita generale dei clitici accusativi è del 45,4%. Tuttavia, quando essi svolgono funzione anaforica o cataforica di elementi dislocati, il discorso cambia: ove il pronome ha funzione anaforica, i dati mettono in evidenza un aumento del 25% rispetto al *corpus* dei monolingui. Ciò significa che deve essere tenuta in conto, nel totale degli accusativi, una perdita ancora maggiore rispetto ai 45% rilevati. Nel caso invece dei clitici cataforici di costituenti dislocati a destra, la perdita riscontrata raggiunge il 54%, superiore quindi a quella media.

Per spiegare questi dati, Raso e Vale hanno condotto una ricerca sulla strut-

turazione dell'enunciato, utilizzando dati di *corpora* di monolingui italiani, bilingui e monolingui brasiliani. La ricerca riguardava la frequenza e la strategia di tematizzazione, ossia quanto e come le due lingue (e i bilingui) mettono in evidenza costituenti con valore tematico, tali cioè da costituire il *topic* dell'enunciato e non l'elemento portatore del *comment*. La tabella sotto rende conto dei risultati relativi a 11.000 parole per ognuno dei *corpora* utilizzati: il NURC<sup>9</sup> di Rio de Janeiro per il PB, il BADIP per l'italiano e il *corpus* Raso-Vale per i bilingui:

	Italiani monolingui	Bilingui	Brasiliani monolingui
dislocazioni a sinistra	12	17	0
Temi sospesi	1	2	4
Topicalizzazioni tematiche	2	7	38
TOTALE	15	26	42

**Tabella 2:** Frequenza e strategia di tematizzazione nei corpora di monolingue e di bilingue.

È importante osservare due grandi differenze: la prima è relativa alla frequenza delle tematizzazioni, molto maggiore in PB che in italiano, e alla tendenza da parte dei bilingui di attestarsi su frequenze intermedie, il che è un segno di erosione. La seconda differenza riguarda le strategie messe in atto per realizzare la tematizzazione. Il PB predilige chiaramente la topicalizzazione, senza nessun legame sintattico tra l'elemento dislocato e il resto della frase: solo l'intonazione segnala la relazione funzionale tra i due costituenti, come nell'esempio *As freiras, a gente morria de rir* (Le suore, noi morivamo dal ridere); al contrario l'italiano predilige la dislocazione con la ripresa anaforica del pronome, come nell'esempio *I vestiti, li ho comprati allo shopping*. La terza strategia possibile, il tema sospeso, costituisce una sorta di strategia intermedia, con la ripresa pronominale ma senza la marca di caso, come nell'esempio *Giovanni, gli ho fatto un bel regalo*. Quanto ai bilingui, pur notandosi una tendenza all'indebolimento delle relazioni sintattiche, è rilevante che

<sup>9</sup> CORPUS NURC-RJ = Projeto norma linguística urbana culta - RJ base di dati. Disponibile all'indirizzo <<http://www.lettras.ufrj.br/nurc-rj/>>.

la strategia di gran lunga dominante rimanga la dislocazione. Di conseguenza Raso e Vale concludono che i bilingui tendono ad aumentare la frequenza delle tematizzazioni, ma a mantenere la struttura di tematizzazione tipica dell'italiano. Ciò spiega il motivo dell'aumento dei pronomi con funzione anaforica di costituenti dislocati: ciò che aumenta non è l'uso del pronome in sé, ma la frequenza di strutture tematizzanti che inducono l'uso del pronome. L'erosione riguarda quindi la strutturazione dell'enunciato; l'aumento dei pronomi ne è una conseguenza.

Per quanto riguarda la diminuzione oltre la media dei clitici cataforici di costituenti dislocati a destra, la spiegazione risiede, secondo Raso (2009), ancora una volta nell'interfaccia tra sintassi e pragmatica. Mentre nel PB la dislocazione a destra viene segnalata solamente dal *break* prosodico (il pronome sopravvive solo in registri molto formali), in italiano la situazione è più complessa. L'obbligatorietà della catafora pronominale (con o senza *break* prosodico) è legata alla struttura argomentale del verbo. Un esempio evidente è dato dai casi di verbi con più strutture argomentali, come *mangiare*, che ha una struttura a due argomenti (qualcuno mangia qualcosa) o a un argomento (mangiare come antonimo di essere a digiuno). Se si usa il verbo nel significato a due argomenti, la catafora è obbligatoria:

*L'hai mangiato, il pollo?*<sup>10</sup>

*Sì, l'ho mangiato, il pollo*

Se invece si usa la struttura a un argomento solo, la dislocazione può essere marcata solo prosodicamente:

*Hai mangiato o sei ancora a digiuno?*

*Ho mangiato, il pollo.*

Inoltre Raso (2009) afferma che la dislocazione a destra può essere confusa con l'eco marcato prosodicamente, che è un fenomeno diverso, anche se strutturalmente molto simile:

*dimmi quando Carlo ha attraversato la strada*

*ecco, adesso ha attraversato, la strada*

10 La virgola segnala il *break* prosodico, che tuttavia non è obbligatoria in italiano.

in cui l'ultimo costituente è pura eco, e non elemento dislocato. Il suo valore informativo è nullo, e si ha solo la marca prosodica.

Il modello brasiliano non pone queste sottili distinzioni, inducendo quindi una perdita della complessità dell'italiano a favore della forma strutturalmente più semplice, senza la catafora pronominale. Ciò spiegherebbe la sua enorme diminuzione nel *corpus* Raso-Vale.

Le conclusioni di Raso e Vale (2009) e di Raso (2009) sono che ogni clítico oggetto di studio ha presentato segnali di erosione linguistica, ma in misura diversa. Essi affermano che, da una parte, l'interferenza del PB può spiegare alcuni fenomeni, ma che, dall'altra, alcuni fenomeni non sono spiegabili con l'interferenza ma sembrano chiamare in causa tendenze generali di ristrutturazione.

### 3. Il nuovo *corpus* Raso-Ferrari: caratteristiche

La ricerca che viene presentata qui ha l'obiettivo di verificare e approfondire la metodologia e l'affidabilità dei risultati delle ricerche precedenti. Un nuovo *corpus* di bilingui è stato raccolto e trascritto in formato CHAT<sup>11</sup>. Gli informatori posseggono le stesse caratteristiche di quelli del *corpus* precedente, tranne per la durata della permanenza in Brasile, che è di gran lunga inferiore: almeno 20 anni nel *corpus* precedente e almeno 8 in quello nuovo. A differenza del *corpus* Raso-Vale, di cui erano disponibili solo le registrazioni della parte raccolta presso l'UFMG, questo nuovo *corpus*, che chiameremo Raso-Ferrari, è corredato da tutte le registrazioni. Il *corpus*, raccolto tra il 2008 e il 2009, è composto da otto testi per un totale di 21.298 parole, con otto informatori diversi. Di questi, due sono gli stessi che compaiono nel *corpus* Raso-Vale, giacché uno dei testi è stato rivisto e riutilizzato. Il *corpus* si compone di cinque dialoghi, due monologhi e una conversazione, presentando una diversificazione situazionale molto maggiore del *corpus* Raso-Vale, costituito solo da interviste, quasi tutte sullo stesso argomento. Nella tabella sotto vengono esplicitate le caratteristiche dei vari testi:

11 MAC WHINNEY (1994) e MAC WHINNEY (2000).

Titolo	numero totale di parole	numero di parole per informatore	Tipologia e descrizione dell'interazione
Pranzo	4408	MRC: 1843 MAS: 2565	Dialogo tra due partecipanti durante un pranzo a casa di MAS: gli argomenti di conversazione variano dal calcio, al vizio delle sigarette e ad un'ampia discussione sui vini (LAF è presente ma parla poco). Registrato con microfono omnidirezionale.
Partita 1	1284	UCR: 533 MAS: 751	Dialogo tra UCR e MAS e poi tra UCR e LAF mentre aspettano l'inizio di una partita in televisione, a casa di MAS. Il primo dialogo riguarda commenti su partite di calcio e persone conosciute; il secondo dialogo è su cosa si fa nei fine settimana. Registrato con microfono omnidirezionale.
Partita 2	3578	UCR: 1348 MAS: 2230	Conversazione tra MAS, UCR e ALG (informatore madrelingua non adeguato alla ricerca) e LAF: sono commenti fatti durante una partita di calcio in televisione, a casa di MAS. Registrato con microfono omnidirezionale.
Sorelle	3691	PAT: 1194 LIV: 2497	Dialogo, diviso in due parti, tra due sorelle. Il primo parla di libri e lezioni di italiano. Il secondo sul cambiamento di casa di LIV e sulla nuova decorazione dell'appartamento. Registrato con un registratore a cassette.
Missionari	1764	MON: 1764	Monologo: è un'intervista ad una missionaria sulla sua esperienza in Brasile (LAF, l'intervistatrice, interviene molto poco). L'intervista è stata fatta nella sede della parrocchia. Registrato con microfono omnidirezionale.
Genitori	1480	ANG: 1480	Monologo con sporadici interventi di altri partecipanti che non sono adeguati alla ricerca. ANG racconta del suo trasferimento in Brasile con la famiglia e delle difficoltà iniziali. Registrato nella parrocchia dove ANG lavora. Registrato con microfono omnidirezionale.
Medici	2144	GIC: 2144	Dialogo tra GIC e LAF (la partecipazione di GIC è molto superiore a quella di LAF), a casa di GIC, nel quale GIC commenta sulla sua esperienza con i medici e racconta vari episodi sull'argomento. Registrato con microfoni a clip.
Cena	2949	MAS: 2949	Dialogo tra MAS e LAF (la presenza di LAF è piuttosto estesa) nella loro residenza. Mentre preparano la cena fanno commenti sugli ingredienti ed il modo di cottura, il risultato della ricetta e vari altri argomenti. Registrato con microfoni a clip.

**Tabella 3:** informazioni sul corpus e ognuno dei testi.

Gli informatori sono stati identificati con sigle; le caratteristiche riguardanti l'età sono state suddivise in base a diverse fasce: dai 18 ai 25 anni, fascia 1; dai 26 ai 39 anni, fascia 2; dai 40 ai 60 anni, fascia 3; oltre i 60 anni, fascia 4. Nella tabella di seguito riportiamo le caratteristiche sociolinguistiche dei partecipanti:

Informatore	Sesso	Età	Provenienza	Titolo di studi	Tempo di residenza	Professione	Torna in Italia
MAS	M	fascia 2	Brescia	scuole superiori e università in Italia	8 anni	insegnante e traduttore	al massimo ogni due anni
UCR	M	fascia 3	Varese	scuole superiori in Italia, università in Brasile	12 anni	insegnante	quasi mai
GIC	M	fascia 3	Torino	scuole superiori in Italia e università (non terminata) in Italia	15 anni	traduttore e grafico	quasi mai
MON	F	fascia 3	Oristano	scuole superiori e università in Italia	11 anni	missionaria	ogni due o tre anni
ANG	M	fascia 4	Bergamo	scuole superiori e università (non terminata) in Italia	20 anni	ragioniere e missionario	quasi tutti gli anni
MRC	M	fascia 3	Milano	scuole superiori e università (non terminata) in Italia; laurea in Brasile	19 anni	insegnante	quasi mai
LIV	F	fascia 3	Chiasso	scuole superiori in Italia, laurea in Brasile	33 anni	insegnante in pensione	ogni tanto
PAT	F	fascia 3	Chiasso	scuole superiori in Italia, laurea e dottorato in Brasile	33 anni	insegnante universitaria	ogni tanto

**Tabella 4:** dati sociolinguistici sugli informatori del *corpus* Raso-Ferrari.

### 3.1 . Limiti del *corpus* Raso-Ferrari

Nella compilazione di questo *corpus* si è cercato di seguire le indicazioni metodologiche della linguistica di *corpus*<sup>12</sup> per quanto riguarda bilanciamento e rappresentatività, ma le difficoltà incontrate non sono state totalmente superate, prima fra tutte il numero limitato degli informatori che rispondono alle caratteristiche richieste e siano disponibili ad autorizzare la registrazione.

Inoltre, come si è detto, uno degli obiettivi della realizzazione di un nuovo *corpus* era quello di poter disporre di dati ricavati da diafasie diverse da quella dell'intervista e realmente spontanee. Registrare interazioni in contesto naturale è di per sé più difficile, ma la difficoltà aumenta enormemente quando si devono registrare situazioni in cui si usa un codice diverso dalla L1 del paese in cui l'interazione si svolge. Bisogna infatti non solo disporre della situazione adatta con gli informatori adatti, ma anche assicurarsi che non sia presente nessuno che faccia mutare il codice comunicativo dalla lingua che si intende studiare alla lingua di contatto. Ciò automaticamente restringe enormemente le situazioni disponibili. Sono stati organizzati alcuni incontri appositi: pranzi, cene, visioni di partite in televisione, oltre alla visita nei domicili o luoghi di lavoro degli informatori. In questo modo è stato possibile ottenere registrazioni per una durata complessiva di circa quattro ore e mezza e 35.000 parole. Tuttavia, per migliorare il bilanciamento, si è preferito operare tagli al materiale raccolto, equilibrando la durata dei testi e riducendo il peso di un informatore che appare più volte. Nonostante ciò, 8.495 parole appartengono ad un unico informatore. Per quanto possibile, abbiamo avuto comunque l'accortezza di tener conto di tutti questi fattori nell'analisi e nell'interpretazione dei dati.

### 3.2 . Risultati

Come si è detto, clitici analizzati sono gli stessi studiati da Raso e Vale (2009). Come *corpus* di paragone si è invece usato un sottocorpus del C-ORAL-ROM italiano, scegliendo quattordici testi tra i più simili ai nostri per tipologia di interazione e livello sociolinguistico dei partecipanti, per un totale di 21.224 parole. I dati dei due *corpora* sono stati normalizzati, dividendoli per il numero di parole dei ri-

12 Citiamo come riferimenti di base: BIBER (1993); ATKINS, CLEAR, OSTÇER (1992); BIBER, CONRAD, REPPEN (1998); MC ENERY, XIAO, TONO (2006); MCENERY; WILSON (1996); MEYER (2004); SARDINHA (2004).

spettivi *corpora* ed in seguito moltiplicandoli per 10.000, ottenendo in questo modo dati confrontabili. Le due tabelle sotto permettono una visualizzazione generale dei risultati ottenuti e una loro comparazione con quelli della ricerca precedente:

Clitici	Corpus Raso-Ferrari		C-ORAL-ROM It		Corpus Raso-Vale		BADIP	
	Valori assoluti	Valori normalizzati	Valori assoluti	Valori normalizzati	Valori assoluti	Valori normalizzati	Valori assoluti	Valori normalizzati
<i>Ci</i> attualizzanti	135	63,38	274	129,09	117	64,71	107	59,18
<i>Ci</i> lessicalizzanti	10	4,69	22	10,36	3	1,65	10	5,53
<i>Ci</i> locativo	3	1,4	19	8,95	24	13,27	39	21,57
<i>lo, la, li, le, l'</i>	237	111,27	310	146,06	166	91,81	304	168,14
<i>Ne</i> (totali)	32	15,02	43	20,26	14	7,74	29	16,03
TOTALE	407	191,09	646	304,37	324	179,18	489	270,46

**Tabella 5:** Occorrenze dei clitici nei *corpora* di monolingui e di bilingui.

Clitici	Variazione percentuale Corpus Raso-Ferrari /C-ORAL-ROM It	Variazione percentuale Corpus Raso-Vale/BADIP	Variazione percentuale Corpus Raso-Ferrari/BADIP	Variazione percentuale Corpus Raso-Vale/C-ORAL-ROM It
<i>Ci</i> attualizzanti	-50,91	9,34	7,09	-49,87
<i>Ci</i> lessicalizzanti	-54,72	-70,16	-15,18	-84,07
<i>Ci</i> locativo	-84,22	-38,47	-93,5	48,26
<i>lo, la, li, le, l'</i>	-23,81	-45,39	33,82	-37,14
<i>Ne</i> (totali)	-25,86	-51,71	-6,3	-61,79
TOTALE	-37,21	-33,74	-29,34	-41,13

**Tabella 6:** Variazione delle occorrenze dei clitici nei *corpora* di monolingui e di bilingui.

Guardando le due tabelle con attenzione, il dato più macroscopico appare essere non tanto la differenza tra i due *corpora* di bilingui ma quella tra i due *corpora* di paragone. Se infatti si paragonano da una parte il *corpus* Raso-Vale col BADIP e dall'altra il *corpus* Raso-Ferrari con il C-ORAL-ROM, si ha l'impressione che le due ricerche giungano a conclusioni molto diverse: per esempio, il *ci* attualizzante diminuisce enormemente nel paragone tra Raso-Ferrari e C-ORAL-ROM, mentre aumenta leggermente nel paragone tra Raso-Vale e BADIP. Se invece paragoniamo i due *corpora* di bilingui allo stesso *corpus* di monolingui, sia esso il BADIP o il C-ORAL-ROM, il quadro è molto più coerente. Fatta questa premessa, passiamo all'esposizione dei dati.

La diminuzione dei clitici nel complesso è confermata, ma in misura minore a quella del *corpus* Raso-Vale. Rispetto al BADIP, Raso-Vale riducevano i clitici del 33,74%, mentre Raso-Ferrari li diminuiscono del 29,34. Bisogna però dire che la differenza è in realtà maggiore di quanto sembri, dal momento che i *ci* locativi si riducono radicalmente nel *corpus* Raso-Ferrari (-93,5% contro una diminuzione di appena 38,47% da parte di Raso-Vale, se paragonati con il BADIP); ciò confermerebbe quanto già previsto negli studi precedenti, e cioè che la tipologia dei testi del *corpus* Raso-Vale induceva un numero di *ci* locativi più alto del normale. Queste considerazioni di ordine generale sembrerebbero indicare che l'erosione, anche per la morfosintassi, continua a crescere al di là del primo decennio di contatto, e non si stabilizza come sostenuto da molti. Ciò, come vedremo, è confermato dal resto dei dati.

Nel nostro studio, il *ci* attualizzante è stato separato da quello lessicalizzante. Chiariamo che in questa ricerca consideriamo i soli *esserci* e *averci* come attualizzanti, mentre tutti gli altri verbi che si uniscono al clitico *ci* saranno definiti lessicalizzanti, indipendentemente dal grado di lessicalizzazione in cui essi si trovino, sia esso incipiente o totalmente grammaticalizzato<sup>13</sup>.

Il *corpus* Raso-Vale, che era stato confrontato con il BADIP, presentava la stessa frequenza di *ci* attualizzanti, ma con una distribuzione molto diversa dal BADIP per quanto riguarda *averci* e *esserci*, mentre nel caso del *ci* lessicalizzante vi era una drastica diminuzione del 70,16% che, nel *corpus* Raso-Ferrari, è di appena il 15,18. Infine il totale dei *ne* conferma la stessa tendenza: rispetto al BADIP, nel *corpus* Raso-Vale la diminuzione è del 51,71%, mentre in quello attuale è del 6,3%. Passiamo ora a un'analisi più accurata.

13 Si veda a questo proposito la discussione sui valori di *ci* unito a verbi, promossa da Russi (2008) e Sabatini (1985 e 1986).

### 3.2.1 . Il pronome *ci* con valore locativo

Come è stato visto sopra, il *ci* locativo è stato tra i clitici quello che ha rilevato la maggiore perdita nel *corpus* Raso-Ferrari. Tuttavia si ricorda che gli stessi Raso e Vale (2009) avevano previsto che, con un *corpus* differente da quello da loro analizzato, il calo sarebbe stato molto più alto. Ciò sarebbe stato dovuto al fatto che il *corpus* precedente era sostanzialmente costituito da interviste semistrutturate, in cui l'argomento principale era il trasferimento degli informatori dall'Italia al Brasile e il racconto delle loro esperienze nei due paesi. È evidente che questi argomenti inducono un uso frequente di espressioni locative e la loro conseguente ripresa anaforica con il pronome *ci*. Aver trovato meno *ci* in Raso-Ferrari rispetto a Raso-Vale non deve quindi essere considerato un elemento di maggiore erosione, ma semplicemente un effetto naturale e atteso della mancanza di variazione negli argomenti dei testi del primo *corpus*.

### 3.2.2 . Il pronome *ne* nelle varie funzioni

Anche nel caso del clitico *ne* si è ritenuto opportuno confrontare i nostri risultati con quelli di Raso e Vale (2009), distinguendone le funzioni e normalizzando tutti i dati.

Clitici	<i>Corpus</i> Raso-Ferrari		C-ORAL-ROM It		<i>Corpus</i> Raso-Vale		<i>BADIP</i>	
	Valori assoluti	Valori normalizzati	Valori assoluti	Valori normalizzati	Valori assoluti	Valori normalizzati	Valori assoluti	Valori normalizzati
Ne partitivo	19	8,92	28	13,19	6	3,31	16	8,84
Ne argoment	13	6,1	15	7,06	8	4,42	10	5,53
Ne locativo	0	0	0	0	0	0	3	1,65
Ne (totali)	32	15,02	43	20,26	14	7,74	29	16,03

**Tabella 7:** Occorrenze dei clitici *ne* nelle loro varie funzioni nei *corpora* di monolingui e di bilingui.

Clitici	Variazione Percentuale <i>Corpus</i> Raso-Ferrari /C-ORAL-ROM It	Variazione percentuale <i>Corpus</i> Raso-Vale/BADIP	Variazione percentuale <i>Corpus</i> Raso-Ferrari/BADIP	Variazione percentuale <i>Corpus</i> Raso-Vale/C-ORAL-ROM It
Ne partitivo	-32,37	-62,55	0,9	-74,9
Ne argoment	-13,59	-20,07	10,3	-37,39
Ne locativo	0	-100	-100	0
Ne (totali)	-25,86	-51,71	-6,3	-61,79

**Tabella 8:** Variazione della frequenza dei clitici *ne* nelle loro varie funzioni nei *corpora* di monolingui e di bilingui.

È evidente come anche i numeri del *corpus* Raso-Ferrari siano alquanto ridotti, ma maggiori comunque rispetto a quelli del *corpus* Raso-Vale. Nel caso dei *ne* totali si passa da 7,74 occorrenze ogni 10.000 parole di Raso-Vale a praticamente il doppio, 15,02 del *corpus* attuale. Ancor più evidente è la differenza per il *ne* partitivo con 3,31 occorrenze nel primo *corpus* contro 8,92 nel secondo. Il *ne* argomentale fa rilevare una differenza meno evidente: 4,42 occorrenze nel *corpus* Raso-Vale e 6,1 in quello Raso-Ferrari. Infine non vi sono differenze nei due *corpora* per quanto riguarda il *ne* locativo, che non compare in nessuno dei due. Vogliamo ricordare che nello studio precedente il BADIP aveva fatto riscontrare solamente tre occorrenze di *ne* locativo e nel C-ORAL-ROM italiano da noi analizzato non ne abbiamo trovata nessuna. Ciò evidenzia la differenza esistente tra i due *corpora* di paragone, diversi per periodo ma anche per tipologia testuale e livello di spontaneità. Difatti il C-ORAL-ROM italiano, essendo più recente e costruito sulla base di situazioni diversificate di parlato effettivamente spontaneo, mostra la realtà attuale del parlato italiano in cui, come sostiene Russi (2008), il *ne* locativo è assai poco diffuso, giacché compare praticamente solo nella lessicalizzazione di verbi in combinazione col clitico *si*. Il fatto che la funzione partitiva continui a prevalere su quella argomentale e ancor più su quella locativa conferma l'ipotesi secondo la quale sarebbe più facile perdere ciò che si acquisisce dopo.

### 3.2.3 . Il pronome *ci* con valore attualizzante

Passiamo ora all'analisi dei dati relativi al clitico *ci* con valore attualizzante, ossia nei casi di *esserci* ed *averci*. Chiariamo che il verbo *esserci* è stato ulteriormente suddiviso tra valore *esistenziale*, ossia quando sostituisce altre forme lessicali che esplicitano una funzione di esistenza<sup>14</sup>, e valore *presentativo*, formato dal verbo *esserci*+*SN*+*che pseudo-relativo*, il quale serve ad alleggerire dal punto di vista cognitivo una struttura completamente nuova e rematica<sup>15</sup>. I dati presentati sono stati tutti normalizzati e quelli del *corpus* Raso-Vale sono stati recentemente rivisti da Vale (2009) nelle suddette funzioni, per cui sono leggermente diversi da quelli presentati anteriormente. Le tabelle sotto presentano il confronto tra i due *corpora* di bilingui.

Clitici	Corpus Raso-Ferrari		C-ORAL-ROM It		Corpus Raso-Vale		BADIP	
	Valori assoluti	Valori normalizzati	Valori assoluti	Valori normalizzati	Valori assoluti	Valori normalizzati	Valori assoluti	Valori normalizzati
esserci esisten.	103	48,36	167	78,68	90	49,77	44	24,33
esserci present.	16	7,51	16	7,53	17	9,04	13	7,19
esserci totale	119	55,87	183	86,22 / 66,9*	107	59,18	57	31,52
averci	16	7,51	91	42,87	10	5,53	50	27,65

\* Dati forniti da Panunzi (2010) sull'intero C-ORAL-ROM italiano di 300.000 parole.

**Tabella 9:** Occorrenze dei *ci* attualizzanti.

14 Si veda Russi (2008).

15 Per intenderci, la struttura *c'è un signore che vuole parlare con te* è cognitivamente più leggera di quella *un signore vuole parlare con te*, dal momento che entrambe le porzioni (il referente *un signore* e ciò che di esso si predica) sono cognitivamente nuove, e la struttura presentativa, staccando le due porzioni, fa in modo che la prima costituisca un appoggio cognitivo per la seconda.

Clitici	Variazione percentuale <i>Corpus</i> Raso-Ferrari /C-ORAL-ROM It	Variazione percentuale <i>Corpus</i> Raso-Vale/BADIP	Variazione percentuale <i>Corpus</i> Raso-Ferrari/BADIP	Variazione percentuale <i>Corpus</i> Raso-Vale/C-ORAL-ROM It
esserci existen.	-38,53	104,56	98,77	-36,74
esserci present.	-0,27	25,73	4,45	20,05
esserci totale	-35,21 / -16,48*	87,75	77,25	-31,36 / -11,54*
averci	-82,48	-80	-72,84	-87,1

\* Dati forniti da Panunzi (2010) sull'intero C-ORAL-ROM italiano di 300.000 parole.

**Tabella 10:** Variazione nella frequenza dei *ci* attualizzanti.

Uno degli aspetti che più saltano all'occhio è che, normalizzando i dati di *esserci* con valore esistenziale, non sembrano constatabili grandi differenze tra le occorrenze del *corpus* Raso-Vale e quello Raso-Ferrari, al contrario di ciò che accade se paragoniamo i due *corpora* di monolingui. Anche nel caso degli *esserci* con valore presentativo non si riscontrano grandi differenze. In questo caso, anche il rapporto fra i due *corpora* di monolingui è costante, a conferma che questa funzione informativa non dipende da variabili di tipologia testuale.

Va comunque spiegata l'altissima occorrenza di *esserci* esistenziale nel C-ORAL-ROM rispetto al BADIP. Difatti Raso e Vale avevano spiegato con l'erosione lessicale il forte aumento di *esserci* esistenziale in rapporto al BADIP. A una maggiore difficoltà nel richiamare la disponibilità lessicale si sarebbe ovviato aumentando l'occorrenza di lessico generico; spia evidente di ciò sarebbe il forte aumento degli *esserci* esistenziali in grado di sostituire la variabilità delle forme di esistenza. Questa spiegazione è compatibile con la ancor più alta occorrenza di *esserci* nel C-ORAL-ROM? A nostro parere non vi è contraddizione tra i due dati. Difatti il C-ORAL-ROM, essendo un *corpus* molto più spontaneo e colloquiale del BADIP, non poteva non presentare un'occorrenza maggiore di forme generiche. Al contrario, i testi dei due *corpora* di bilingui, nonostante lo sforzo del *corpus* Raso-Ferrari di aumentare la spontaneità e la colloquialità delle situazioni comunicative, rimangono probabilmente più comparabili con il BADIP. Infine, va considerato che, se si considerano i dati, molto più estesi, presenti in Panunzi (2010), la differenza tra

BADIP e C-ORAL-ROM è ben più ridotta rispetto a quanto appare dal campione da noi esaminato.

Nel caso di *averci*, ciò che accade è in controtendenza rispetto agli altri fenomeni esaminati. Infatti, contrariamente a quanto Raso e Vale prevedevano, anche nel nuovo *corpus* di bilingui, con informatori giunti in Brasile in tempi più recenti, si conferma una fortissima riduzione di *averci*. Non si può quindi spiegare il fenomeno, come era stato fatto in precedenza, con il fatto che questa forma era stata accettata nella nuova norma in tempi più recenti. Si tratta senza dubbio di un aspetto da approfondire. Tuttavia, si può avanzare una considerazione che potrebbe aiutare a spiegare il fenomeno. Quasi tutti gli informatori del *corpus* Raso-Ferrari, e una parte di quelli del *corpus* Raso-Vale, sono docenti di italiano. È possibile che ciò induca a una sorvegliatezza maggiore e a una forma di evitamento di forme più colloquiali, come è il caso di *averci*, in assenza di *input* costante da parte della comunità di parlanti nativi. Un docente di italiano all'estero è indotto a controllare la propria produzione e probabilmente a evitare forme che sono o possono essere percepite come troppo colloquiali, soprattutto se manca loro l'*input* della comunità di L1, che mostra costantemente come queste forme siano ormai accettate anche in registri di formalità media. In ogni caso, questo risultato mostra come sia difficile, proprio perché soggetto a un grande numero di variabili, lo studio dell'erosione linguistica.

### 3.2.4 . Il pronome *ci* con valore lessicalizzante

Sarà ora esposta la situazione, nei vari *corpora*, relativa al pronome *ci* con valore lessicalizzante, ossia unito a un verbo che così viene modificato semanticamente in maniera più o meno forte. Normalizzando i dati, si nota, come segnalato nelle Tabelle 5 e 6, che, rispetto al BADIP, la perdita di *ci* lessicalizzante è del 70% nel *corpus* Raso-Vale e solo del 15% nel *corpus* Raso-Ferrari, confermando, ancora una volta, che l'erosione non sarebbe un processo che si esaurisce dopo il primo decennio<sup>16</sup>. Il confronto fra i due *corpora* di monolingui mostra ancora una volta differenze evidenti fra i *corpora* di paragone, con il C-ORAL-ROM che presenta una frequenza di *ci* lessicalizzanti doppia rispetto al BADIP. Le tabelle 11 e 12 qui sotto mostrano invece quali sono i *types* che compaiono nei 4 *corpora* e la loro frequenza

16 Cfr. Soesman (1997), Waas (1996) e le affermazioni di Hutz (2004) sull'inconclusività delle ricerche del peso che ha il fattore tempo nel processo di erosione. A questo proposito, si veda anche Keijzer (2007).

in *tokens*. È facile notare come l'erosione, nella misura illustrata sopra, si manifesti sia con una diminuzione dei lessemi sia con una riduzione delle occorrenze.

<i>Types lessicali</i>	<i>Corpus Raso-Ferrari</i>	<b>C-ORAL-ROM italiano</b>
<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
<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
<b>TOTALE TYPES</b>	<b>7</b>	<b>12</b>
<b>TOTALE TOKENS</b>	<b>10</b>	<b>22</b>

**Tabella 11:** *Types* lessicalizzanti del clitico *ci* e occorrenze dei *tokens* nei *corpora* Raso-Ferrari e C-ORAL-ROM italiano.

<i>Types lessicali</i>	<i>Corpus Raso-Vale</i>	<b>BADIP</b>
<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
<b>TOTALE TYPES</b>	<b>3</b>	<b>6</b>
<b>TOTALE TOKENS</b>	<b>3</b>	<b>10</b>

**Tabella 12:** *Types* lessicalizzanti del clitico *ci* e occorrenze dei *tokens* nei *corpora* Raso-Vale e BADIP.

### 3.2.5 . Gli accusativi di terza persona

Passiamo ora all'analisi dei pronomi accusativi di terza persona. La tabella sotto mostra i dati riscontrati nel *corpus* Raso-Ferrari ed il confronto con quelli dello studio Raso-Vale, con la relativa suddivisione dei clitici in base alle loro funzioni di anaforici di costituenti dislocati a sinistra, cataforici di costituenti dislocati a destra e quando non sono forici di costituente dislocato.

Clitici	<i>Corpus</i> Raso-Ferrari		C-ORAL-ROM It		<i>Corpus</i> Raso-Vale		<i>BADIP</i>	
	Valori assoluti	Valori normalizzati	Valori assoluti	Valori normalizzati	Valori assoluti	Valori normalizzati	Valori assoluti	Valori normalizzati
lo, la, li, le, l' non forici di costit disloc.	202	94,84	246	115,9	124	68,58	253	139,93
lo, la, li, le, l' anaforici di costit. disloc a sinistra	23	10,79	29	13,66	29	16,03	23	12,72
lo, la, li, le, l' cataforici di costit disloc a destra	12	5,63	35	16,49	13	7,19	28	15,48
lo, la li, le, l' TOTALI	237	111,27	310	146,06	166	91,81	304	168,14

**Tabella 13:** Occorrenze dei pronomi accusativi nelle varie funzioni.

Clitici	Variazione Percentuale <i>Corpus</i> Raso-Ferrari/C-ORAL-ROM It	Variazione percentuale <i>Corpus</i> Raso-Vale/ <i>BADIP</i>	Variazione percentuale <i>Corpus</i> Raso-Ferrari/ <i>BADIP</i>	Variazione percentuale <i>Corpus</i> Raso-Vale/C-ORAL-ROM It
lo, la, li, le, l' non forici di costit disloc.	-18,17	-50,98	-32,22	-40,82
lo, la, li, le, l' anaforici di costit. disloc a sinistra	-21,01	26,02	-15,17	17,34
lo, la, li, le, l' cataforici di costit disloc a destra	-65,85	-53,55	-63,63	-56,39
lo, la li, le, l' TOTALI	-23,81	-45,39	-33,82	-37,14

**Tabella 14:** Variazione di frequenza dei pronomi accusati vi nelle varie funzioni.

Analizziamo prima i dati relativi ai due *corpora* di monolingui. Questa volta essi si mostrano abbastanza coerenti nei dati, soprattutto nelle funzioni informativamente marcate, come anafora o catafora di costituenti dislocati. La maggiore occorrenza di accusativi nel BADIP non è statisticamente significativa. Consideriamo ora i due *corpora* di bilingui quanto ai dati relativi agli accusativi non forici di costituente dislocato, giacché sono pronomi, per così dire, informativamente neutri, non dovuti, cioè, a particolari strategie informative. In questo caso si mantiene la tendenza generale che abbiamo osservato finora: il *corpus* Raso-Ferrari conferma l'esistenza di un processo di erosione, ma in misura chiaramente inferiore a quella del *corpus* Raso-Vale. Di nuovo, si può sostenere che l'erosione non sarebbe un processo che si conclude dopo il primo decennio. Essa, al contrario, continua a manifestarsi anche dopo e in maniera evidente.

Il discorso si fa invece più complesso se esaminiamo ciò che accade a proposito degli accusativi forici di costituenti dislocati. Discutiamo i dati dei due *corpora* di bilingui in rapporto al BADIP. Gli studi di Raso e Vale avevano mostrato un aumento considerevole dei pronomi anaforici di costituenti dislocati a sinistra, al contrario di ciò che accadeva per i pronomi non forici di costituenti dislocati, che diminuivano molto, e soprattutto dei pronomi cataforici di dislocazioni a destra, che diminuivano ancor di più. Ora, invece, i dati del *corpus* Raso-Ferrari, mentre confermano le attese per quanto riguarda i pronomi non forici di dislocazioni, presentano risultati diversi per quanto riguarda le forme foriche di costituenti dislocati. I pronomi anaforici di costituenti dislocati a sinistra, che nel *corpus* Raso-Vale aumentavano, ora diminuiscono, sia pure in percentuali minori rispetto agli altri accusativi, mentre quelli cataforici di costituenti dislocati a destra diminuiscono ancora di più che nel *corpus* Raso-Vale. Bisogna concludere che le analisi operate da Raso e Vale vengono smentite dai dati del nuovo *corpus* o esiste una spiegazione compatibile con questi nuovi dati e con le analisi degli studi anteriori? A nostro avviso una spiegazione esiste, e potrà facilmente essere verificata operando un confronto fra testi scelti per questo obiettivo.

I costituenti tematizzati, che in italiano vengono ripresi con il pronome anaforico, hanno una precisa funzione informativa: instaurare per il destinatario un referente cognitivo che costituisca l'ambito semantico di riferimento di un'illocazione, cioè di un atto linguistico<sup>17</sup>. Se un parlante deve realizzare un'illocazione, deve ritenere che l'ascoltatore sappia a quale dominio semantico tale illocazione si ap-

17 Per una teoria *corpus based* del riferimento cognitivo dell'illocazione, si veda Cresti (2000).

plica. Se, per esempio, dovessimo dire, a proposito di questa ricerca, che l'abbiamo realizzata a partire da una tesi di "mestrado", non potremmo dire semplicemente: *l'abbiamo realizzata a partire da una tesi di "mestrado"*. Ciò sarebbe possibile solo se si sta già parlando della ricerca, ma non se la ricerca non è, per così dire, attiva nel modello discorsivo degli interlocutori. In questo caso dovremmo prima instaurare il referente *questa ricerca* nel modello discorsivo, e solo poi potremmo compiere un'ilocuzione che si riferisca ad essa. Una maniera molto comune per farlo sarebbe dire: *questa ricerca, l'abbiamo realizzata a partire da una tesi di "mestrado"*. In tal modo tematizzeremmo il costituente che rappresenta l'ambito di applicazione dell'ilocuzione, per poi riprenderlo con il pronome durante la realizzazione dell'ilocuzione stessa. In sintesi, la tematizzazione è frequente quando è necessario instaurare un referente che serva da ambito di applicazione dell'ilocuzione che si intende eseguire, mentre è molto meno frequente quando questo costituente è già cognitivamente instaurato. Se, per esempio, mentre siamo con una persona, arriva un comune amico di nome, poniamo, *Carlo*, non è naturale dire *Carlo, l'ho invitato a cena*, ma piuttosto diremmo semplicemente *l'ho invitato a cena* oppure *ho invitato Carlo a cena*. Non sarebbe, cioè, normale instaurare, tematizzandolo, un referente cognitivamente già instaurato situazionalmente.

Ora, se paragoniamo i testi del *corpus* Raso-Vale a quelli del *corpus* Raso-Ferrari, notiamo che nel primo gli intervistati raccontano di esperienze proprie, in cui i referenti non sono noti all'interlocutore, ma devono essere instaurati perché, a proposito di essi, si possano compiere atti linguistici. Lo stesso accade per i testi del NURC utilizzati da Raso e Vale per confrontare le strategie di tematizzazione. Al contrario, nel *corpus* Raso-Ferrari gran parte dei testi sono testi dialogici che avvengono in situazione: persone che assistono a una partita di calcio in televisione o che cenano insieme o che discutono liberamente. Ciò, pragmaticamente, aumenta moltissimo la quantità di illocuzioni che fanno riferimento a referenti dati dalla situazione stessa. Per esempio, un atto linguistico che si riferisce a un calciatore che ha appena sbagliato una giocata semplice non viene realizzata instaurando il riferimento al calciatore e poi dicendo su di lui qualcosa, ma l'ilocuzione è realizzata direttamente, perché il riferimento è ovvio. Oppure se a cena viene offerto un dolce particolare, non sarà necessario dire: *questo dolce, l'hai fatto tu o l'hai comprato?* Si dirà direttamente: *l'hai fatto tu o l'hai comprato?* Il contesto pragmatico che prevale

nei due *corpora* spiega quindi la maggiore frequenza di tematizzazioni (e di conseguenza di pronomi anaforici) nel *corpus* Raso-Vale, dove la situazione comunicativa gioca un ruolo marginale a paragone con il *corpus* Raso-Ferrari.

Le dislocazioni a destra possono avere più funzioni: sia quella di riprendere un costituente già dato situazionalmente sia, più raramente, quella di instaurare, per così dire, in ritardo un referente solo parzialmente dato e che quindi opportuno disambiguare. L'effetto della situazione è minore che nelle tematizzazioni con dislocazioni a sinistra, ma esiste. Ciò spiegherebbe quindi una diminuzione delle dislocazioni a destra nel *corpus* Raso-Ferrari, e anche la misura più contenuta di tale diminuzione.

## 4 . Conclusioni

La ricerca si poneva i seguenti obiettivi:

1. Costituire un nuovo *corpus*, con caratteristiche in parte uguali e in parte diverse dal precedente, per aumentare e diversificare i dati relativi all'erosione linguistica degli italiani colti in contatto prolungato col PB.
2. Cambiare il *corpus* di paragone e verificare se i dati offerti da esso fornivano una conferma, e fino a che punto, ai dati del BADIP.
3. Verificare i risultati delle ricerche di Raso e Vale, sia per controllare la tenuta dei dati sia per vedere come reagivano a un *corpus* con caratteristiche diverse.

I risultati ottenuti sono interessanti sia per le conferme che offrono, sia perché mostrano che metodologie diverse consentono di vedere più in profondità problemi già in parte noti, sia per le prospettive di studio che aprono.

In primo luogo, i due *corpora* di monolingui mostrano divergenze nei dati della maggior parte dei pronomi. Metodologicamente ciò significa che la composizione dei *corpora* condiziona in maniera significativa i dati. La comparabilità dei dati non è un fatto puramente quantitativo ma anche qualitativo. Quanto accade per i pronomi accusativi anaforici di dislocamenti a sinistra è solo uno degli esempi di ciò che questa ricerca offre. L'aver aumentato il *corpus* Raso-Vale con l'aggiunta di quello Raso-Ferrari deve quindi essere visto solo come il primo passo per la costruzione di un più ampio *corpus* che permetta analisi più affidabili e presenti situazioni più

diversificate, tali da costituire subcorpora che consentano di vedere la dipendenza dei dati linguistici dai contesti pragmatici e sociolinguistici.

L'aver costruito un nuovo *corpus* con informatori che presentino le stesse caratteristiche di quelli del *corpus* precedente ad eccezione del tempo di contatto con la L2 e di carenza di *input* della L1 ci ha permesso di formulare un'ipotesi teorica importante. Sembra, infatti, e lo confermerebbero tutti i dati considerati, che l'erosione prosegua nel tempo anche oltre il primo decennio, non soltanto nel lessico, come si sostiene nella maggioranza delle ricerche, ma ben più a fondo nella struttura linguistica. La minore erosione in praticamente tutti i pronomi esaminati riscontrata nel *corpus* Raso-Ferrari non sembra spiegabile diversamente.

Inoltre, ciò che accade con le forme di *averci* nel *corpus* Raso-Ferrari mette in crisi la spiegazione offerta per la forte riduzione di queste forme nel *corpus* Raso-Vale. Aver riscontrato gli stessi dati in un *corpus* i cui informatori non possedevano le caratteristiche chiamate in causa per spiegare il fenomeno impone che si continui la ricerca su altri dati. Una spiegazione che sembrava naturale e che non aveva trovato obiezioni di sorta è stata confutata da dati raccolti con criteri diversi. La spiegazione qui avanzata in via puramente ipotetica e come eventuale concausa deve essere verificata e integrata.

Al contrario, proprio un *corpus* diverso ha permesso di confermare le ipotesi degli studi precedenti a proposito dei *ci* locativi.

In conclusione, se l'erosione pronominale come fenomeno generale viene in sostanza confermata, se si osserva che il fenomeno si protrae nel tempo ben oltre i primi dieci anni di contatto e di assenza di *input*, e se si comprende con chiarezza quanto *corpora* diversi inducano esiti diversi, si mostra anche la necessità di continuare la raccolta di dati con variabili mirate, non solo per aumentare i dati ma anche per renderli sensibili a fattori sociolinguistici diversi.

## Riferimenti bibliografici

- ATKINS S.; CLEAR J.; OSTLER N. Corpus Design Criteria. *Literary and Linguistic Computing*, Oxford, v. 7, n. 1, 1992, pp. 1-16.
- BERRETTA M. I pronomi clitici nell'italiano parlato. In: HOLTUS; RADKE (Orgs.). *Gesprochenes Italienisch in Geschichte und Gegenwart*. Tübingen: Narr, 1985, pp. 185-504.

- BERRETTA, M. 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, pp. 329-352.
- BIBER, D. Representativeness. In *Corpus Design. Literary and Linguistic Computing*, Oxford, v. 8, n. 4, 1993, pp. 1-15.
- BIBER, D.; CONRAD, S.; REPPEN R. *Corpus Linguistics: Investigating Language structure and Use*. Cambridge: CUP, 1998.
- CORDIN P.; CALABRESE, A. I pronomi personali. In: RENZI, L. *Grande grammatica di consultazione*. Bologna: Il Mulino, 1988, v. I., pp. 535-592.
- CRESTI, E. *Corpus di italiano parlato*. Firenze: Accademia della Crusca, 2000. 2v.
- CRESTI, E.; MONEGLIA, M. (Orgs.). *C-ORAL-ROM, Integrated Reference Corpora for Spoken Romance Language*. Amsterdam-Philadelphia: John Benjamin, 2005.
- DE MAURO, T. et al. *Lessico di frequenza dell'italiano parlato*. Milano: EtasLibri, 1993.
- GIACALONE RAMAT, A. Italiano di stranieri. In: SOBRERO, A. (Org.). *Introduzione all'italiano contemporaneo: la variazione e gli usi*. Roma-Bari: Laterza, 1993, pp. 341-410.
- HUTZ, M. Is there a natural process of decay? A longitudinal study of language attrition. In: KOPKE, B.; SCHMID, M. S. *First Language Attrition. Interdisciplinary perspectives on methodological issues*. Amsterdam/Philadelphia: John Benjamin, 2004, pp. 189-206.
- ISURIN, L. Deserted island or a child's first language forgetting. *Bilingualism: Language and Cognition*, Cambridge University Press, n. 3, 2000, pp. 151-166.
- JAKOBSON, R. *Il farsi e il disfarsi del linguaggio: linguaggio infantile e afasia*. Torino: Einaudi, 1971 (1941).
- KELJZER, M. *Last in first out? An investigation of the regression hypothesis in Dutch emigrants in Anglophone Canada*. Amsterdam: Vrije Universiteit, 2007.
- KOPKE, B.; SCHMID, M.S. *First Language Attrition: Interdisciplinary perspectives on methodological issues*. Amsterdam/Philadelphia: John Benjamin Publishing Company, 2004.
- LEONE, F. *I pronomi personali di terza persona: l'evoluzione di un microsistema nell'italiano di fine millennio*. Roma: Carocci, 2003.
- MACWHINNEY, B. *The CHILDES project: tools for analyzing 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.
- 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.
- PANUNZI, A. *La variazione semantica del verbo essere nell'italiano parlato: uno studio su corpus*. Firenze: Firenze University Press, 2010.
- SARDINHA, T. B. *Linguística de corpus*. Baruerí: Manole, 2004.
- SOESMAN, A. An experimental study on native language attrition in Dutch adult immigration in Israel. In: J.

- KLATTER-FOLMER; KROON, S. (Eds.). *Dutch overseas: Studies in maintenance and loss of Dutch as an immigrant language* Tilburg: Tilburg University Press, 1997, pp. 181-194.
- SHARWOOD SMITH, M.; VAN BUREN, P. First language attrition and the parameter setting model. In: SELINGER, H. W. & VAGO, R. M. *First Language Attrition*. Cambridge: CUP, 1991, pp. 17-31.
- RASO, T. L'italiano parlato a San Paolo da madrelingua colti: primi sondaggi e ipotesi di lavoro. *Revista de Italianística*, São Paulo, n. 8, 2003, pp. 9-49.
- \_\_\_\_\_. Erosione dei clittici e strutture tematizzanti in italiani colti in contatto prolungato col portoghese brasiliano. In: CONGRESSO DELLA SOCIETÀ INTERNAZIONALE DI LINGUISTICA E FILOLOGIA ITALIANA, 10, 2009, Basileia. *Atti: Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione, giustapposizione*. Basileia. Società Internazionale di Linguistica e Filologia Italiana, 2009, pp. 384-399.
- RASO, T.; VALE, H. P. A erosão linguística em italianos cultos em contato prolongado com o português do Brasil: os clíticos e alguns efeitos na estrutura do enunciado. *Revista de Italianística*, v. 16, 2009, pp. 1-22.
- REVISTA DE ITALIANÍSTICA. São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas da USP, n. 5, ano V, 1997.
- 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. In: HOLTUS, G; RÄDKE, E. (Orgs.). *Gesprochenes Italienisch in Geschichte und Gegenwart*. Tübingen: Narr, 1985, pp. 154-184.
- VALE, H. P. *A erosão linguística dos italianos cultos em contato prolongado com o português do Brasil: os clíticos*. Monografia (Graduação) - Faculdade de Letras, Universidade Federal de Minas, Belo Horizonte, 2007.
- VALE, H. P. A erosão dos clíticos verificada em um novo corpus: *esserci, averci e ci* lexicalizante. 2009. (Apresentação de Trabalho Comunicação). XIII Congresso da ABPI, VIII Encontro Internacional de Estudos Italianos, II Jornada de Italianística da América Latina. Belém, 14 a 19 de setembro de 2009.
- WAAS, M. *Language Attrition Downunder: German Speakers in Australia*. Frankfurt: Peter Lang, 1996.