

I SOGGETTI CLITICI IN FIORENTINO: UN'ANALISI DIACRONICA *CORPUS BASED*

TOMMASO RASO *

LÚCIA DE ALMEIDA FERRARI* *

ABSTRACT: L'articolo tratta della questione della supposta perdita dei clitici soggetto nel vernacolo fiorentino attraverso uno studio *corpus based*. I lavori su questo argomento (GALLI DE' PARATESI, 1984; SOBRERO, 1997; BERRUTO, 1997, 2005; BINAZZI, 1997, 1998, 2007; CARDINALETTI & MUNARO, 2009) sostengono che la pressione dell'italiano standard, a partire dagli anni del secondo dopoguerra, abbia contribuito alla perdita, da parte del vernacolo fiorentino, di molte delle sue caratteristiche. Tuttavia, fino a questo momento, non era stata effettuata nessuna indagine di tipo quantitativo e qualitativo sul sistema pronominale basata su dati. L'opportunità per tale tipo di lavoro è stata resa possibile grazie alla pubblicazione rimasterizzata del *Corpus Stammerjohann* (1970) e del *Corpus per il Confronto Diacronico* (SCARANO, 2005), entrambi *corpora* di lingua parlata spontanea, registrati a Firenze a quarant'anni circa di distanza. Il lavoro qui presentato ha analizzato l'espressione dei clitici soggetto di terza persona singolare *gl', l', la, e* in concomitanza con il verbo *essere* al presente indicativo. I risultati raggiunti in questo primo momento del lavoro indicano una riduzione nell'uso dei clitici soggetto nel tempo, ma fanno pensare anche ad una retrodatazione del periodo di inizio della perdita, data l'età dei parlanti. Le analisi dovranno continuare in futuro con un maggior numero di dati affinché il fenomeno possa essere meglio compreso.

*Docente all'Universidade Federal de Minas Gerais (UFMG), ricercatore CNPq, Capes e Fapemig.

**Dottoranda all'Universidade Federal de Minas Gerais (UFMG), borsista Capes.



PAROLE CHIAVE: vernacolo fiorentino; clítics soggetto; *corpus*; parlato spontaneo.

RESUMO: O artigo trata da questão da suposta perda de clítics sujeito no vernáculo florentino através de um estudo baseado em *corpus*. Os trabalhos sobre este assunto (GALLI DE' PARATESI, 1984; SOBRERO, 1997; BERRUTO, 1997, 2005; BINAZZI, 1997, 1998, 2007; CARDINALETTI & MUNARO, 2009) afirmam que a pressão do italiano padrão, a partir dos anos do segundo pós-guerra, contribuiu para a perda, por parte do vernáculo florentino, de muitas de suas características. Contudo, até o presente momento não havia sido feita qualquer investigação de tipo quantitativo e qualitativo sobre o sistema pronominal baseada em dados. A oportunidade para tal tipo de trabalho foi possível graças à publicação remasterizada do *Corpus Stammerjohann (1970)* e do *Corpus per il Confronto Diacronico (SCARANO, 2005)*, ambos corpora de língua falada espontânea gravados em Florença com cerca de quarenta anos de distância. O trabalho que aqui apresentamos analisou a expressão dos clítics sujeito de terceira pessoa do singular *gl'*, *l'*, *la*, e em concomitância com o verbo *essere (ser)* no presente do indicativo. Os resultados obtidos nesta primeira etapa do trabalho apontam para uma redução no uso dos clítics sujeito no tempo, mas também fazem vislumbrar a retrodatação do período de início de tal perda, devido à idade dos informantes. As análises deverão continuar no futuro com um maior número de dados a fim de poder compreender melhor o fenômeno.

PALAVRAS-CHAVE: vernáculo florentino; clítics sujeito; *corpus*; fala espontânea.

ABSTRACT: The article analyzes the supposed loss of subject clitics in vernacular Florentine through a corpus-based study. Works about this theme (GALLI DE' PARATESI, 1984; SOBRERO, 1997; BERRUTO, 1997, 2005; BINAZZI, 1997, 1998, 2007; CARDINALETTI & MUNARO, 2009) claim that the pressure of standard

*Italian, from the years after the Second World War, contributed to the loss of many of the characteristics of vernacular florentine. However, until now, there was no quantitative nor qualitative data-based investigation on the pronominal system. The occasion for this kind of work was possible thanks to the remastered publication of Corpus Stammerjohann (1970) and of Corpus per il Confronto Diacronico (SCARANO, 2005), both of them spontaneous spoken corpora taped in Florence with approximately forty years of time distance. The investigation that we present analyzed the expression of the third person singular subject clitics *gl'*, *l'*, *la*, *e* with the present indicative tense of the verb *essere* (to be). Results obtained in this first stage of the research indicate that there was a reduction in the use of subject clitics over time, but also suggest that the beginning of that loss must be dated back, due to the generation of the informants. Researches should proceed in the future with a larger amount of data to better understand this phenomenon.*

KEYWORDS: *vernacular Florentine; subject clitics; corpus; spontaneous speech.*

1. Introduzione

Il portoghese brasiliano (PB), con particolare evidenza in alcune delle sue varietà, ha sviluppato una serie di soggetti clitici, tra cui *cê* che è il caso più studiato. Tuttavia, il percorso di grammaticalizzazione delle forme clitiche con funzione di soggetto in PB è molto discusso e certamente non è compiuto. Non esiste infatti una doppia serie di forme, una tonica e una clitica, che si distingua chiaramente a livello segmentale. Questa considerazione ci ha stimolato a uno studio contrastivo che ci permettesse di comprendere meglio lo statuto del rapporto fra *você* e *cê* (FERRARI, 2013), anche grazie al confronto con lo studio dei pronomi clitici con funzione di soggetto presenti nel fiorentino.

La realizzazione di questa doppia ricerca è ora possibile grazie all'esistenza di 3 *corpora* di parlato spontaneo che permettono di osservare i fenomeni in questione con nuovi strumenti. Si

tratta del *corpus* C-ORAL-BRASIL (RASO & MELLO, 2012), per quanto riguarda il PB, e dei *corpora Stammerjohann* restaurato e allineato (SCARANO & SIGNORINI, 2003) e C-ORAL-ROM (CRESTI & MONEGLIA, 2005), per quanto riguarda l'analisi diacronica dei pronomi fiorentini.

In questo lavoro ci limitiamo a esporre i primi risultati relativi allo studio diacronico dei clitici soggetto nel fiorentino.

1.1 Il vernacolo fiorentino e i soggetti clitici

La tradizione degli studi linguistici italiani sostiene che la forte pressione dell'italiano standard, a partire dagli anni del secondo dopoguerra, abbia contribuito in modo sostanziale alla perdita, da parte del vernacolo fiorentino, di molte delle sue peculiarità. Ciò sarebbe avvenuto specialmente a causa della scolarizzazione obbligatoria e della grande diffusione dei mass media, in particolar modo della televisione (GALLI DE' PARATESI, 1984; SOBRERO, 1997; BERRUTO, 1997, 2005; BINAZZI, 1997, 1998, 2007; CARDINALETTI & MUNARO, 2009).

Ci soffermeremo in questa ricerca su alcuni aspetti del particolare sistema del fiorentino (RENZI, 1983, 1992; GIANNELLI, 2000), ovvero l'espressione obbligatoria del soggetto e la possibilità di esprimerlo attraverso un soggetto clitico, a volte reduplicato. In modo specifico ci occuperemo qui dell'espressione dei clitici di terza persona singolare: *gl'*, *l'*, *la*, *e*.

È nota la forte tendenza del fiorentino ad esprimere il soggetto, contrariamente a quanto avviene per l'italiano standard, in cui esso è marcato pragmaticamente (PALERMO, 1997; RENZI, 1983). La motivazione sintattica del soggetto in fiorentino si traduce spesso nella presenza di un soggetto pronominale, che può essere costituito da un clitico. Di fatto, l'espressione pronominale del soggetto presenta una doppia serie: una tonica, come l'italiano, ed una clitica; esse sono chiaramente distinte sotto il profilo segmentale. Inoltre, la serie clitica può accompagnare sia il soggetto lessicale che quello pronominale, e addirittura quello clitico nella cosiddetta reduplicazione del clitico, come negli esempi fittizi che proponiamo sotto:

1. La Maria va a casa.
2. Lei va a casa.
3. La Maria/lei *la* va a casa.
4. *E la* va a casa.

Nell'esempio (3) il clitico *la* riprende il soggetto lessicale (la Maria) o quello pronominale tonico (lei). Nell'esempio (4) si ha la reduplicazione del clitico con i pronomi *e* e *la*, entrambi clitici, in funzione di soggetto (RENZI, 1983).

Non siamo tuttavia a conoscenza di studi che verifichino le affermazioni sulla perdita delle strutture clitiche a partire da dati di parlato consultabili, anche perché ciò implica la necessità di *corpora* di parlato di tipo diacronico.

Tale opportunità, probabilmente unica, è stata resa possibile grazie ai lavori svolti dal laboratorio linguistico LABLITA del Dipartimento di Italianistica dell'Università degli Studi di Firenze, diretto da Emanuela Cresti. Presso LABLITA, nel 2001, sono state rimasterizzate, analizzate, ricomposte e allineate le registrazioni effettuate dallo studioso tedesco Harro Stammerjohann nel 1965, nelle prossimità del quartiere di Santo Spirito a Firenze (SCARANO & SIGNORINI, 2003). Per realizzare ricerche in prospettiva diacronica, nel 2005 è stato preparato il *Corpus per il Confronto Diacronico* (SCARANO, 2005) utilizzando testi estratti dal C-ORAL-ROM italiano (CRESTI & MONEGLIA, 2005) e dai *corpora* LABLITA (CRESTI, 2000). È stato così possibile avere a disposizione materiale rappresentativo di due momenti cruciali della storia linguistica fiorentina: il primo intorno alla metà degli anni Sessanta, quando la pressione dell'italiano standard era da poco avviata, e l'altro circa quarant'anni dopo, con testi comparabili del periodo che va dal 2000 al 2005 circa, quando la standardizzazione avrebbe dovuto essere conclusa o comunque molto avanzata.

Sinora gli studi di tipo diacronico si sono limitati all'analisi del lessico (MONEGLIA et al., 2008). Il nostro obiettivo è verificare cosa accade all'espressione del soggetto, in particolare di quello clitico della terza persona singolare, ponendo i due *corpora* a confronto.

1.2 Il paradigma clitico della terza persona singolare

Prima di procedere alla descrizione dei *corpora* e della metodologia adottata in quest'analisi, è bene soffermarci sul paradigma canonico dell'espressione del soggetto in fiorentino. Gli studi classici di italiano e varietà locali sono quelli di Boström (1972), di Brandi e Cordin (1981, 1983, 1989), e di Renzi e Vanelli (1983). Tuttavia, riprenderemo qui più in dettaglio le analisi fatte da Renzi (1983, 1992) e da Giannelli (2000), che si occupano direttamente della questione del fiorentino in prospettiva sincronica e diacronica.

Come è noto, in italiano i pronomi soggetto sono tonici e facoltativi, in quanto motivati pragmaticamente. In fiorentino invece il soggetto pronominale sarebbe obbligatorio (RENZI,

1983), in secondo luogo esso può essere espresso da un clitico che occupa una posizione fissa dinanzi al verbo, o da uno o più clitici che a loro volta precedono il verbo. La formula è quella che segue (RENZI, 1983, p. 224):

5. (pron, SN) X cl V

dove i due simboli tra parentesi sono opzionali e si escludono a vicenda; *X* indica la possibilità di inserzione di materiale lessicale; *cl* indica il clitico e *V* il verbo. Spieghiamo la formula con un esempio fittizio alla terza persona singolare, che è quella che verrà analizzata in questa ricerca:

- 6. **Lui** oggi **l'**ha mangiato.
- 7. **l' bambino** oggi **l'**ha mangiato.
- 8. **Lei** oggi **la** preparerà il pranzo.
- 9. **La ragazza** oggi **la** preparerà il pranzo.

Nell'esempio (6) il pronome soggetto *lui* viene ripreso dal clitico soggetto *l'* (che può essere usato solo davanti a vocale), mentre nell'esempio (7) è il SN *il bambino* ad essere ripreso dal clitico *l'*. Gli esempi (8) e (9) mostrano la stessa situazione nel caso del femminile, ripreso dal clitico *la*.

Nel caso della negazione, il pronome clitico viene immediatamente preceduto da essa. Siccome prende un posto fisso, la negazione stessa diventa un clitico.

10. **Lui 'un l'**ha cantato.

Si ha qui il pronome soggetto *lui* duplicato dal clitico *l'*, preceduto dalla negazione *'un*. L'eccezione è la terza persona femminile, in cui l'ordine viene invertito:

11. **(Lei) la 'un** canta.

dove il soggetto pronominale *lei* viene ripreso dal clitico *la*, questa volta seguito dalla negazione *'un*.

Infine è necessario chiarire la questione del clitico *e*¹. Gli studiosi sono concordi nell'affermare che il suo uso è praticamente sparito da tempo nel parlato.

Analizzandone la forma sintattica, Renzi (1983) sostiene che la sua funzione come soggetto è piuttosto recente ma il suo uso non è mai obbligatorio. Questa forma può comparire con tutte le persone *e*, sempre in maniera facoltativa, può essere incontrata con gli impersonali, i meteorologici e nella costruzione con il soggetto posposto. Una sua particolarità è che, nel caso sia presente la *e*, non compare il pronome tonico, mentre per quello clitico può avvenire la reduplicazione. Di seguito vengono esemplificati i suoi usi:

12. **e** parlo / **(e)** tu parli / **(e)** la parla / **(e)** 'un parliamo / **(e)** **vu** parlate / **e** parlano
13. **(e)** fa caldo / **(e)** bisogna fare presto.
14. **(e)** piove / **(e)** tuonò.
15. **(e)** **la** mi fa male **la** testa.
16. **E gl'**è andato a casa.
17. **E la** 'un capisce niente.

La parentesi negli esempi sopra indica facoltatività. L'esempio (12) mostra la forma *e* come soggetto di tutte le persone: nel caso della prima singolare l'unico clitico esistente è *e*; nella seconda il clitico è *tu*, giacché in fiorentino la forma tonica è *te*; per la terza persona è presente il clitico femminile *la*; per la prima plurale, non essendo possibile utilizzare la forma tonica, il clitico eventualmente è la negazione 'un; la forma *vu* è quella clitica per la seconda persona plurale; infine, la terza plurale ha come forma clitica, davanti a vocale, *e*.

Gli esempi (13) e (14), rispettivamente, mostrano *e* come soggetto facoltativo di verbi impersonali e meteorologici.

Nell'esempio (15) si ha la costruzione con il soggetto posposto: il clitico *e* è facoltativo, seguito dal clitico *la*, reduplicazione del soggetto posposto SN *la* testa.

Gli esempi (16) e (17) mostrano la reduplicazione del clitico al maschile (16), con le forme *e* e *gl'*, entrambi clitici soggetto, e al femminile (17) con le forme *e* e *la*, clitici soggetto, seguite dal clitico di negazione 'un.

Concludendo, il paradigma dei clitici soggetto di terza persona è come segue:

¹ In questo testo utilizzeremo la grafia *e*, senza apostrofo, la stessa che compare nei *corpora* consultati e negli scritti della maggior parte degli studiosi. Negli scritti più antichi e in citazioni di essi è possibile trovare la grafia *e'*, con apostrofo, che tuttavia si riferisce ad un'altra origine pronominale (si veda a questo proposito Renzi, 1983, p. 226).

MASCHILE	FEMMINILE
<i>gl'</i> (dinanzi a vocale)	<i>l'</i> (dinanzi a vocale)
<i>e</i> (dinanzi a consonante)	<i>la</i> (dinanzi a consonante)

Tabella 1: forme clitiche di terza persona singolare nel vernacolo fiorentino.

Passeremo ora alla metodologia utilizzata, prima con la descrizione dei *corpora* consultati e poi con le tappe seguite in questo lavoro.

2. Metodologia

2.1 I corpora utilizzati

Il *Corpus Stammerjohann* è considerato il primo esempio di *corpus* di parlato spontaneo in lingua italiana (e non solo). Dal 29 gennaio al 26 febbraio 1965, lo studioso tedesco Harro Stammerjohann, allora dottorando in romanistica, si recò a Firenze, dove effettuò 30 registrazioni per un totale di più di 40 ore. Il suo obiettivo era raccogliere dati per uno studio sull'italiano parlato a Firenze. Il presidente dell'Accademia della Crusca dell'epoca, Giovanni Nencioni, lo presentò a varie famiglie di artigiani e professionisti del quartiere popolare di Santo Spirito e dintorni. Fatto totalmente inedito in quegli anni, le registrazioni volevano cogliere la lingua viva, prodotta spontaneamente in contesto naturale, in interazioni che ne documentassero le varietà d'uso più significative. Per ottenere ciò, Stammerjohann si avvale di equipaggiamenti assai sofisticati per quegli anni, il che ha garantito, tra l'altro, la preservazione di gran parte del materiale raccolto fino alla sua rimasterizzazione nel 2001. Il totale di più di 40 ore di audio è composto per la metà da registrazioni ambiente: il registratore fu collocato in un punto nascosto di un barbiere in via Faenza mentre i clienti entravano ed uscivano. È stato possibile dunque ottenere registrazioni in situazioni comunicative differenziate, e su diversi temi. L'altra metà delle registrazioni è composta da un testo per ogni audio, senza cambiamenti di situazione o di interlocutori. Queste ultime sono state realizzate in ambiente familiare/privato o pubblico, e illustrano vari argomenti: i generi prevalenti sono l'intervista e la conversazione in contesto familiare.²

² La descrizione dettagliata di tutti i testi, partecipanti e durata delle registrazioni può essere incontrata in Scarano e Signorini (2003), responsabili per l'attualizzazione del *corpus*.

I preziosi documenti che Stammerjohann ha lasciato³ e donato ad Emanuela Cresti perché li mettesse a disposizione della comunità scientifica non costituivano un vero e proprio *corpus*. La sua trasformazione in questo formato, come abbiamo detto, avvenne nel 2001, ad opera del laboratorio LABLITA. Furono riversati in digitale gli audio delle bobine, convertiti in formato *.wavPCM*. Tutto il materiale fu quindi rimasterizzato ed ascoltato integralmente per suddividere i testi e classificarli secondo le stesse caratteristiche, sociolinguistiche e di architettura, adottate nel C-ORAL-ROM⁴ (CRESTI & MONEGLIA, 2005).

Scartando gli audio danneggiati e obbedendo a criteri di corretto bilanciamento e rappresentatività diafasica, è stato così compilato un *corpus* di 86.252 parole e 32 testi suddivisi in contesto familiare, privato e pubblico. Ogni contesto è ulteriormente ripartito in base al tipo di interazione, regolata o non regolata⁵, ognuna delle quali è nuovamente distribuita secondo la struttura dell'evento comunicativo: monologo, dialogo e conversazione. Tutti i testi sono trascritti secondo gli stessi criteri del C-ORAL-ROM e allineati al suono tramite il software *WinPitch* (MARTIN, 2005). In questo modo i ricercatori hanno accesso ad un *corpus* con dati del 1965 in cui possono essere effettuate ricerche con le moderne risorse informatiche e con la possibilità di verificare facilmente l'audio allineato al testo.

Sotto inseriamo i dettagli del *corpus* per numero di testi, parole e tipologie di interazione:

3 L'autore ha lasciato un'unica pubblicazione basata sul *corpus*: Stammerjohann (1970), pubblicata sulla rivista *Studi di Filologia Italiana*, in cui compaiono alcune trascrizioni delle registrazioni.

4 I dettagli dei criteri e procedimenti adottati si trovano in: Scarano e Signorini (2003) e Signorini e Tucci (2004).

5 L'interazione è regolata quando vi sono norme sociali che stabiliscono lo scambio di turno tra i parlanti, come in situazioni familiari convenzionalizzate o di lavoro; non regolata è quando non è possibile prevedere la risposta dell'interlocutore e lo scambio di turno avviene in maniera casuale. Nella pratica, le interazioni regolate corrispondono alla situazione dell'intervista, mentre quelle non regolate ad interazioni libere.

CONTESTO	TIPO	INTERAZIONE	NUMERO DI TESTI
FAMILIARE 22.517 parole	NON-REGOLATO 22.517 parole	MONOLOGO: 0 parole	0
		DIALOGO: 9.310 parole	2
		CONVERSAZIONE: 13.207 parole	4
PRIVATO 36.817 parole	NON-REGOLATO 25.606 parole	MONOLOGO: 0 parole	0
		DIALOGO: 0 parole	0
		CONVERSAZIONE: 25.606 parole	8
	REGOLATO 11.211 parole	MONOLOGO: 3.023 parole	1
		DIALOGO: 8.188 parole	4
		CONVERSAZIONE: 0 parole	0
PUBBLICO 23.196 parole	NON-REGOLATO 16.278 parole	MONOLOGO: 0 parole	0
		DIALOGO: 0 parole	0
		CONVERSAZIONE: 16.278 parole	4
	REGOLATO 6.918 parole	MONOLOGO: 1.561 parole	3
		DIALOGO: 5.357 parole	4
		CONVERSAZIONE: 0 parole	0
TELEFONICO 3.722 parole		3.722 parole	2
TOTALE		86.252 PAROLE	32

Tabella 2: Struttura del *Corpus Stammerjohann*, con numero di parole e tipologie dei testi ed interazioni.

Perché fosse possibile utilizzare in modo proficuo ed in prospettiva diacronica il *Corpus Stammerjohann*, presso LABLITA è stato compilato nel 2005 il *Corpus per il Confronto Diacronico* (SCARANO, 2005; MONEGLIA et al., 2008; SCARANO & SIGNORINI, 2003; SCARANO & MONEGLIA, 2006). Sono stati selezionati testi provenienti dal C-ORAL-ROM italiano e dai *corpora* LABLITA in modo da replicare la struttura, il numero di parole e le

varie suddivisioni di testi e variazioni diafasiche del *Corpus Stammerjohann*. I ricercatori del LABLITA hanno quindi dovuto fare una cernita delle situazioni le cui particolarità diafasiche, diastratiche e di contenuto garantissero un uso il più vicino possibile alla varietà del fiorentino.

Il nuovo *corpus* è riuscito a mantenere la struttura originale ed approssimativamente anche il numero di parole, 85.908. L'unica variante è il totale di testi, che ammonta a 55 (SCARANO & SIGNORINI, 2003; SIGNORINI & TUCCI, 2004).

2.2 La ricerca

Per la ricerca che viene qui presentata abbiamo utilizzato tutti i testi del *Corpus Stammerjohann*. Tuttavia, siccome il *Corpus per il Confronto Diacronico* è composto da contributi di parlanti in genere toscani, ma non tutti fiorentini e, dato che ciò che è di nostro interesse è un fenomeno che ha declinazioni diverse anche all'interno della Toscana, è stato necessario operare un taglio nel *corpus*, scartando i parlanti che non presentavano le caratteristiche che ci interessavano, cioè che non fossero propriamente fiorentini. In questo modo abbiamo estratto 40 dei 55 testi del *corpus*, per un totale di 58.299 parole⁶, e abbiamo chiamato il *corpus* risultante *Corpus di Fiorentini*. Per ovvi motivi, il bilanciamento originale è stato in parte compromesso ma, nell'elaborazione delle conclusioni sulle ricerche svolte, questo fattore verrà tenuto in considerazione e i suoi effetti verranno commentati.

La tabella 3 riporta i dati di questo nuovo *corpus*.

⁶ Per i dettagli di tale scelta e le specificazioni dei testi utilizzati si veda Ferrari, L. A. (Tese de Doutorado, UFMG, em preparação).

CONTESTO	TIPO	INTERAZIONE	NUMERO DI TESTI
FAMILIARE 21.232 parole	NON-REGOLATO 21.232 parole	MONOLOGO: 0 parole	0
		DIALOGO: 9.415 parole	6
		CONVERSAZIONE: 11.817 parole	6
PRIVATO 24.862 parole	NON-REGOLATO 16.641 parole	MONOLOGO: 0 parole	0
		DIALOGO: 0 parole	0
		CONVERSAZIONE: 16.641 parole	12
	REGOLATO 8.221 parole	MONOLOGO: 1.958 parole	2
		DIALOGO: 6.263 parole	4
		CONVERSAZIONE: 0 parole	0
PUBBLICO 8.862 parole	NON-REGOLATO 7.308 parole	MONOLOGO: 0 parole	0
		DIALOGO: 0 parole	0
		CONVERSAZIONE: 7.308 parole	4
	REGOLATO 1.554 parole	MONOLOGO: 0 parole	0
		DIALOGO: 1.554 parole	1
		CONVERSAZIONE: 0 parole	0
TELEFONICO 3.343 parole		3.343 parole	5
TOTALE		58.299 PAROLE	40

Tabella 3: Struttura del *Corpus di Fiorentini*, con numero di parole e tipologie dei testi ed interazioni.

La tabella 4 invece comprende i dati essenziali sui quali abbiamo lavorato:

	<i>CORPUS STAMMERJOHANN</i>	<i>CORPUS DI FIORENTINI</i>
N. PAROLE	86.252	58.299
N. TESTI	32	40

Tabella 4: Dati sui *corpora* utilizzati nella ricerca.

La ricerca si è concentrata, come è stato detto, sui pronomi soggetto clitici di terza persona singolare: *gl'*, *la*, *l'*, *e*. Per individuare le forme all'interno dei *corpora* è necessario verificare tutti i verbi alla terza persona singolare, giacché, come si è visto, sono vari i modi di esprimere il soggetto. Si è ritenuto quindi che sarebbe stato poco produttivo lavorare su tutti i verbi del *corpus*, sia per motivi di tempo, vista l'enormità del lavoro che ciò comporta, sia perché l'opzione per verbi ad alta frequenza e con varietà di valori semantici può permettere di raggiungere comunque risultati affidabili e rappresentativi dell'insieme. La scelta è ricaduta, in questo primo momento della ricerca, sul verbo *essere* al presente indicativo. Questo verbo è il secondo lemma a maggior frequenza riscontrato all'interno del C-ORAL-ROM italiano. Uno studio *corpus-driven* di Panunzi (2010) sulla sua variazione semantica ne ha individuate circa 14.000 occorrenze. Ciò equivale a 4,6% del TOTALE di *tokens* del C-ORAL-ROM italiano e fa di *essere* il verbo a più alta frequenza della lingua italiana. Oltre alla sua importanza dal punto di vista quantitativo, dallo studio di Panunzi (2010) sono emerse importanti considerazioni che hanno fatto ricadere la nostra scelta su questo verbo: non solo *essere* può venire utilizzato come ausiliare e come verbo autonomo, ma va fatta anche una distinzione tra *essere* e l'altro lessema che da esso si origina, cioè *esserci*. Inoltre, gli usi verbali di *essere* sono vari e qualitativamente diversi: i principali sono quello copulare, quello identificativo e quello predicativo a base locativa⁷. Per il lavoro sono stati esclusi per il momento sia *esserci* sia, naturalmente, *essere* con funzione di ausiliare.

⁷ La questione è piuttosto complessa e lunga, per cui rimandiamo a Panunzi (2010). Esemplicando brevemente, essere con valore copulare compare in frasi del tipo: "Questa è una bella giornata". Con valore identificativo si avrebbe: "Lui è l'avvocato"; e con valore predicativo a base locativa: "Qui è casa mia".

3. I dati

La nostra ricerca non è ancora conclusa e qui presenteremo solo i dati riguardanti i clitici in concomitanza con la forma *è* utilizzata come verbo non ausiliare.

Sono state analizzate 1.977 occorrenze della forma nel *Corpus Stammerjohann*, di cui 1.126 sono risultate appartenere a *essere* come verbo non ausiliare. Nel *Corpus di Fiorentini* abbiamo riscontrato un totale di 1.347 occorrenze della forma, di cui 697 appartenenti a *essere* come verbo non ausiliare.

Lo spoglio dei dati nei due *corpora* è stato effettuato utilizzando il software *AntConc*. Il programma rileva la forma ricercata insieme al contesto. Nella tabella sotto riportiamo i risultati incontrati. Siccome il numero totale di parole dei due *corpora* è diverso, per poterli confrontare i dati sono stati normalizzati per 100.000, ossia, nella colonna di destra il lettore incontrerà i valori ogni 100.000 parole.

Forme	CORPUS STAMMERJOHANN		CORPUS DI FIORENTINI	
	Occorrenze assolute	Occorrenze Normalizzate	Occorrenze assolute	Occorrenze Normalizzate
TOTALE delle occorrenze “è” verbo autonomo	1136	1317	697	1196
<i>L'</i>	238	276	133	228
<i>Gl'</i>	44	51	39	67
<i>La</i>	2	2	1	2
<i>E</i>	0	0	0	0
Raddoppiamento clitico <i>e l'</i>	2	2	0	0
Raddoppiamento clitico <i>e gl'</i>	2	2	0	0
Raddoppiamento clitico <i>la l'</i>	1	1	0	0
Sogg. <i>Æ</i>	382	443	244	419

Soggetto lessicale preverbale	187	217	129	221
Soggetto lessicale postverbale	118	137	93	160
Soggetto Pronome Dimostrativo	158	183	129	221
Pronome tonico	17	20	12	21

Tabella 5: Forme di soggetto incontrate nei due corpora a confronto, in termini assoluti e normalizzati.

Riportiamo ora, per ogni tipologia di soggetto, esempi tratti dai due *corpora*, affinché il lettore possa meglio rendersi conto di ciò di cui stiamo parlando.

- Soggetto *l'* clitico:

18. *ART: [...] // [431] / no / **l'** è / hhh Castiglion della Pescaia // (Corpus Stammerjohann, arti)⁸

19. *LUC: [392] [<] <ma bellina> / guarda // [393] vedessi come **l'** è bellina // (Corpus Confronto Diacronico, fam1)

- Soggetto *gl'* clitico:

20. *ART: [469] non so se **gl'** è Porto Santo Stefano // (Corpus Stammerjohann, arti)

21. *LOR: [279] <eh> // [280] ho vinto / ora **gl'** è poco // (Corpus Confronto Diacronico, ipubdl03)

- Soggetto *la* clitico:

⁸ Spieghiamo brevemente la simbologia adottata nel *corpus*: tra parentesi è indicato il *corpus* e la sigla del testo di provenienza dell'esempio. I vari simboli sono quelli utilizzati nella famiglia C-ORAL-ROM: la barra semplice (/) denota l'apertura e la chiusura di unità tonale; la barra doppia (//) la frontiera prosodica di finale di enunciato. Il simbolo (+) specifica che l'enunciato è stato interrotto; parole tra i simboli (<>) designano sovrapposizioni nel parlato; i numeri tra parentesi quadre [] indicano il numero dell'enunciato; tre lettere maiuscole dopo un asterisco la sigla del parlante; il simbolo ([/no]) indica *retracting* o errore nell'esecuzione dell'enunciato; quello (&), inizio di parola interrotta; (&he) è utilizzato in caso di esitazione; (hhh) quando vi è un fenomeno paralinguistico. Per evitare di citare enunciati troppo lunghi, nei casi in cui ne sono state soppresse alcune parti abbiamo utilizzato il simbolo [...].

22. *CL2: [164] e poi lo sai icché c' è ? c' è la &zia [///] c' è che l' è tedesca // [165] ecco // [166] l' è [///] **la** non è italiana // (Corpus Stammerjohann, barb02)

23. *CLE: [...] // [42] sì // \$ [43] perché **la** 'un è roba mia // (Corpus Confronto Diacronico, ipubdl01b)

- Raddoppiamento del soggetto clítico e l'⁹:

24. *MOG [...] // [249] perché / lui / **e l'** è un ragazzo tedesco / no / Giorgio Ottmar // (Corpus Stammerjohann, fami1)

- Raddoppiamento del soggetto clítico e gl'¹⁰:

25. *VIT: [14] sì sì // [15] **e / gl'** è bellino / vai // [16]<guarda> // (Corpus Stammerjohann, vita1)

- Raddoppiamento del soggetto clítico la l'¹¹:

26. *GIO: [32] sta qui / no ? [33] sicchè / **la l'** è più <l' è più / turbata> degli altri / voglio dire // (Corpus Stammerjohann, port)

- Soggetto Æ:

27. *BRA: [...] // [158] lo impiccano / però / siccome Æ è di legno / (Corpus Stammerjohann, pino)

- Soggetto lessicale preverbale:

28. *AND: [864] mah / senti / **il francese** è più facile / eh / dell' inglese // (Corpus Stammerjohann, fran)

29. *GUG: [263] **Vito** / è di' quarantacinque // (Corpus Confronto

9 Non sono stati incontrati esempi di questo tipo di raddoppiamento nel *Corpus di Confronto Diacronico*, per cui citiamo solo l'esempio del *Corpus Stammerjohann*.

10 *Idem*.

11 *Idem*.

Diacronico, fam1)

- Soggetto lessicale postverbale:

30. *AMI: [150] è contento **il gatto** / eh / che [/] che [/] che il lupo sia morto // (Corpus Stammerjohann, vita2)

31. *VAL: [406] certo // [407] 'un è mica sposata / **l' Ornella** // (Corpus Confronto Diacronico, fam1)

- Soggetto pronominale di altro tipo:

32. *FRA: [859] **questa** è una professoressa / sai // (Corpus Stammerjohann, fran)

33. *NIL: [116] **quella** è la procedura che stabilisce il giudice // (Corpus Confronto Diacronico, ifamnm09)

- Pronome tonico soggetto:

34. *MOG: [...] / [237] anche **lei** è a letto // (Corpus Stammerjohann, fami1)

35. *TIZ: [128] magari **lei** è una femmina // (Corpus Confronto Diacronico, ifamd108)

Ci riferiremo, per semplificazione, ai dati normalizzati, più facilmente confrontabili tra loro. Il *Corpus Stammerjohann* presenta un totale di 1.317 occorrenze (ogni 100.000 parole) della forma è quando è un verbo non ausiliare, ossia quando possiede valore copulare, identificativo o predicativo a base locativa, contro le 1.196 occorrenze del *Corpus di Fiorentini*.

Per quanto riguarda l'espressione del soggetto, la prima impressione fornita dai dati è che il calo delle forme clitiche apparentemente non sia avvenuto: il clitico *l'* presenta 276 occorrenze

nel *Corpus Stammerjohann* e 228 nel *Corpus di Fiorentini*. Una diminuzione del 17%, non così forte come ci si sarebbe aspettati. Nel caso di *gl'* si sarebbe avuto persino un aumento, passando da 51 occorrenze nel *Corpus Stammerjohann* a 67 in quello di *Fiorentini*. Per quanto riguarda le forme *la* ed *e* non ci sono stati cambiamenti. Le occorrenze di *la* sono alquanto basse, due per ogni *corpus*. Per quanto riguarda *e*, non si sono riscontrate occorrenze di questa forma neanche nel *Corpus Stammerjohann*, confermando la sua quasi estinzione già nel dopoguerra, come vari studiosi e lo stesso Renzi (1983) già sostenevano.

Sono stati verificati anche i casi di raddoppiamento di soggetto clitico. In questo contesto il calo ci sarebbe stato, ma i numeri sono troppo piccoli: dalle 2 occorrenze per i raddoppiamenti *e l'*, *e gl'* e *la l'* del *Corpus Stammerjohann* a 0 nel *Corpus di Fiorentini*.

L'analisi dei dati si è soffermata sulle altre tipologie di espressione del soggetto. L'ipotesi iniziale è che, oltre al calo dei clitici nel *corpus* più recente, sarebbe stato possibile osservare una generale riduzione dell'espressione del soggetto a favore della forma Æ . Tuttavia, si sarebbe passati dalle 443 occorrenze del *Corpus Stammerjohann* alle 419 del *Corpus di Fiorentini*, con una riduzione di appena il 5%. Ma quello che più ha destato sorpresa è stata la pressoché totale stabilità nell'espressione del soggetto lessicale e pronominale, ed in alcuni casi un suo leggero aumento. Difatti, negli esempi di soggetto lessicale¹² preverbale, il *Corpus Stammerjohann* ha presentato 217 occorrenze e quello dei *Fiorentini* 221. Praticamente non vi sono state alterazioni. I soggetti lessicali postverbali sono passati dalle 137 occorrenze del *Corpus Stammerjohann* alle 160 del *Corpus di Fiorentini*, con un incremento del 17%, ma non sono oggetto di questo lavoro perché la loro funzione è assai diversa e la loro presenza è collegata a ragioni informative che andrebbero sottoposte a un'indagine qualitativa accurata. Invece parrebbe che pronomi dimostrativi abbiano avuto un certo incremento: nel *Corpus Stammerjohann* le occorrenze sono 183 mentre nel *Corpus di Fiorentini* sono 221, un aumento del 21%. Infine, i pronomi tonici sarebbero rimasti praticamente invariati: se ne contano 20 nel *Corpus Stammerjohann* e 21 nel *Corpus di Fiorentini*.

Apparentemente, quindi, il calo dei pronomi clitici soggetto, e in generale dell'espressione del soggetto in fiorentino, non avrebbe risentito dei quarant'anni di pressione dell'italiano standard.

Tuttavia, un'immagine complessiva cambia disaggregando i dati sulla base dei diversi

12 Nel computo dei soggetti lessicali, per praticità, sono state inserite anche le proposizioni soggettive (11 nel *Corpus Stammerjohann* e 4 nel *Corpus di Fiorentini*) e gli indefiniti (3 nel *Corpus Stammerjohann* e 6 nel *Corpus di Fiorentini*).

parametri e confrontandoli in modo più accurato. In primo luogo, osserviamo unicamente i clitici:

Forme	<i>Corpus Stammerjohann</i>		<i>Corpus di Fiorentini</i>	
	Occorrenze assolute	Occorrenze Normalizzate	Occorrenze assolute	Occorrenze Normalizzate
Lessicale preverbale + <i>l'</i>	0	0	12	21
Lessicale preverbale + <i>gl'</i>	1	1	1	1
Lessicale postverbale + <i>l'</i>	1	1	6	10
Lessicale postverbale + <i>gl'</i>	1	1	1	1
Questo + <i>l'</i>	21	24	4	7
Questo (posposto) + <i>l'</i>	1	1	1	1
Quello + <i>l'</i>	15	17	1	1
Questo + <i>gl'</i>	1	1	1	1
Codesto + <i>l'/gl'</i>	1	1	1	1
Quale + <i>l'</i>	0	0	1	1
Chi + <i>l'</i>	0	0	3	5
Lei + <i>l'</i>	3	3	4	7
Lei/lui + <i>gl'</i>	4	5	0	0
Lei/lui (posposto) + <i>l'</i>	1	1	3	5
Lui (posposto) + <i>gl'</i>	1	1	0	0
TOTALE	51	7	39	2

Tabella 6: Forme di soggetto reduplicato incontrate nei due *corpora* a confronto, in termini assoluti e normalizzati.

Senza dilungarci troppo in una spiegazione dettagliata della tabella, vorremmo far notare che i casi di soggetto lessicale, sia preverbale che postverbale, ripresi da clitico soggetto *l'*,

hanno fatto registrare un sorprendente aumento dei casi. Nel *Corpus Stammerjohann* difatti ne incontriamo un unico esempio, mentre nel *Corpus di Fiorentini* ve ne sono ben 31: un incremento apparentemente enorme.

La situazione opposta avviene per i pronomi dimostrativi (*questo* e *quello*) ripresi dal clitico soggetto *l'*: mentre nel *Corpus Stammerjohann* le occorrenze totali sono 41, in quello di *Fiorentini* sono solamente 8.

Per quanto riguarda i pronomi tonici ripresi da clitici, sembrano esserci dei cambiamenti, con un certo aumento nel loro uso nel *Corpus di Fiorentini*, ma i numeri sono troppo bassi per avanzare ipotesi.

Si direbbe invece che, dai dati rilevati, si possa affermare che sembra essere avvenuto, nel tempo, una variazione da parte dei parlanti nell'utilizzo della forma clitica: negli anni Sessanta la preferenza sarebbe stata per la ripresa dei dimostrativi, mentre in tempi recenti la ripresa si verificherebbe con i soggetti lessicali e talora con i tonici.

Diamo un esempio per ognuno dei casi, tratti dai *corpora* consultati:

- Pronome dimostrativo con clitico soggetto *l'*:

36. *CRI: [...] / [77]<**quello l'** è chiaro> // (Corpus Stammerjohann, cust)

- Soggetto lessicale con clitico soggetto *l'*:

37. *LOR: [...] // [39] e poi / sì / **la cravatta l'** è poco personale / pe' un omo // (Corpus Confronto Diacronico, ifamd103)

Vediamo ora cosa succede se prendiamo in considerazione i dati relativi all'età dei parlanti e osserviamo prima cosa accade all'interno di ogni *corpus* e poi mettendo i due *corpora* a confronto, facendo così scoprire alcune cose molto importanti.

Essi seguono i criteri di distribuzione per età della famiglia del C-ORAL-ROM: con la fascia A vengono indicati i parlanti tra i 18 e i 25 anni di età; la fascia B include quelli tra i 26 ed i 39 anni; la fascia C quelli tra 40 e 60 anni; infine la fascia D quelli al di sopra dei 60 anni. Quando non si conosce l'età del parlante è utilizzato il simbolo X. Una particolarità del *Corpus Stammerjohann* è che tre dei quattro parlanti della fascia A, quelli che pronunciano i clitici

soggetto di cui ci occupiamo, sono bambini. L'altro parlante della fascia A non presenta clitici.

Il *corpus Stammerjohann* è composto di 59 parlanti mentre quello di *Fiorentini* ne ha 72.

Lo spoglio minuzioso di tutte le forme di espressione del soggetto per ogni parlante, in cui sono state suddivise le forme lessicali, pre e postverbal, i vari pronomi tonici e dimostrativi, oltre ai clitici soggetto e ai diversi abbinamenti delle reduplicazioni, ha permesso di ottenere un quadro alquanto interessante¹³. Dopo aver terminato il conteggio, sono state escluse le forme di soggetto *Æ*, potendo così estrarre un totale di tutte le forme di soggetto espresso. Quindi sono stati contati i casi in cui è stata usata una delle forme di soggetto clitico pronominale. Successivamente è stata calcolata la percentuale dei soggetti clitici sul totale. I risultati sono presentati nella tabella 7:

	FASCIA A	FASCIA B	FASCIA C	FASCIA D
<i>CORPUS STAMMERJOHANN</i>	38,3%	17%	37,7%	-
<i>CORPUS DI FIORENTINI</i>	11,8%	15,9%	31,6%	58,1%

Tabella 7: totale dei soggetti clitici incontrati nei due *corpora* a confronto, in percentuali sui soggetti espressi, in base all'età dei parlanti.

Nel *Corpus Stammerjohann*, del totale di tutte le forme possibili di espressione del soggetto, quelle con i clitici rappresentano il 38,3% nella fascia A. È importante ricordare che tutte queste forme sono pronunciate da bambini, i quali sono figli di informatori della classe C. Nella fascia B i clitici sono un'alternativa molto meno utilizzata: è incontrata solo nel 17% dei casi. Infine, la fascia C presenta una percentuale simile alla fascia A (37,7%). Mancano i parlanti della classe D. Questi risultati inducono le considerazioni seguenti: si sarebbe verificata una chiara riduzione dei clitici soggetto nella fascia B, corrispondente ai parlanti nati durante la guerra. Se la fascia A non lo conferma sarebbe solo perché si tratta di bambini che riproducono l'uso dei genitori (parlanti della fascia C) perché questo è ancora il loro unico modello. È facile ipotizzare che crescendo siano stati sottoposti a modelli alternati e abbiano seguito o incrementato le tendenze osservabili per la fascia B.

Il *Corpus di Fiorentini* mostra una gradualità nell'utilizzo delle forme clitiche soggetto: la fascia A preferisce altri modi di esprimere il soggetto e si serve dei clitici solo nell'11,8% dei

13 Per ovvi motivi di spazio, non riproduciamo qui le tabelle che si riferiscono ad ogni parlante. Esse saranno disponibili in Ferrari, L. A. (Tese de Doutorado, UFMG, em preparação).

casi. Anche la fascia B è piuttosto restia nell'avvalersi dei clitici come soggetto, utilizzandoli nel 15,9% degli esempi incontrati. Un aumento notevole avviene nella fascia C che se ne avvale nel 31,6% dei casi. La fascia D infine è quella che li adopera maggiormente, nel 58,1% dei casi. È opportuno osservare che la fascia D del C-ORAL-ROM corrisponde grosso modo a quella B del *corpus Stammerjohann*, cioè ai nati durante la guerra.

Sono stati verificati anche altri fattori, come il livello scolastico e la professione dei parlanti, ma nessuno di essi ha mostrato differenze nell'uso delle forme di espressione del soggetto così evidenti come il fattore età.

Se si osservano i due *corpora* separatamente sembra quindi che vi sia stata una costante riduzione dei clitici: nello *Stammerjohann* la fascia B ne riduce fortemente l'uso e la fascia A, data la sua particolare costituzione, non farebbe testo perché i bambini ancora riproducono l'uso dei genitori. Quindi le fasce su cui basarsi sarebbero solo due. A sua volta, nel *Corpus di Fiorentini* ci sarebbe stata una progressiva riduzione dei clitici dalla fascia D alla C e alla B, e ormai in misura molto ridotta tra la fascia B e la A. Insomma, analizzando così i dati sembrerebbe che, dalla generazione corrispondente alla fascia C dello *Stammerjohann* in poi, la perdita dei clitici sarebbe stata costante. La fascia C dello *Stammerjohann* corrisponde all'incirca ai nati fra il 1925 e il 1940. Non abbiamo dati per le generazioni anteriori e quindi, se l'analisi dei dati fosse corretta, dovremmo concludere che perlomeno a partire dai nati intorno al 1940 ci sarebbe stata una progressiva e costante riduzione dei clitici, e che questa riduzione potrebbe essere iniziata anche prima.

Tuttavia, guardando con attenzione, i dati presentano una situazione alquanto strana. Come sarebbe possibile che i parlanti della fascia B dello *Stammerjohann* presentino i clitici solo nel 17% dei casi e i parlanti della fascia D del *Corpus di Fiorentini*, cioè i parlanti della fascia di età che nello *Stammerjohann* appartenevano alla fascia B, li presentino nel 58% dei casi? Strano anche il fatto che i parlanti della fascia C del *Corpus di Fiorentini*, quindi più giovani dei parlanti della fascia B dello *Stammerjohann*, presentino una percentuale di clitici del 31,6%.

Una situazione del genere non può restare senza risposta. Siamo allora andati a verificare in dettaglio la composizione dei due *corpora*, ipotizzando che i tagli operati per mantenere nel *Corpus di Fiorentini* la composizione con parlanti esclusivamente fiorentini e non generalmente toscani come nel *corpus LABLITA* e nel C-ORAL-ROM potesse aver compromesso la comparabilità con lo *Stammerjohann* in altri aspetti. Tuttavia non ci è parso che le diverse componenti dei due *corpora* presentino differenze significative, e certamente non sufficienti

a giustificare la situazione contraddittoria presentata fin qui. Abbiamo allora approfondito l'analisi dei parlanti, per verificare se qualcuno di quelli presenti nel *Corpus di Fiorentini* fosse sovrarappresentato. Ne abbiamo trovati 3, due appartenenti alla fascia C e uno alla D:

- *ELA, della fascia C, è presente in 5 testi, con 21 clitici;
- *LUC, della fascia C, è presente in 1 testo, ma con ben 18 clitici;
- *LUC, della fascia D, è presente in 3 testi, con 19 clitici.

Inoltre, sia ELA che LUC appartengono al limite massimo della fascia C.

Abbiamo allora escluso questi 3 parlanti dal conteggio dei clitici. Ne è risultato un quadro molto diverso e più verosimile. Senza questi 3 parlanti, sia la fascia D che la fascia C del *Corpus di Fiorentini* presentano una percentuale di clitici di appena il 17%, cioè sostanzialmente la stessa della fascia B dello *Stammerjohann*. Se questi dati devono essere considerati affidabili, una conclusione coerente, per quanto ancora provvisoria, sarebbe che tra la fascia C e la fascia B dello *Stammerjohann* ci sarebbe stata una forte riduzione dei clitici (non possiamo sapere se cominciata già prima) e che poi la riduzione sarebbe proseguita con minore evidenza. Ciò indicherebbe che la riduzione sarebbe databile almeno alla generazione nata durante la guerra, se non prima. Sarebbe insomma visibile il salto fra la generazione della fascia D dello *Stammerjohann*, nata intorno al 1925, e la generazione successiva. Nulla possiamo dire della generazione ancora precedente, ma apparentemente il fenomeno di riduzione dei clitici non sarebbe poi proseguito con la stessa intensità.

4. Conclusioni

Chiariamo innanzi tutto che questo studio non è ancora concluso e che in questa sede stiamo presentando dati parziali, resi ancora più parziali dalla necessità di escludere i 3 parlanti sovrarappresentati.

Le ipotesi illustrate devono quindi essere confermate dalla raccolta di dati relativi ad altri verbi. Lo spoglio relativo all'espressione dei soggetti con i verbi *fare*, *andare* e *venire* è già in corso, ma potrebbe essere necessario ricorrere ad altri verbi per ottenere un quadro sufficientemente affidabile.

È possibile anche che con maggiori dati si possano non solo trarre conclusioni a proposito della generale presenza dell'espressione del soggetto e della sua espressione attraverso clitici, ma anche su eventuali cambiamenti nella distribuzione di ciascun clitico.

Sotto il profilo metodologico, questo lavoro mostra tuttavia quanto delicata sia la compilazione di un *corpus* veramente rappresentativo, soprattutto trattandosi di *corpora* di parlato spontaneo, cioè di *corpora* che per la loro stessa natura faticano ad assumere dimensioni che sopportino l'incrocio di troppe variabili.

Riferimenti bibliografici

BERRUTO, G. Dialect/standard convergence, mixing, and models of language contact: the case of Italy. In: AUER, P.; HINSKENS, F.; KERSWILL, P. (a cura di). *Dialect change. Convergence and divergence in European languages*. Cambridge: Cambridge University Press, 2005, pp. 81-97.

_____. Linguistica del contatto e aspetti dell'italianizzazione dei dialetti: appunti di creolistica casalinga. In: *Italica et Romanica*. Festschrift für Max Pfister zum 65. Geburtstag, hrsg. von G. Holtus, J. Kramer & W. Schweickard, Tübingen, Niemeyer, 3 voll., vol. 3°, pp. 13-29, 1997.

BINAZZI, N. 'Italiano' e 'dialetto' a Firenze: 50 giovani giudicano il loro lessico d'uso. In: MULJAČIĆ, Ž. (a cura di). *L'italiano e le sue varietà linguistiche*. Sauerländer: Verlag für deutsch-italienische Studien, 1998, pp. 13-32.

_____. *Le parole dei giovani fiorentini: variazione linguistica e variazione sociale*. Roma: Bulzoni, 1997.

_____. L'appartenenza rivelata. Lessico e tradizioni del discorso nel parlato Fiorentino. In: *Quaderni del Dipartimento di Linguistica dell'Università di Firenze*, Firenze: 17, pp. 137-168, 2007.

BOSTRÖM I. *La morfologia dei pronomi personali: soggetti della terza persona in italiano e in fiorentino*. Contributo allo studio storico dei rapporti fra l'italiano standard e la varietà fiorentina, Volume 5 di Acta Universitatis Stockholmiensis. Stockholm: Almqvist & Wiksell International, 1972.

BRACCO, C.; BRANDI, L.; CORDIN, P. Sulla posizione soggetto in italiano e in alcuni dialetti dell'Italia centro-settentrionale. In: *Sintassi e morfologia della lingua italiana d'uso*. Teorie e applicazioni descrittive. Atti del XVII Convegno Internazionale di Studi, Urbino, 11-13 settembre 1983, SLI, Società di Linguistica Italiana.

BRANDI, L.; CORDIN, P. Two Italian dialects and the null subject parameter. In: JAEGGLI, O.; SAFIR, K. (Eds.). *The Null Subject Parameter*. Dordrecht: Kluwer, 1989, pp. 111-142.

- _____. Dialetti italiani: un confronto sul parametro del soggetto nullo. In: *Rivista Di Grammatica Generativa*, 9781556080876, Vol. 6, pp. 33-87, 1981.
- _____. Subject clitics in two Italian dialects: a perspective on the Null subject Parameter. In: *Journal of Italian Linguistics*, Vol. 8.2, pp. 1-24, 1983.
- CARDINALETTI, A.; MUNARO, N. (a cura di). *Italiano, italiani regionali e dialetti*. Milano: FrancoAngeli, 2009.
- CONTINI, G. Per una interpretazione strutturale della cosiddetta “gorgia” toscana. In: *Boletim de Filologia*, p. 263-281, 1960.
- CRESTI, E. *Corpus di italiano parlato*, 2 voll., vol. I Introduzione; vol. II Campioni. Firenze: Accademia della Crusca, 2000.
- CRESTI, E.; MONEGLIA, M. (a cura di). *C-Oral-Rom. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: John Benjamins, 2005.
- FERRARI, L. A. As formas cê(s) e você(s) na fala espontânea do PB: uma análise baseada em corpora. In: *Domínios de Linguagem*, v.7, n. 2, ISSN 1980-5799, 2013.
- _____. Tese de Doutorado em Estudos Linguísticos, Faculdade de Letras, UFMG, em preparação.
- GALLI DE' PARATESI, N. *Lingua toscana in bocca ambrosiana. Tendenze verso l'italiano standard: un'inchiesta sociolinguistica*. Bologna: il Mulino, 1984.
- GIANNELLI, L. Toscana. In: *Profilo dei Dialetti Italiani*, a cura di Alberto Zamboni, Consiglio Nazionale delle ricerche, Istituto di fonetica e dialettologia. Ospitaletto: Pacini Editore, 2000.
- GIANNELLI, L.; SAVOIA, L. L'indebolimento consonantico in Toscana III. *Rivista italiana di dialettologia* 2/4: 23-48/38-101, 1978 e 1979-1980.
- MARTIN, P. WinPitch Corpus: A text-to-speech analysis and alignment tool. In: CRESTI, E.; MONEGLIA, M. (Org.). *C-ORAL-ROM: integrated reference corpora for spoken Romance languages*. Amsterdam: John Benjamins, 2005, pp. 40-51.
- MONEGLIA, M.; BINAZZI, N.; CELLA, R.; SCARANO, A.; PANUNZI, A.; FABBRI, M. L'incidenza del lessico fiorentino nella lingua d'uso a Firenze. Un confronto tra il corpus Stammerjohann del 1965 e un corpus di parlato contemporaneo. In CRESTI, E. (a c. di) *Prospettive nello studio del lessico Italiano*, Atti del IX Congresso internazionale della Società di Linguistica e Filologia Italiana. Firenze: Firenze University Press, 2008, pp. 99-106.
- PALERMO, M. *L'espressione del pronome personale soggetto nella storia dell'italiano*. Roma: Bulzoni, 1997.
- PANUNZI, A. *La variazione semantica del verbo essere nell'italiano parlato. Uno studio su corpus*.

Firenze: Firenze University Press, 2010.

RASO, T.; MELLO, H. *C-ORAL-BRASIL I: Corpus de referênciã do português brasileiro falado informal*. Belo Horizonte: UFMG, 2012.

RENZI, L. Fiorentino e italiano: storia dei pronomi personali soggetto. In: *Italia linguistica. Idee, storia, strutture*, a cura di F. Albano Leoni et al. Bologna: il Mulino, 1983, p. 223-239.

_____. I pronomi soggetto in due varietà substandard: fiorentino e français avancé. In: POLZIN-HAUMANN, C.; SCHWEICKARD, W. (eds.). *Zeitschrift für romanische Philologie. Volume 108, Issue 1-2 (Jan 1992)*, De Gruyter, 1992.

RENZI, L.; VANELLI, L. I pronomi soggetto in alcune varietà romanze. In: *Scritti linguistici in onore di G.S. Pellegrini*. Pisa: Pacini, 1983.

RICCA, D. *Italianizzazione dei dialetti*. Consultabile al sito: <http://www.treccani.it/enciclopedia/italianizzazione-dei-dialetti/> Accesso il 2 dicembre 2013.

SCARANO, A. (a cura di). *Corpus di confronto diacronico LABLITA*, 2005. <http://lablita.dit.unifi.it/corpora/descriptions/corpusconfronto/> Accesso il 2 novembre 2012.

SCARANO, A.; MONEGLIA, M. Il Corpus Stammerjohann. Il primo corpus di italiano parlato, in rete nella base dati di LABLITA. In PETTORINO, M. (a cura di). *Atti del Congresso Internazionale "La comunicazione Parlata"*. Napoli, 23-25 Febbraio 2006.

SCARANO, A.; SIGNORINI, S. Corpus linguistics and diachronic variability. A study on Italian spoken language corpora from the 1960s until nowadays. In: KABATEK, J.; PUSCH, C. D.; RAIBLE, W. (eds). *Korpora und diachrone Sprachwissenschaft. Romance Corpus Linguistics II: Corpora and Diachronic Linguistics*. Tübingen: Narr, p. 191-202, 2005.

SCHIEL, F. et al. *The validation of speech corpora*. 2004. <http://www.phonetik.uni-muenchen.de/forschung/BITS/TP2/Cookbook/> Accesso: 20 dicembre 2012.

SIGNORINI, S.; TUCCI, I. Il restauro e l'archiviazione elettronica del primo corpus di italiano parlato: il corpus Stammerjohann. In: DE DOMINICIS, A.; MORI, L.; STEFANI, M. (a cura di). *Costituzione, gestione e restauro di corpora vocali*, Atti delle XIV Giornate del GFS, Collana degli atti dell'associazione italiana di acustica. Viterbo, 4-6 dicembre 2003. Roma: Esagrafica, 2003, pp. 119-126.

SOBRERO, A. Italianization of the dialects. In: MAIDEN M.; PARRY, M. (eds.). *The dialects of Italy*. London; New York: Routledge, pp. 412-418, 1997.

STAMMERJOHANN, H. Strukturen der Rede. Beobachtungen an der Umgangssprache von Florenz. In: *Studi di Filologia Italiana*, XXVIII, pp. 295-397, 1970.