

Un'analisi diacronica *corpus-based* dei soggetti clitici di terza persona singolare in fiorentino

Lúcia de Almeida Ferrari, Tommaso Raso
Universidade Federal de Minas Gerais

The article focuses on the supposed loss of clitic subjects in Florentine vernacular, through a corpus-based study. Previous works (Galli De' Paratesi 1984; Sobrero 1997; Berruto 1997, 2005; Binazzi 1997, 1998, 2007; Cardinaletti e Munaro 2009) claim that pressure of standard Italian, in post-war years, contributed to the loss of many of the characteristics of vernacular Florentine. We present a corpus-based analysis through the re-masterized publication of *Corpus Stammerjohann* (1970) and the *Corpus per il Confronto Diacronico* (Scarano 2005), both of them spontaneous spoken corpora recorded in Florence. The time span between the two corpora is approximately forty years. The investigation analyzes mainly the expression of third person singular clitic subject *gl', l', la, e* in the indicative present of *essere* (to be), *andare* (to go), *venire* (to come) and *fare* (to do). Results obtained in this first stage of the research indicate that there was a loss of clitic subjects along the time; it additionally suggests that the beginning of that loss must be dated back. Further research should be pursued in the future taking into account a larger amount of data to better understand this phenomenon.

Keywords: clitics, florentine, spoken corpora, diachrony

1. Introduzione

In questo contributo verranno presentati i primi risultati di un'analisi diacronica *corpus-based* dei pronomi clitici soggetto di terza persona singolare *gl', l', la, e* nel vernacolo fiorentino. Obiettivo del lavoro è verificare la questione dell'indebolimento dei clitici soggetto in fiorentino dovuta, secondo buona parte degli studiosi, alla pressione dell'italiano standard, specialmente a partire dalla fine degli anni Sessanta, con la scolarizzazione di massa e la diffusione dei mass-media, in particolare della televisione (Galli De' Paratesi 1984; Sobrero 1997; Berruto 1997, 2005; Binazzi 1997, 1998, 2007; Cardinaletti e Munaro 2009).

La novità dello studio risiede nel fatto che, per la prima volta, ci si è avvalsi di dati di parlato spontaneo in diacronia per verificare il calo dei pronomi clitici soggetto in fiorentino e osservare eventuali riconfigurazioni dell'espressione del soggetto. I corpora utilizzati sono il *Corpus Stammerjohann* (Stammerjohann 1970; Signorini & Tucci 2005; Scarano & Moneglia 2006) e il *Corpus per il confronto diacronico LABLITA* (Scarano 2005). La possibilità di avere a disposizione corpora che forniscono dati diacronici comparabili è unica, sfruttata sinora specialmente per studi di tipo lessicale (Moneglia et al. 2008). Si ritiene dunque che questo lavoro offra nuovi dati per la comprensione della riduzione dei clitici soggetto nel vernacolo fiorentino.

L'articolo è suddiviso come segue: in 2 saranno esposte le particolarità morfosintattiche dei pronomi soggetto del vernacolo fiorentino rispetto all'italiano. In 3 verranno presentati i corpora consultati: il *Corpus Stammerjohann* e il *Corpus per il Confronto Diacronico*. In 4 si esporrà la metodologia utilizzata; 5 è dedicato all'analisi dei dati. Infine, la conclusione presenta una sintesi degli argomenti trattati e una riflessione sugli sviluppi futuri possibili a partire dalla ricerca qui presentata.

2. Il vernacolo fiorentino e i soggetti clitici

L'italiano standard è una lingua a espressione libera del soggetto: essa è facoltativa e motivata pragmaticamente (si vedano a questo proposito gli studi di Rizzi 1982, 1986a, 1986b; Kayne 1989, 1991; Palermo 1997; Renzi 1983). D'altra parte i dialetti del nord Italia ed il vernacolo fiorentino, ma non gli altri dialetti toscani, secondo diversi autori, esprimono obbligatoriamente il soggetto¹ (Brandi & Cordin 1981, 1983, 1989; Bracco, Brandi & Cordin 1985; Renzi & Vanelli 1983; oltre al volume I di Manzini & Savoia 2005). Nel fiorentino il soggetto può essere espresso:

- da una forma lessicale o da un pronome tonico;
- da un pronome clitico;

¹ La questione dei soggetti clitici è oggetto di studio in varie altre lingue oltre ai dialetti italiani. Il portoghese brasiliano ad esempio evidenzia in maniera sempre più forte l'uso della forma pronominale di seconda persona singolare contratta *cê* a discapito di quella standard *você*, come mostrano sia gli studi di matrice generativista (Vital 1996; Ramos 1997; Vital & Ramos 2006) sia studi quantitativi e qualitativi su corpus di parlato spontaneo (Ferrari 2015).

- da una forma lessicale o da un pronome tonico reduplicati dal clitico²;
- da un clitico reduplicato da un altro clitico³.

Sono presentati di seguito alcuni esempi fittizi che illustrano la situazione: (1) e (2) mostrano rispettivamente il soggetto lessicale e quello pronominale tonico. In (3) si ha il soggetto clitico fiorentino di terza persona singolare femminile. Gli esempi (4) e (5) rappresentano rispettivamente la reduplicazione del soggetto lessicale e del soggetto pronominale tonico con il clitico femminile *la*. In (6) è rappresentata la reduplicazione del clitico con i pronomi *e* e *la*, entrambi clitici, in funzione di soggetto (Renzi 1983).

- (1) La Maria va a casa
- (2) Lei va a casa
- (3) La va a casa
- (4) La Maria la va a casa
- (5) Lei la va a casa
- (6) E la va a casa

L'espressione pronominale del fiorentino presenta una doppia serie: una tonica e una clitica, distinte dal punto di vista segmentale.

² Sottolineiamo che, nell'ambito della teoria relativa ai parametri della Grammatica Universale, in Brandi & Cordin (1981, 1983, 1989) e in particolare in Bracco, Brandi & Cordin (1985: 195 e segg.), si sostiene che le differenze di comportamento dei dialetti friulano, torinese e fiorentino rispetto all'italiano siano dovute a fatti di ordine superficiale e non profondo. Ciò porterebbe questi dialetti ad essere considerati lingue a soggetto nullo, come l'italiano. In fiorentino, nelle frasi a soggetto \emptyset il clitico assorbirebbe il ruolo tematico del soggetto preverbale (sia esso agente, sperimentatore, ecc.), fungendo quindi da vero e proprio soggetto. Nel caso di reduplicazioni del soggetto, lessicale o pronominale, con il clitico, secondo il principio del Criterio Tematico (Bracco, Brandi & Cordin 1985: 195), che stabilisce che ogni SN referenziale deve possedere uno e un solo ruolo tematico, sarebbe l'elemento lessicale a veicolare il ruolo tematico espresso dal soggetto preverbale, per cui il clitico fungerebbe semplicemente da morfema di flessione verbale.

³ Nel caso del clitico reduplicato dal clitico, Bracco, Brandi & Cordin (1985) non esplicitano l'interpretazione, ma si potrebbe considerare che il primo clitico assorbirebbe il ruolo tematico del soggetto preverbale, fungendo da soggetto, ed il secondo sarebbe una semplice marca di flessione verbale.

Si osservi la tabella 1⁴.

Tabella 1. Doppia serie pronominale soggetto nel vernacolo fiorentino

Pronome tonico soggetto	Pronome atono soggetto
Io	<i>e</i> ⁵ (può introdurre tutte le persone a eccezione della terza singolare)
Te	<i>tu</i>
lui (maschile), lei (femminile)	<i>gl'</i> (maschile prima di vocale), <i>e</i> (maschile prima di consonante), <i>l'</i> (maschile prima di vocale) <i>la</i> (femminile), <i>l'</i> (femminile prima di vocale)
Noi	<i>e</i> quando accompagna un altro clitico
Voi	<i>vu</i>
Loro	<i>gl'</i> (maschile prima di vocale), <i>e</i> (maschile prima di consonante) <i>le</i> (femminile) <i>l'</i> (femminile prima di vocale)

Secondo Giannelli (2000) i pronomi atoni possono accompagnare i tonici ed essere, a loro volta, accompagnati da *e*, che darebbe una “sfumatura enfatica” all’enunciato⁶.

Nella bibliografia consultata non sono state riscontrate affermazioni sulla possibilità di uso di *l'* come maschile singolare prima di vocale ma lo stesso Renzi (1983: 224) lo cita tra i suoi esempi⁷. D'altronde l'analisi dei corpora ne ha riscontrato varie occorrenze come nell'esempio sotto:

- (7) // [38] questo qui *l'* è il piolo / quello a Santo Spirito // (Corpus Stammerjohann, arti)

Gli studi di Renzi (1983, 1992) e di Giannelli (2000) si occupano direttamente della questione del fiorentino in prospettiva sincronica e diacronica. Secondo Renzi (1983), in fiorentino il soggetto pronominale sarebbe obbligatorio: potrebbe essere espresso da un clitico che occupa una posizione fissa dinanzi al

⁴ Parzialmente tratta da Giannelli (2000: 38-39).

⁵ 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: 226).

⁶ Come in: *lui e l'è ito*, con significato di *proprio lui è andato*.

⁷ Tale possibilità ci è stata confermata anche da Binazzi in una comunicazione personale (2014).

verbo, o da uno o più clitici che a loro volta precedono il verbo. La formula è quella che segue (Renzi 1983: 224):

(8) (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 esempi fittizi alla terza persona singolare, che è quella che verrà analizzata in questa ricerca:

(9) *Lui l'*arriva.

(10) *I' bambino l'*arriva⁸.

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

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

Nell'esempio (9) il pronome soggetto *lui* viene reduplicato dal clitico soggetto *l'* (che può essere usato solo davanti a vocale), mentre nell'esempio (10) è il SN *i' bambino* ad essere reduplicato dal clitico *l'*. Gli esempi (11) e (12) mostrano la stessa situazione nel caso del femminile, reduplicato dal clitico *la*.

Renzi (1983) sostiene che questo schema include anche la prima persona plurale (Renzi 1983: 224):

(13) (noi) *si* canta.

Il pronome tra parentesi è facoltativo e il verbo viene coniugato alla terza persona singolare. In questo caso *si* sarebbe il clitico soggetto del tonico noi.

Vi sono tuttavia contesti nei quali non avviene la reduplicazione del clitico; in essi il fiorentino presenta lo stesso comportamento dell'italiano standard. Ciò è riscontrabile nei seguenti casi:

- a. alla terza persona maschile, anche nel caso di soggetto lessicale, quando il verbo seguente inizia con consonante:

⁸ Gli esempi (9) e (10) sono estratti da Renzi (1983: 224) che cita anche l'esempio seguente, "(lui, i bambino) l'ha cantato", come casi di soggetto nominale facoltativo e clitico obbligatorio di terza persona maschile singolare dinanzi a vocale. Si noti innanzitutto che Renzi utilizza qui il clitico *l'*, ed in secondo luogo che l'interpretazione della frase è di soggetto reduplicato da clitico e non, come potrebbe sembrare a prima vista, di soggetto lessicale e clitico oggetto.

(14) (lui , i' bambino) canta

- b. alla terza persona singolare, quando il soggetto è posposto e il verbo inizia con consonante:

(15) viene le bambine⁹

Sempre secondo Renzi (1983: 225), nel caso della negazione il pronome clitico viene immediatamente preceduto da essa. Siccome prende un posto fisso, la negazione stessa diventerebbe un clitico.

(16) *Lui 'un gl'è arrivato.*

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

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

dove l'eventuale soggetto pronominale *lei* viene reduplicato dal clitico *la*, questa volta seguito dalla negazione *'un*.

È bene chiarire la questione del clitico *e*. Gli studiosi sono concordi nell'affermare che il suo uso è praticamente scomparso da tempo nel parlato. Analizzandone la forma sintattica, Renzi (1983: 226) 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 *e* sia presente, non compare il pronome tonico, mentre può esserci la reduplicazione clitica. Di seguito vengono esemplificati i suoi usi (la parentesi negli esempi indica facoltatività):

(18) (*e*) parlo / (*e*) tu parli / (*e*) la parla / (*e*) parliamo / (*e*) vu parlate / (*e*) parlano

(19) (*e*) fa caldo / (*e*) bisogna fare presto.

⁹ Nel caso il verbo inizi per vocale, il clitico viene anteposto ad esso: *l'è venuto le bambine*.

- (20) (e) piove / (e) tuonò.
 (21) (e) la mi fa male la testa.
 (22) (e) *gl'*è andato a casa.
 (23) (e) *la* 'un capisce niente.

L'esempio (18) 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, viene utilizzato facoltativamente il clitico *e*; la forma *vu* è quella clitica per la seconda persona plurale; infine, la terza plurale ha come forma clitica facoltativa, davanti a consonante, *e*. Gli esempi (19) e (20), rispettivamente, mostrano *e* come soggetto facoltativo di verbi impersonali e meteorologici. Nell'esempio (21) si ha la costruzione con il soggetto posposto: il clitico *e* è facoltativo, viene seguito dal clitico *la*, reduplicazione del soggetto posposto SN *la testa*. Gli esempi (22) e (23) mostrano la reduplicazione del clitico al maschile (22), con le forme *e* e *gl'*, entrambi clitici soggetto, e al femminile (23) con le forme *e* e *la*, clitici soggetto, seguite dal clitico di negazione 'un.

Per concludere, il paradigma dei clitici soggetto di terza persona nel vernacolo fiorentino è riportato nella tabella 2.

Tabella 2. Forme clitiche di terza persona singolare nel vernacolo fiorentino

maschile	femminile
<i>gl'/l'</i> (dinanzi a vocale)	<i>la</i> (dinanzi a consonante)
<i>e</i> (dinanzi a consonante)	<i>l'</i> (dinanzi a vocale)

3. Metodologia

Come è stato detto, gli studi effettuati sinora sostengono che la forte pressione dell'italiano standard, a partire dagli anni del secondo dopoguerra e ancor più dagli anni Sessanta-Settanta, abbia contribuito in modo sostanziale alla perdita, da parte del vernacolo fiorentino, di varie delle sue caratteristiche sintattiche, morfologiche e lessicali.

Non siamo tuttavia a conoscenza di lavori che verifichino tali affermazioni a partire da dati di parlato consultabili, anche perché ciò renderebbe necessario l'uso di corpora di parlato di tipo diacronico, materiale assai raro.

Questa opportunità, probabilmente unica, è resa possibile grazie ai lavori svolti dal laboratorio LABLITA del Dipartimento di Italianistica dell'Università degli Studi di Firenze, diretto da Emanuela Cresti e da Massimo Moneglia. 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 2005). 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 (Binazzi 1997).

L'obiettivo di questo studio è verificare cosa accade all'espressione del soggetto ponendo i due corpora a confronto.

Come è stato specificato, la ricerca si è concentrata sui pronomi soggetto clitici di terza persona singolare: *gl'*, *la*, *l'*, *e*. È stata valutata, prima in maniera sincronica e poi diacronica, la vitalità di dette forme e dei vari modi di espressione del soggetto. In particolare si è voluto verificare:

- a. la presenza di soggetto \emptyset o di soggetto espresso¹⁰;
- b. la presenza o assenza del clitico *e*, ed in quali casi esso compare;
- c. la presenza delle forme *l'* e *gl'*, per il maschile;
- d. la presenza delle forme *l'* e *la*, per il femminile;
- e. la presenza della reduplicazione clitica;
- f. la presenza di forme di reduplicazione clitica del soggetto;
- g. l'eventuale concomitanza di forme diverse di espressione del soggetto.

Verranno di seguito descritti i corpora sui quali si è concentrato il lavoro.

¹⁰ Dato che entrambi i corpora presentano una segmentazione prosodica in enunciati e unità tonali, in modo da poter essere studiati secondo la Language into Act Theory (L-AcT) (per i dettagli della teoria rimandiamo a Cresti 2000 e a Moneglia & Raso 2014), dopo aver controllato ed eventualmente rivisto la segmentazione, abbiamo individuato i casi di *Topic*, in modo da non inserirli fra quelli di soggetto. Cresti (2010) ha valutato l'importanza dei clitici e delle loro relazioni anaforiche nell'articolazione dell'informazione, mostrando come le relazioni che occorrono all'interno di un'unica unità informativa tra una testa e un clitico linearizzato sono osservabili solamente in tre unità informative di tipo testuale: *Comment*, *Topic* e *Parentetico*.

3.1 I corpora utilizzati

Il *Corpus Stammerjohann* è considerato il primo esempio di corpus di parlato spontaneo in lingua italiana.

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 ore¹¹. La finalità di tale lavoro 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 avvalse di attrezzature assai sofisticate per quegli anni¹², il che ha garantito fino alla sua rimasterizzazione nel 2001 la preservazione di gran parte del materiale raccolto. Circa la metà delle quaranta ore di audio è composto da registrazioni d'ambiente: il registratore venne collocato in un punto nascosto della bottega di un barbiere in via Faenza, mentre i clienti entravano, si fermavano e uscivano. È stato così possibile ottenere registrazioni di situazioni comunicative differenziate e su diversi temi. L'altra metà delle registrazioni è composta da lunghe interazioni, perlopiù dialogiche o monologiche, senza cambiamenti di situazione o 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 familiare¹³.

I preziosi documenti che Stammerjohann ha lasciato¹⁴ 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

¹¹ Per la precisione 41 ore, 15 minuti e 34 secondi.

¹² Fu utilizzato un registratore Rk 34 Philips a bobine a quattro piste e velocità di 4,75 cm al secondo.

¹³ La descrizione dettagliata di testi, partecipanti e durata delle registrazioni può essere incontrata in Scarano & Signorini (2005), responsabili per l'attualizzazione del corpus.

¹⁴ L'autore ha lasciato un'unica pubblicazione sul corpus, Stammerjohann (1970), presentata negli *Studi di Filologia Italiana*, in cui compaiono alcune trascrizioni delle registrazioni.

testi e classificarli secondo le stesse caratteristiche sociolinguistiche e di architettura adottate nel corpus C-ORAL-ROM¹⁵ (Cresti & Moneglia 2005).

Scartando gli audio danneggiati e seguendo i criteri di corretto bilanciamento e rappresentatività diafasica, è stato successivamente 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 regolata¹⁶, 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 adottati nel corpus C-ORAL-ROM e allineati al suono tramite il software *WinPitch* (Martin 2005). È possibile così accedere 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. Nella tabella 3 inseriamo i dettagli del corpus per numero di testi, parole e tipologie di interazione.

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

contesto	tipo	interazione	numero di testi
Familiare 22.517 parole	Non-regolato 22.517 parole	dialogo: 9.310 parole	2
		conversazione: 13.207 parole	4
Privato 36.817 parole	Non-regolato 25.606 parole	conversazione: 25.606 parole	8
		Regolato 11.211 parole	1
	Regolato 11.211 parole	monologo: 3.023 parole	1
		dialogo: 8.188 parole	4
Pubblico 23.196 parole	Non-regolato 16.278 parole	conversazione: 16.278 parole	4
		Regolato 6.918 parole	3
	Regolato 6.918 parole	monologo: 1.561 parole dialogo: 5.357 parole	4
Telefonico 3.722 parole		3.722 parole	2
Totale		86.252 parole	32

Per rendere utilizzabile in modo proficuo ed in prospettiva diacronica il Corpus Stammerjohann, è stato compilato nel 2005, sempre presso LABLITA, il *Corpus*

¹⁵ I dettagli dei criteri e procedimenti adottati si trovano in Scarano & Signorini (2005) e Signorini & Tucci (2004).

¹⁶ L'interazione è regolata quando vi sono norme sociali che stabiliscono lo scambio di turno tra i parlanti, come quelle che avvengono in situazioni di lavoro o di una narrazione guidata; non regolata è quando non è possibile prevedere la risposta dell'interlocutore e il cambio di turno avviene in maniera non pianificata. In questo corpus, le interazioni regolate corrispondono alla situazione dell'intervista, mentre quelle non regolate ad interazioni libere.

per il *Confronto Diacronico* (Scarano 2005; Moneglia *et al.* 2008; Scarano & Signorini 2005; Scarano & Moneglia 2008). Per la compilazione, sono stati selezionati testi provenienti da C-ORAL-ROM italiano e dai corpora LABLITA, in modo da replicare la struttura, il numero di parole, le varie suddivisioni interne e la variazione diafasica del *Corpus Stammerjohann*. I ricercatori del LABLITA hanno quindi dovuto operare una cernita fra i testi degli altri corpora per garantire la comparabilità diafasica, diastratica e di contenuto con il *Corpus Stammerjohann*.

Il nuovo corpus è riuscito a mantenere la struttura originale ed approssimativamente anche il numero di parole, 85.908. La variante principale è il totale di testi, che ammonta a 55 (Scarano & Signorini 2005; Signorini & Tucci 2005).

Per la ricerca che viene qui presentata abbiamo utilizzato tutti i testi del *Corpus Stammerjohann*. Tuttavia, dato che il *Corpus per il Confronto Diacronico* è composto da contributi di parlanti in genere toscani ma non tutti fiorentini, e poiché questo lavoro è interessato a un fenomeno che ha declinazioni diverse all'interno della Toscana, è stato necessario operare un taglio nel corpus, eliminando i parlanti non fiorentini. In questo modo sono stati estratti 40 dei 55 testi del corpus, per un totale di 58.299 parole¹⁷: il corpus risultante è stato chiamato *Corpus di Fiorentini*. La tabella 4 riporta i dati del *Corpus di Fiorentini*.

Tabella 4. Struttura del *Corpus di Fiorentini*, con numero di parole e tipologie dei testi e delle interazioni

contesto	tipo	interazione	numero di testi
Familiare 21.232 parole	Non-regolato	dialogo: 9.415 parole	6
	21.232 parole	conversazione: 11.817 parole	6
Privato 24.862 parole	Non-regolato	conversazione: 16.641 parole	12
	Regolato	monologo: 1.958 parole	2
		8.221 parole	dialogo: 6.263 parole
Pubblico 23.196 parole	Non-regolato	conversazione: 7.308 parole	4
	Regolato	dialogo: 1.554 parole	1
Telefonico 3.343 parole		3.343 parole	5
Totale		58.299 parole	40

In termini di numero di parole il *Corpus di Fiorentini* rappresenta circa il 68% del totale del *Corpus per il Confronto Diacronico*. Dal punto di vista del bilan-

¹⁷ Per i dettagli di tale scelta e le specificazioni dei testi utilizzati si veda Ferrari (2015).

ciamento interno, nel contesto privato sono stati mantenuti valori prossimi all'originale: nel *Corpus per il Confronto Diacronico* il contesto privato rappresenta il 42,55% del numero totale di parole, mentre nel *Corpus di Fiorentini* è il 42,65%. Nel contesto privato quindi la proporzione è prossima al corpus originale: le interazioni non regolate rappresentano il 30,13% del totale di parole del *Corpus per il Confronto Diacronico* e il 28,54% del *Corpus di Fiorentini*; quelle regolate il 12,42% nel primo caso e il 14,10% nel secondo.

Le maggiori differenze riguardano i contesti familiari e pubblici, come è possibile visualizzare nella tabella 5.

Tabella 5. Confronto, in termini percentuali, tra il *Corpus per il Confronto Diacronico* e il *Corpus di Fiorentini*, per ogni campo della struttura.

contesto	tipo	interazione
Familiare Corpus per il Confronto Diacronico: 26,57 Corpus di Fiorentini: 36,42	Non-regolato Corpus per il Confronto Diacronico: 26,57	Dialogo Corpus per il Confronto Diacronico: 10,98 Corpus di Fiorentini: 16,15
	Corpus di Fiorentini: 36,42	Conversazione Corpus per il Confronto Diacronico: 15,59 Corpus di Fiorentini: 20,27
Privato Corpus per il Confronto Diacronico: 42,55 Corpus di Fiorentini: 42,65	Non-regolato Corpus per il Confronto Diacronico: 30,13	Conversazione Corpus per il Confronto Diacronico: 30,13 Corpus di Fiorentini: 28,54
	Corpus di Fiorentini: 28,54	Monologo Corpus per il Confronto Diacronico: 4,05 Corpus di Fiorentini: 3,36
	Regolato Corpus per il Confronto Diacronico: 12,42	Dialogo Corpus per il Confronto Diacronico: 8,37 Corpus di Fiorentini: 10,74
	Corpus di Fiorentini: 14,10	
Pubblico Corpus per il Confronto Diacronico: 26,97 Corpus di Fiorentini: 15,20	Non-regolato Corpus per il Confronto Diacronico: 18,77	Conversazione Corpus per il Confronto Diacronico: 18,77 Corpus di Fiorentini: 12,54
	Corpus di Fiorentini: 12,54	Monologo Corpus per il Confronto Diacronico: 1,42 Corpus di Fiorentini: 0
	Regolato Corpus per il Confronto Diacronico: 8,2	Dialogo Corpus per il Confronto Diacronico: 6,78 Corpus di Fiorentini: 2,67
	Corpus di Fiorentini: 2,67	
Telefonico Corpus per il Confronto Diacronico: 3,9 Corpus di Fiorentini: 5,73		Corpus per il Confronto Diacronico: 3,9 Corpus di Fiorentini: 5,73
Totale		100

La tabella mostra che il contesto familiare rappresenta il 26,57% del totale di parole nel *Corpus per il Confronto Diacronico*, mentre nel *Corpus di Fiorentini* esso è responsabile per il 36,42% dell'insieme. Il contesto pubblico rappresenta il 26,97% del totale di parole nel *Corpus per il Confronto Diacronico*, mentre nel *Corpus di Fiorentini* solo il 15,2%. Infine, nel caso delle interazioni telefoniche, il numero di parole è identico, non essendo stati fatti tagli di nessun tipo. Tuttavia, nel *Corpus per il Confronto Diacronico*, questa sezione rappresenta il 3,9% del totale di parole, mentre nel *Corpus di Fiorentini* il 5,73%.

Dati i tagli subiti, il bilanciamento originale è stato sicuramente in parte compromesso ma, nell'elaborazione delle conclusioni sulle ricerche svolte, questo fattore è stato tenuto in considerazione e i suoi effetti verranno commentati.

Concludendo la presentazione del materiale analizzato, la tabella 6 presenta i dati essenziali sui quali abbiamo lavorato.

Tabella 6. Dati sui corpora utilizzati nella ricerca

	Corpus Stammerjohann	Corpus di Fiorentini
n. parole	86.252	58.299
n. testi	32	40

3.2 La scelta dei verbi

Per individuare le forme clitiche oggetto di studio all'interno dei corpora è necessario verificare i verbi alla terza persona singolare, dato che il soggetto può essere espresso in vari modi. Sono stati scelti i verbi di maggiore frequenza, cioè *essere, fare, andare e venire*, alla terza persona singolare del presente indicativo.

La selezione dei verbi sopra citati ha seguito alcuni criteri di base: sono stati innanzitutto individuati i lemmi verbali ad alta frequenza. In seguito si è deciso di scartare quelli la cui alta frequenza era dovuta soprattutto a specifici usi pragmatici in espressioni fisse, come “vedi, dice”.

Il primo verbo selezionato è stato *essere*, il secondo lemma a maggior frequenza riscontrato all'interno del C-ORAL-ROM italiano¹⁸, equivalente al 4,6% del totale di *tokens* del corpus: *essere* sarebbe il verbo a più alta frequenza dell'italiano. Oltre alla sua importanza dal punto di vista quantitativo, dallo studio di Panunzi (2010) sono emerse importanti considerazioni che hanno fatto ricadere la 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

¹⁸ Uno studio *corpus-driven* di Panunzi (2010) sulla sua variazione semantica ne ha individuate circa 14.000 occorrenze.

vari e qualitativamente diversi: i principali sono quello copulare, quello identificativo e quello predicativo a base locativa¹⁹. Per le finalità di questo studio sono stati esclusi, per il momento, sia *esserci* sia *essere* con funzione di ausiliare.

La scelta delle forme verbali *fa*, *va* e *viene* è dovuta al fatto che limitarsi alla forma *è* avrebbe compromesso i risultati relativi al clitico *e*, vista l'impossibilità della sua cooccorrenza davanti a vocale. Sono state quindi scelte forme inizianti per consonante, mantenendo il criterio di base dell'alta frequenza ed evitando le espressioni fisse.

4. Analisi dei dati

Prima di presentare e analizzare i dati relativi alle forme di espressione del soggetto, è opportuno esplicitare che dai dati sono stati ritirate tutte le forme che devono essere considerate Topic e non forme di soggetto. I corpora sono già segmentati in enunciati e unità tonali, cosa che facilita l'identificazione dei Topic. In ogni caso, abbiamo ascoltato tutti gli enunciati interessati e in qualche caso abbiamo corretto la segmentazione. Ci basiamo sulla concezione di Topic della Language into Act Theory (Cresti 2000; Moneglia & Raso 2014)²⁰.

La tabella 7 illustra, relativamente alla forma *è*, tutti i casi in cui compaiono Topic e le possibili combinazioni con l'espressione del soggetto. Le percentuali si riferiscono a tutte le posizioni preverbalì. La tabella 8 presenta invece i Topic davanti alle forme di *fa*, *va* e *viene*. Nella tabella 7 si può osservare che il Topic è presente nell'11,7% delle forme verbali nel *Corpus Stammerjohan* e nel 12,4% delle forme verbali del *Corpus di Fiorentini*, cioè sostanzialmente con la stessa frequenza. Inoltre, la disaggregazione per fasce d'età non presenta alcuna tendenza nell'evoluzione d'uso. Vedremo che la situazione sarà diversa per quanto riguarda l'espressione del soggetto. Nella tabella 8 sono presentati i numeri relativi alle occorrenze di Topic per gli altri verbi. Trattandosi di numeri molto piccoli non presentiamo i dati disaggregati.

Presentiamo ora i risultati delle analisi delle forme verbali *è*, *fa*, *va* e *viene* nel *Corpus Stammerjohann* e nel *Corpus di Fiorentini*, con la classificazione delle varie forme di espressione del soggetto.

¹⁹Per un approfondimento su questi usi si rimanda a Panunzi (2010). Esemplicando brevemente, essere con valore copulare compare in frasi del tipo “Questa è una bella giornata”; con valore identificativo in “Lui è l'avvocato”; e con valore predicativo a base locativa in “Qui è casa mia”.

²⁰Per studi specifici sul Topic in italiano, si vedano Firenzuoli & Signorini (2003) e Signorini (2003, 2005).

Tabella 7. Topic davanti a *è*, e cooccorrenza con le diverse forme di espressione del soggetto, divise per fasce d'età. Numeri assoluti e percentuali.

	Corpus Stammerjohann			Corpus di Fiorentini			
	A	B	C	A	B	C	D
Topic lessicale + sogg. \emptyset	6 (7,3%)	12 (12,2%)	64 (8,6%)	3 (2,9%)	25 (12,7%)	12 (3,9%)	5 (10%)
Topic lessicale + sogg. clit.	1 (1,2%)	1 (1%)	11 (1,5%)	0 (0%)	2 (1%)	11 (3,5%)	3 (6%)
Topic pronomi per- sonale + sogg. \emptyset	0 (0%)	2 (2%)	1 (0,1%)	0 (0%)	2 (1%)	0 (0%)	1 (2%)
Topic pronomi per- sonale + sogg. clit.	0 (0%)	0 (0%)	1 (0,1%)	0 (0%)	1 (0,5%)	1 (0,3%)	0 (0%)
Topic pronomi per- sonale + sogg. radd. clit.	0 (0%)	0 (0%)	1 (0,1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Topic pronomi non personale + sogg. \emptyset	0 (0%)	1 (1%)	8 (1,1%)	8 (7,8%)	4 (2%)	5 (6%)	0 (0%)
Topic pronomi non personale + sogg. clit.	1 (1,2%)	1 (1%)	3 (0,4%)	0 (0%)	1 (0,5%)	1 (0%)	0 (0%)
TOTALE	114 (11,7%)			85 (12,4%)			

Tabella 8. Topic in valori assoluti nelle posizioni preverbalì con i verbi fare, andare e venire.

Tipi di topic/verbo	Corpus Stammerjohann				Corpus di Fiorentini			
	fa	va	viene	totale	fa	va	viene	totale
lessicale	1	7	2	10	4	0	0	4
pronomi dimostrativo	0	0	0	0	0	0	0	0
pronomi personale tonico	0	3	0	3	0	0	0	0

4.1 Verbo *essere*

Lo spoglio dei dati nei due corpora è stato effettuato utilizzando il software *AntConc*. Il programma rileva la forma ricercata insieme al contesto. Nel *Corpus Stammerjohann* sono state riscontrate ed analizzate 1.977 occorrenze della forma *è* di cui 1.126 sono risultate appartenere a *essere* come verbo non ausiliare. Nel *Corpus di Fiorentini* il totale di occorrenze della forma *è* è di 1.347 occorrenze, di cui 697 appartenenti a *essere* come verbo non ausiliare. La ricerca ha con-

siderato gli enunciati interrotti che non erano passibili di analisi, escludendoli dal computo totale.

Nella tabella 9 riportiamo le occorrenze analizzate: 970 per il *Corpus Stammerjohann* e 686 per il *Corpus di Fiorentini*. Dato che il numero totale di parole dei due corpora è diverso, per poterli confrontare, i dati sono stati normalizzati per 100.000; nella colonna di destra il lettore incontrerà i valori ogni 100.000 parole.

Tabella 9. Forme soggetto incontrate nei due corpora, in termini assoluti e normalizzati.

forme	Corpus Stammerjohann		Corpus di Fiorentini	
	occorrenze assolute	occorrenze normalizzate	occorrenze assolute	occorrenze normalizzate
totale occorrenze di è verbo autonomo	970	1124,8	686	1176,1
Ø	503	583,1	350	600,4
Lessicale preverbale	63	73,1	50	85,7
Lessicale postverbale	47	54,4	27	46,2
Pronom. non personale (Dimostrativo, indefinito, possessivo)	81	93,9	82	140,7
Pron. pers. tonico	7	8,2	6	10,2
l'	165	191,3	111	190,3
gl'	29	33,6	36	61,7
la	1	1,2	1	1,7
e	0	0	0	0
e + l'	6	7	1	1,7
la + l'	1	1,2	0	0
e + gl'	1	1,2	0	0
Dimostrativo + l'	34	39,5	3	5,1
Dimostrativo + gl'	2	2,3	1	1,7
Lessicale prev. + l'	14	19,7	6	10,3
Lessicale prev. + gl'	4	4,6	1	1,7
l'+ Lessicale postv.	2	2,3	6	10,2
gl'+ Lessicale postv.	1	1,2	1	1,7
Tonico + l'	3	3,5	4	6,8
Tonico + gl'	3	3,5	0	0

Riportiamo ora, per ogni tipologia di soggetto, esempi tratti dai due corpora.

a. Soggetto *l'* clitico:

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

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

b. Soggetto *gl'* clitico:

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

(27) *LOR: [279] <eh> // [280] ho vinto / ora *gl'* è poco // (Corpus Confronto Diacronico, ipubl03)

c. Soggetto *la* clitico:

(28) *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)

(29) *CLE: [...] // [42] sì // [43] perché *la* 'un è roba mia // (Corpus Confronto Diacronico, ipubl01b)

d. Raddoppiamento del soggetto clitico *e l'*:

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

(31) *ELA: [84] <eh> // [85] *e l'* è una responsabilità / eh // (Corpus Confronto Diacronico, ifamd08)

e. Raddoppiamento del soggetto clitico *e gl'*²¹:

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

f. Raddoppiamento del soggetto clitico *la l'*:

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

²¹ Questo tipo di raddoppiamento e quello illustrato dall'esempio successivo non sono stati incontrati nel Corpus di Confronto Diacronico, per cui citiamo solo l'esempio del Corpus Stammerjohann.

g. Soggetto Ø:

- (34) *BRA: [...] // [158] lo impiccano / però / siccome Ø è di legno / è chiaro / l'operazione di asfissia dura &tu [///] diverse ore //(Corpus Stammerjohann, pino)

h. Soggetto lessicale preverbale:

- (35) *AND: [864] mah / senti / *il francese* è più facile / eh / dell' inglese // (Corpus Stammerjohann, fran)
- (36) *ART: [112] <col mastice> // [113] *l'ingrediente* principale è il mastice // (Corpus Confronto Diacronico, ifamdl04)

i. Soggetto lessicale postverbale:

- (37) *AMI: [150] è contento *il gatto* / eh / che [/] che [/] che il lupo sia morto // (Corpus Stammerjohann, vita2)
- (38) *MAX: [184] no / <non è questa *la strada*> // (Corpus Confronto Diacronico, ifamdl19)

j. Soggetto pronominale (dimostrativo, indefinito, possessivo):

- (39) *FRA: [859] *questa* è una professoressa / sai // (Corpus Stammerjohann, fran)
- (40) *NIL: [116] *quella* è la procedura che stabilisce il giudice // (Corpus Confronto Diacronico, ifammm09)
- (41) *CAP:[...] [151] *uno* è / quello che riguarda / la mozione che abbiamo fatto / il febbraio scorso // (Corpus Confronto Diacronico, inatpd02)
- (42) *NIL: [...] [34] però di fatto poi / non [///] *nessuno* ne [/] ne è a conoscenza / e quindi / il contribuente / niente / a volte lo dice lui / a dei parenti hhh // (Corpus Confronto Diacronico, ifammm09)
- (43) *PAO: [32] sì / mi sa che *la tua* è una scusa bella e buona // (Corpus Stammerjohann, paol)

k. Pronome tonico soggetto:

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

(45) *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 1124,8 occorrenze (ogni 100.000 parole) della forma *è* quando è un verbo non ausiliare, contro le 1.176,1 occorrenze del *Corpus di Fiorentini*.

La prima impressione fornita dai dati, per ciò che riguarda l'espressione del soggetto, è che non sia avvenuto il calo delle forme clitiche: il clitico *l'* presenta 191,3 occorrenze nel *Corpus Stammerjohann* e 190,3 nel *Corpus di Fiorentini*. Una diminuzione dello 0,52%, pressoché nulla. Nel caso di *gl'* sembra essere avvenuto persino un aumento, si passa da 33,6 occorrenze nel *Corpus Stammerjohann* a 61,7 in quello *di Fiorentini*, un incremento di quasi il doppio delle occorrenze. Le occorrenze assolute di *la* sono piuttosto basse, due per il *Corpus Stammerjohann* e una per il *Corpus di Fiorentini*. Per *e* non sono state riscontrate occorrenze neanche nel *Corpus Stammerjohann*.

Invece sembrerebbe esserci stato un calo nella reduplicazione clitica, ma i numeri sono troppo bassi per tirare conclusioni: si passa dalle 2 occorrenze per i raddoppiamenti *e gl'* e *la l'* del *Corpus Stammerjohann* a nessuna nel *Corpus di Fiorentini*. Il raddoppiamento clitico *e l'* passa da 6 occorrenze assolute nel *Corpus Stammerjohann* a una sola nel *Corpus di Fiorentini*.

A questo punto l'analisi dei dati ha verificato le altre tipologie di espressione del soggetto. L'ipotesi iniziale sosteneva 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 \emptyset . Tuttavia, la riduzione non sembra essere avvenuta: il soggetto viene espresso con forme lessicali, pronominali, clitiche o reduplicate in 541,7 occorrenze del *Corpus Stammerjohann* e in 575,7 nel *Corpus di Fiorentini*. Per il soggetto lessicale²² preverbale, il *Corpus Stammerjohann* ha registrato 73,1 occorrenze e il *Corpus di Fiorentini* 85,7. I soggetti lessicali postverbali sono passati dalle 54,4 occorrenze del *Corpus Stammerjohann* alle 46,2 del *Corpus di Fiorentini*. Tuttavia tali forme non sono oggetto di questo studio perché la loro funzione è assai diversa e la loro presenza

²² Nel computo dei soggetti lessicali sono state inserite anche le proposizioni soggettive (11 nel *Corpus Stammerjohann* e 4 nel *Corpus di Fiorentini*).

è collegata a ragioni che andrebbero sottoposte a un'indagine qualitativa accurata. Le forme pronominali non personali, come gli indefiniti e i dimostrativi (nella maggior parte) e possessivi, fanno invece riscontrare un incremento notevole: si passa da 93,9 occorrenze nel *Corpus Stammerjohann* a 140,7 nel *Corpus di Fiorentini*. Infine, i pronomi personali tonici rimangono praticamente invariati: se ne contano 8,2 nel *Corpus Stammerjohann* e 10,2 nel *Corpus di Fiorentini*.

Apparentemente quindi, il calo dei pronomi clitici soggetto così atteso, e in generale dell'espressione del soggetto in fiorentino, non avrebbe risentito dei quarant'anni di pressione dell'italiano standard. Tuttavia, quando vengono disaggregati e confrontati i dati è possibile ottenere un quadro diverso.

Osservando unicamente i clitici si ottiene la tabella seguente:

Tabella 10. Forme di soggetto clitico e reduplicato con clitico incontrate nei due corpora, in termini assoluti e normalizzati.

forme	Corpus Stammerjohann		Corpus di Fiorentini	
	occorrenze assolute	occorrenze normalizzate	occorrenze assolute	occorrenze normalizzate
l'	165	191,3	111	190,3
gl'	29	33,6	36	61,7
la	1	1,2	1	1,7
e	0	0	0	0
e + l'	6	7	1	1,7
la + l'	1	1,2	0	0
e + gl'	1	1,2	0	0
Dimostrativo + l'	34	39,5	3	5,1
Dimostrativo + gl'	2	2,3	1	1,7
Lessicale prev. + l'	14	19,7	6	10,3
Lessicale prev. + gl'	4	4,6	1	1,7
l' + lessicale postv.	2	2,3	6	10,2
gl' + lessicale postv.	1	1,2	1	1,7
Tonico + l'	3	3,5	4	6,8
Tonico + gl'	3	3,5	0	0
Totale	266	312,1	171	292,9

È bene notare che le occorrenze di soggetto lessicale preverbale riprese dai clitici soggetto *l'* e *gl'* hanno fatto registrare una diminuzione dei casi; nei soggetti lessicali postverbalmente accompagnati da *l'* invece è avvenuto un certo aumento. Nel *Corpus Stammerjohann* difatti ne sono stati incontrati 2,3 esempi, mentre nel *Corpus di Fiorentini* ve ne sono 10,2. Tuttavia i valori, almeno per il primo caso, sono esigui per poter formulare ipotesi. Riportiamo sotto un esempio che chiarisce questo tipo di occorrenza:

- (46) *ELA: [354] <scusa &i> [/] indove l' è &ca [/] *Castiglion de' Pepoli* ? // (Corpus Confronto Diacronico, ifamcv01)

Nel caso dei pronomi dimostrativi (*questo/a* e *quello/a*) reduplicati dai clitici soggetto *l* e *gl'* si registra certamente un calo: mentre nel *Corpus Stammerjohann* le occorrenze totali sono 41,8, nel *Corpus di Fiorentini* sono solamente 6,8. Un esempio tratto dai corpora può chiarire il tipo di occorrenza:

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

Quanto ai pronomi tonici ripresi da clitici, si nota un certo aumento nel *Corpus di Fiorentini*, ma i numeri sono troppo bassi per avanzare ipotesi.

Sulla base dei dati rilevati, si potrebbe supporre che sia avvenuta, nel tempo, una variazione da parte dei parlanti nell'utilizzo della forma clitica: negli anni Sessanta la preferenza sarebbe stata per la reduplicazione dei dimostrativi, mentre in tempi recenti la reduplicazione si verificherebbe, in forma minore, con i soggetti lessicali posposti.

4.1.1 *Analisi sociolinguistica*

L'analisi è quindi continuata con l'osservazione dei dati sociolinguistici dei parlanti. Il fattore che è emerso come più rilevante è stato quello dell'età²³. Sono stati confrontati tali dati in entrambi i corpora.

I criteri di distribuzione per età seguono quelli della famiglia 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 40 anni; la fascia C quelli tra 40 e 60 anni; infine la fascia D quelli al di sopra dei 60 anni²⁴. Una particolarità del *Corpus Stammerjohann* è che i quattro parlanti della fascia A che usano i clitici soggetto di cui ci occupiamo sono bambini²⁵, aspetto, come vedremo, importante per valutare i dati.

Il *Corpus Stammerjohann* è composto da 59 parlanti mentre il *Corpus di Fiorentini* ne ha 72.

²³ Sono stati verificati anche altri fattori sociolinguistici, come il livello scolastico e la professione dei parlanti, ma nessuno di essi ha mostrato effetti nell'uso delle forme di espressione del soggetto così evidenti come il fattore età.

²⁴ Quando non si conosce l'età del parlante è utilizzato il simbolo X.

²⁵ È stato possibile dedurre questo dato ascoltando i testi nella loro interezza: innanzitutto ne emergono voci chiaramente infantili; in secondo luogo, nel caso degli informatori *FRA, *ALE e *AND, dai loro interventi si desume che frequentano la quinta elementare. Nel caso dell'informatore *BAM non è chiara quale sia l'età ma si tratta comunque di un bambino più giovane degli altri tre, che sta raccontando una storia al maestro.

È stato effettuato uno spoglio minuzioso di tutte le forme di espressione del soggetto per ogni parlante: sono state suddivise le forme Ø, quelle lessicali, pre e postverbal, i vari pronomi tonici (dimostrativi, indefiniti e possessivi) oltre ai clitici soggetto e ai diversi abbinamenti delle reduplicazioni²⁶. Terminato il conteggio, sono state calcolate le percentuali di ogni forma di possibile espressione del soggetto per ogni fascia di età²⁷. I soggetti clitici sono stati sommati alle diverse forme di reduplicazioni clitiche; tutte le forme pronominali non clitiche compaiono come “altro pronome”; i soggetti lessicali preverbal e postverbal sono stati mantenuti separati. La fascia X, che indica i parlanti di cui non si conosce l'età è stata esclusa da quest'analisi²⁸. I risultati sono presentati nella tabella 11.

Tabella 11. Totale, in percentuale, delle forme di espressione del soggetto distribuite in base all'età dei parlanti

	Corpus stammerjohann			Corpus di Fiorentini			
	A	B	C	A	B	C	D
Soggetto Ø	50	66,3	49,9	62,1	65	40,8	36
Sogg. lessicale preverbale	6,1	8,2	6,5	7,8	6,6	6,8	10
Sogg. lessicale postverb.	4,9	4,1	4,6	7,8	4,1	2,9	0
Sogg altro pronome	4,9	8,2	9,5	10,7	11,6	16,8	2
Soggetto clitico	34,1	13,2	29,5	11,6	12,7	32,7	52

Nel *Corpus Stammerjohann*, del totale di tutte le forme possibili di espressione del soggetto, nella fascia A, quelle con i clitici rappresentano il 34,1%. Si ricorda che tutte le occorrenze sono pronunciate da bambini, figli di alcuni degli informatori della fascia C. Nella fascia B, invece, i clitici sono una alternativa meno utilizzata: è stata rinvenuta solo nel 13,2% dei casi. Infine, la fascia C presenta una percentuale simile alla fascia A (29,5%). Mancano del tutto parlanti della fascia D.

Una prima osservazione delle fasce d'età ha perciò indotto le seguenti considerazioni: si sarebbe verificata una chiara riduzione dei clitici soggetto nella fascia B, corrispondente ai parlanti nati possibilmente tra gli anni Trenta e Quaranta.

²⁶ Per motivi di spazio, non riproduciamo qui le tabelle che si riferiscono ad ogni parlante.

²⁷ Includo le forme Ø.

²⁸ I dati di questa fascia sono comunque stati analizzati, ma non sono stati rilevanti per questa ricerca.

Se non vi è conferma di ciò nella fascia A sarebbe solo perché si tratta di bambini, che riproducono l'uso dei genitori (parlanti della fascia C), visto che questo è ancora il loro principale modello. Si può ipotizzare che crescendo siano stati sottoposti a modelli diversi 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'11,6% dei casi. Anche la fascia B è piuttosto restia nell'avvalersi dei clitici come soggetto, utilizzandoli nel 12,7% degli esempi incontrati. Un uso molto maggiore si registra nella fascia C che li usa nel 32,7% dei casi. La fascia D infine è quella che se ne avvale maggiormente, nel 52% dei casi²⁹.

Osservando 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, presenta un tipo di produzione che ancora riproduce le forme d'uso dei genitori, vale a dire della fascia C. Quindi le fasce su cui basarsi sarebbero solo due: la B e la C. Analogamente, nel *Corpus di Fiorentini* sembra essersi registrata una progressiva riduzione dei clitici dalla fascia D alla C e alla B e, in misura molto meno significativa, anche tra la fascia B e la A.

Nel *Corpus Stammerjohann* la fascia C corrisponde all'incirca a coloro che sono nati fra il 1925 e il 1940. Non sono disponibili dati per le generazioni anteriori e quindi si potrebbe concludere che, perlomeno a partire dai nati intorno al 1940, ci sarebbe stata una progressiva e costante riduzione dei clitici, ma che questa riduzione potrebbe senz'altro essere iniziata prima.

Tuttavia, osservando attentamente i dati, la situazione è più complessa. I parlanti della fascia B del *Corpus Stammerjohann* appartengono alla stessa generazione di quelli della fascia D del *Corpus di Fiorentini*. Solo che i primi presentano un uso dei clitici del 13,2% mentre i secondi del 52%. Inoltre, i parlanti della fascia C del *Corpus di Fiorentini*, più giovani dei parlanti della fascia B dello *Stammerjohann*, presentano una percentuale di clitici del 32,7% di contro al 34,1% dei parlanti della fascia A del *Corpus Stammerjohann*.

Si è quindi verificata nei dettagli la composizione dei due corpora e dei vari testi, ipotizzando che i tagli operati per mantenere nel *Corpus di Fiorentini* la composizione con parlanti esclusivamente del capoluogo, e non generalmente

²⁹ È opportuno osservare che la fascia D di C-ORAL-ROM corrisponde grosso modo a quella B del corpus *Stammerjohann*, cioè ai nati durante la guerra.

toscani come nel *Corpus di Confronto Diacronico*, potesse aver compromesso la comparabilità con lo *Stammerjohann* in altri aspetti.

Le diverse tipologie testuali dei due corpora non sembrano presentare differenze significative, perlomeno non sufficienti a giustificare la situazione contraddittoria sopra esposta. Si è allora analizzata nuovamente la composizione dei parlanti e dei vari testi in cui compaiono per verificare se qualcuno di essi fosse sovrarappresentato. In effetti, nel *Corpus di Fiorentini*, sono stati incontrati tre informatori che compaiono in molti testi o pronunciano un numero di clitici molto maggiore degli altri, due appartenenti alla fascia C e uno alla D:

- *ELA, della fascia C, è presente in 5 testi, con 25 clitici;
- *LUC, della fascia C, è presente in 1 testo, ma con ben 21 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.

Quindi, questi 3 parlanti sono stati esclusi dal conteggio dei clitici. Ne è risultato un quadro molto diverso che presentiamo nella tabella 12.

Tabella 12. Totale dei soggetti incontrati nei due corpora a confronto, in percentuale, in base all'età dei parlanti, con esclusione degli informatori *ELA e *LUC (fascia C) e *LUC (fascia D) del *Corpus di Fiorentini*

	Corpus stammerjohann			Corpus di Fiorentini			
	A	B	C	A	B	C	D
Soggetto o	50	66,3	49,9	62,1	65	46,7	56,5
Sogg. lessicale preverbale	6,1	8,2	6,5	7,8	6,6	9,3	13,1
Sogg. lessicale postverb.	4,9	4,1	4,6	7,8	4,1	4	0
Sogg altro pronomi	4,9	8,2	9,5	10,7	11,6	16,9	0
Soggetto clitico	34,1	13,2	29,5	11,6	12,7	23,1	30,4

Tuttavia, anche senza questi parlanti, la presenza dei clitici nel *Corpus di Fiorentini* continua a essere superiore a quella dello *Stammerjohann*: la fascia D del corpus più recente presenta una percentuale di uso dei clitici del 23,1%, superiore alla percentuale di clitici della fascia corrispondente dello *Stammerjohann* (cioè la fascia B, composta dai nati negli stessi anni dei rappresentanti della fascia D nel *Corpus di Fiorentini*), che presenta una percentuale del 13,2%.

In ogni caso, se analizzati separatamente, entrambi i corpora confermano la riduzione dei clitici nel tempo. Naturalmente questo dato è in evidente contradd-

dizione con la maggiore presenza di clitici nel corpus più tardo rispetto a quello più antico, ma allo stesso tempo non si può pensare che ciò che si riduce nel tempo dentro due corpora sia poi aumentato nel tempo. Non resta che supporre che aspetti più fini delle interazioni nel *Corpus di Fiorentini* siano responsabili di una maggiore induzione all'uso dei clitici mentre le caratteristiche corrispondenti del *Corpus Stammerjohann* inibiscono (rispetto all'altro corpus) l'uso delle stesse forme. Questo resta un punto da verificare in futuro, indagando più a fondo le strutture che inducono l'uso dei clitici ed estendendo la ricerca a tutti i verbi e ad altri pronomi soggetto clitici.

Allo stato, dobbiamo mettere in rilievo la costanza con cui in entrambi i corpora ogni generazione mostra un aumento dei soggetti \emptyset e una riduzione delle forme clitiche. L'unica eccezione è la fascia A del *Corpus Stammerjohann*, che, come abbiamo visto, è composta di bambini, che, quindi, riproducono l'uso dei genitori.

Analizzando i dati in questo modo, si può ipotizzare che la riduzione dei clitici e l'aumento dei soggetti \emptyset fossero tendenze già in atto nel parlato della generazione nata intorno al 1935 e il 1940; che questa tendenza sia proseguita con un'intensità simile nella generazione nata intorno al 1955 e il 1960; e che la tendenza, pur proseguendo, mostri una riduzione di intensità nei nati intorno al 1970 e il 1975.

Se ipotizziamo che l'uso proprio della generazione nata intorno al 1935-40 si manifesti pienamente intorno alla metà degli anni Cinquanta, dobbiamo concludere che il fenomeno certamente non sarebbe cominciato dopo questa data ma potrebbe essere cominciato prima e anche molto prima.

4.2 Verbi *fare*, *andare* e *venire*

La scelta del verbo *essere* all'indicativo della terza persona singolare influisce decisamente sui risultati del clitico *e*, che non viene utilizzato dinanzi a vocale. Si è deciso quindi di analizzare altri verbi ad alta frequenza in cui si potesse effettivamente valutare l'uso di *e*. Un'attenta analisi dei lemmi verbali ad alta frequenza ha fatto ricadere la scelta sui verbi: *fare*, *andare* e *venire*. Per garantire maggiore coerenza ai dati, si è deciso di concentrarsi solo sulle forme al presente indicativo. Il numero totale delle forme è comunque esiguo, soprattutto per quanto riguarda i clitici.

4.2.1 Analisi di *fa*

La ricerca sul *Corpus Stammerjohann* ha rilevato 294 occorrenze della forma *fa*. Successivamente sono stati scartati dal computo:

- gli errori di trascrizione;
- le frasi interrotte o totalmente incomprensibili (tutti i casi sono stati ascoltati per verificare l'effettiva comprensibilità);
- i casi di frasi fisse come “x tempo fa”;
- i casi di espressioni meteorologiche come “fa caldo/fa freddo”, senza soggetto;
- i casi di espressione del tipo “si fa” che in fiorentino equivale a “noi facciamo”;
- i casi di elisione, in cui *fa'* indica l'infinito o la seconda persona singolare dell'imperativo;
- i casi in cui *fa* è inserito nei commenti dei trascrittori.

In questo modo il totale delle occorrenze analizzabili è sceso a 42 casi. Lo stesso trattamento dei dati è stato adottato per il *Corpus di Fiorentini*, dove le occorrenze di *fa* riportate dal software *AntConc* sono scese da 239 a 69.

Tabella 13. Forme soggetto incontrate nei due corpora a confronto, in termini assoluti e normalizzati.

forme	Corpus Stammerjohann		Corpus di Fiorentini	
	occorrenze assolute	occorrenze normalizzate	occorrenze assolute	occorrenze normalizzate
totale occorrenze di <i>fa</i>	42	48,7	69	118,3
Ø	34	39,42	47	80,6
Lessicale preverbale	0	0	2	3,44
Lessicale postverbale	0	0	5	8,6
Pronom.non personale (Dimostrativo, indefi- nito, possessivo)	5	5,8	6	10,3
Sogg. pronome tonico	2	2,32	3	5,1
L'	0	0	1	1,7
Gl'	0	0	0	0
La	0	0	5	8,6
E	0	0	0	0
Dimostrativo + la	1	1,16	0	0

Si osservi che, dal primo al secondo corpus, si segnala un aumento nell'uso del clitico soggetto *la*. Comunque, non è stato incontrato nessun esempio della forma *e* in nessuno dei due corpora. Nel caso del soggetto Ø, passando dal corpus più antico a quello più recente, si osserva un incremento notevole.

Sono stati infine verificati i dati relativi ai parlanti che utilizzano le forme clitiche nel *Corpus di Fiorentini*: uno appartiene alla fascia B, uno alla C e uno

alla D. La più anziana, *LUC, appartenente alla fascia D, è colei che pronuncia il maggior numero di clitici, ben quattro su sei.

4.2.2 *Analisi di va*

Presentiamo ora i dati relativi all'analisi della forma verbale *va*. La metodologia utilizzata è la stessa adottata per *fa*. Dopo una prima ricerca con il software *Ant-Conc*, che ha fatto segnalare nel *Corpus Stammerjohann* 278 occorrenze di *va*, sono stati scartati dal computo:

- gli errori di trascrizione;
- le frasi interrotte o totalmente incomprensibili (tutti i casi sono stati ascoltati per verificare l'effettiva comprensibilità);
- i casi di frasi fisse come “*va beh, va bene*”;
- i casi di elisione, in cui *va* indica la seconda persona singolare dell'imperativo;
- i casi in cui *va* è utilizzato con valore deontico, come “*va detto*” nel senso di “*deve essere detto*”;
- i casi in cui *va* è inserito nei commenti dei trascrittori.

In questo modo il totale delle occorrenze analizzabili è sceso a 92 casi. Lo stesso trattamento dei dati è stato adottato per il *Corpus di Fiorentini*, dove le occorrenze di *va* riportate dal software *AntConc* sono scese da 192 a 66.

Anche in questo caso le cifre sono state normalizzate per 100.000 e sono stati analizzati i dati linguistici, che vengono presentati nella tabella sotto.

Tabella 14. Forme soggetto incontrate nei due corpora, in termini assoluti e normalizzati.

forme	Corpus Stammerjohann		Corpus di Fiorentini	
	occorrenze assolute	occorrenze normalizzate	occorrenze assolute	occorrenze normalizzate
totale occorrenze di <i>va</i>	92	106,6	66	113,2
Ø	53	61,4	42	72
Lessicale preverbale	11	12,75	5	8,6
Lessicale postverbale	2	2,3	8	13,7
Pronom. non personale (Dimostrativo, indefi- nito, possessivo)	6	7	3	5,1
Sogg. pronomi tonico	4	4,6	2	3,4
L'	0	0	0	0
Gl'	0	0	0	0
La	12	13,9	6	10,3
E	3	3,5	0	0
Tonico + e	1	1,1	0	0

La forma a soggetto Ø ha mostrato un incremento nel corpus più recente, come sosteneva l'ipotesi iniziale. In genere le varie forme di espressione del soggetto hanno subito un calo in diacronia, ad eccezione di quella lessicale postverbale³⁰. I soggetti clitici sono utilizzati in pochi casi: di *la* sono state registrate 12 occorrenze nello *Stammerjohann* e la metà del *Corpus di Fiorentini*; *e* è stato incontrato solo quattro volte, sempre nel *Corpus Stammerjohann*.

Per quanto riguarda i dati sociolinguistici, nuovamente registriamo la predominanza di *LUC della fascia D che pronuncia tre delle sei occorrenze di clitici soggetto *la* trovati nel *Corpus di Fiorentini*.

4.2.3 *Analisi di viene*

La ricerca con il software *AntConc* ha rilevato 66 occorrenze della forma *viene* nel *Corpus Stammerjohann* e 44 occorrenze nel *Corpus di Fiorentini*. Sono stati scartati dal computo:

- gli errori di trascrizione;
- le frasi interrotte o totalmente incomprensibili (tutti i casi sono stati ascoltati per verificare l'effettiva comprensibilità);
- i casi in cui *viene* è utilizzato con valore di diatesi passiva;
- i casi in cui *viene* è inserito nei commenti dei trascrittori.

In questo modo il totale delle occorrenze analizzabili è sceso a 30 casi nel *Corpus Stammerjohann* e 18 nel *Corpus di Fiorentini*.

I dati sono quindi stati normalizzati per 100.000 e analizzati, come mostra la tabella 15.

Si osserva un aumento in diacronia dell'uso del clitico *la* e della forma Ø. Sono presenti due unici casi del clitico soggetto *e*, tutti nel *Corpus Stammerjohann*. Nel *Corpus Stammerjohann*, appare un unico caso di reduplicazione, nel quale il tonico è duplicato dal clitico *la*.

Per quanto riguarda l'età dei parlanti, nel *Corpus Stammerjohann* tutti i clitici sono pronunciati dalla fascia C, mentre nel *Corpus di Fiorentini*, dalle due fasce più anziane, C, con due occorrenze e D, con tre³¹.

³⁰ I soggetti lessicali postverbal non verranno commentati in questa sede giacché, come è stato spiegato anteriormente per la forma verbale *è*, la loro motivazione è di natura diversa.

³¹ Pronunciate ancora una volta dall'informatore *LUC della fascia D.

Tabella 15. Forme soggetto incontrate nei due corpora a confronto, in termini assoluti e normalizzati.

forme	Corpus Stammerjohann		Corpus di Fiorentini	
	occorrenze assolute	occorrenze normalizzate	occorrenze assolute	occorrenze normalizzate
totale occorrenze di <i>va</i>	30	34,8	18	30,9
Ø	9	10,4	9	15,4
Lessicale preverbale	1	1,1	2	3,4
Lessicale postverbale	15	16,1	1	1,7
Pronom.non personale (Dimostrativo, indefi- nito, possessivo)	0	0	1	1,7
Sogg. pronome tonico	0	0	0	0
L'	0	0	0	0
Gl'	0	0	0	0
La	2	2,3	5	8,6
E	2	2,3	0	0
Tonico + la	1	1,1	0	0

4.2.4 Conclusioni su *fa*, *va* e *viene*

Osserviamo adesso congiuntamente i dati relativi a *fa*, *va* e *viene*. Ancora una volta le caratteristiche che spiccano sono quelle che riguardano le fasce d'età dei parlanti. Naturalmente la fascia X, che indica i parlanti di cui non si conosce l'età, è stata esclusa. I risultati sono osservabili nella tabella seguente.

Tabella 16. Totale dei soggetti incontrati nei due corpora a confronto, in percentuale, in base all'età dei parlanti

	Corpus Stammerjohann			Corpus di Fiorentini			
	A	B	C	A	B	C	D
Ø	76	43,5	56,7	82,1	59,6	60,5	55,6
Lessicale prev.	4	13	7,8	0	8,5	7	3,7
Lessicale postv.	0	17,4	10,2	17,9	12,8	4,6	7,4
Pronome non per- sonale	12	17,4	9	0	17	14	3,7
Prono- meclitico	8	8,7	16,3	0	2,1	13,9	29,6

La situazione non è del tutto corrispondente a quella ricavata dai dati della forma verbale *è*. Nel caso del *Corpus Stammerjohann* esistono differenze consistenti nelle percentuali dei parlanti della fascia A (che ricordiamo essere bambini) e quelli della fascia C (corrispondenti ai genitori). Tuttavia bisogna sottolineare che, con questi verbi, la partecipazione della fascia A nei testi è minima.

Ciò può spiegare il calo così consistente da una fascia all'altra, a differenza di ciò che era stato registrato per la forma *è*.

Tralasciando questo fattore, si può comunque confermare la tendenza già osservata nel comportamento dei parlanti delle fasce B e C: gli informatori della fascia C utilizzano il doppio di clitici degli appartenenti alla fascia B.

Nel *Corpus di Fiorentini*, si osservi che i clitici soggetto sono pronunciati soprattutto dalla fascia D, quella dei più anziani, confermando parzialmente l'analisi di *è*. Sono stati controllati gli informatori di questa fascia e la loro partecipazione nei testi analizzati: si è osservato che la maggioranza dei clitici soggetto realizzati appartengono ad un unico informatore, *LUC. Se tale parlante venisse ritirato dagli esempi, la percentuale verrebbe ridotta notevolmente e la situazione diventerebbe più simile a quella osservata nel caso di *è*.

5. Conclusioni

Lo studio ha presentato un'analisi delle forme clitiche soggetto *gl'*, *l'*, *la*, *e* del vernacolo fiorentino su due corpora ritenuti comparabili, in uno studio diacronico. L'obiettivo era valutare l'effettiva vitalità di tali clitici nell'italiano dei fiorentini del 1965 e di quello degli anni 2000-2005 partendo dall'ipotesi che con il passare degli anni si sarebbe registrata una forte riduzione nel loro uso.

L'ipotesi di base era che il calo, se non la quasi scomparsa, nell'uso dei clitici soggetto del fiorentino, dovuta alla pressione dell'italiano standard, si sarebbe registrata specialmente tra la fine degli anni Sessanta e gli anni Settanta, come effetto della scolarizzazione di massa, della diffusione della televisione e della maggiore mobilità della popolazione.

Un'analisi dei dati, che ha tenuto in particolar conto le caratteristiche degli informatori e specialmente la loro fascia d'età³², ha evidenziato che, effettivamente, si segnala un calo nella vitalità dei clitici soggetto e, più in generale, l'aumento del soggetto \emptyset . Tuttavia, l'ipotesi sul periodo di inizio della diminuzione non è stata confermata dai dati: osservando attentamente il *Corpus Stammerjohann*, la perdita di vitalità dei clitici soggetto è già percepibile nella generazione nata prima della seconda guerra mondiale, intorno al 1925, nella quale si nota una riduzione notevole a confronto della generazione che la precede. Non è possibile escludere che il fenomeno abbia avuto inizio ancor prima, visto che gli informatori più anziani del *Corpus Stammerjohann* sono di una sola generazione anteriore a quelli della fascia che mostra che la riduzione era già in corso. La fi-

³² Oltre alla sovrarappresentazione di alcuni di essi nel Corpus di Fiorentini.

ne degli anni Sessanta e gli anni Settanta, quindi, non sembrano costituire il punto di partenza del fenomeno e nemmeno il momento di accelerazione di esso. I dati qui presentati mostrano al contrario che tale situazione compare almeno a partire dalla seconda generazione presente nel corpus e continua, in maniera meno intensa, nelle generazioni successive.

Questo studio non può comunque dare risposte definitive né differenziare il comportamento dei diversi clitici a causa dell'eseguità dei dati. Le conclusioni esposte dovranno essere confermate e precisate da ricerche successive su tutti i verbi e su altre persone. Mentre questo articolo sta andando in stampa possiamo aggiungere che una ricerca recente fatta sui pronomi di seconda persona (Cardoso 2015) conferma la sostanza delle nostre conclusioni.

Ringraziamenti

Questa ricerca è stata possibile grazie ai finanziamenti della Capes - Comissão de Aperfeiçoamento de Pessoal do Nível Superior (Programa PDSE, processo n. 99999.012164/2013-05), della Fapemig e del CNPq.

Riferimenti bibliografici

- Berruto, G. 1997. Linguistica del contatto e aspetti dell'italianizzazione dei dialetti: appunti di creolistica casalinga. In G. Holtus, J. Kramer & W. Schweickard (eds), *Italica et Romanica. Festschrift für Max Pfister zum 65. Geburtstag*, Vol. III. Tübingen: Niemeyer, 13-29.
- Berruto, G. 2005. Dialect/standard convergence, mixing, and models of language contact: the case of Italy. In P. Auer, F. Hinskens & P. Kerswill (eds), *Dialect change. Convergence and divergence in European languages*. Cambridge: Cambridge University Press, 81-97.
- Binazzi, N. 1997. *Le parole dei giovani fiorentini: variazione linguistica e variazione sociale*. Roma: Bulzoni.
- Binazzi, N. 1998. 'Italiano' e 'dialetto' a Firenze: 50 giovani giudicano il loro lessico d'uso. In Ž. Muljačić (eds), *L'italiano e le sue varietà linguistiche*. Sauerländer: Verlag für deutsch-italienische Studien, 13-32.
- Binazzi, N. 2007. L'appartenenza rivelata. Lessico e tradizioni del discorso nel parlato Fiorentino. *Quaderni del Dipartimento di Linguistica dell'Università di Firenze XVII*: 137-168.
- Bracco, C., Brandi, L. & Cordin, P. 1985. Sulla posizione soggetto in italiano e in alcuni dialetti dell'Italia centro-settentrionale. In A. Franchi de Bellis & L.M. Savoia (eds), *Sintassi e morfologia della lingua italiana d'uso: Teorie e applicazioni descrittive - Atti del XVII Convegno Internazionale di Studi, Urbino 11-13 settembre 1983*. Roma: Bulzoni.

- Brandi, L. & Cordin, P. 1981. Dialetti italiani: un confronto sul parametro del soggetto nullo. *Rivista Di Grammatica Generativa* 6: 33-87.
- Brandi, L. & Cordin, P. 1983. Subject clitics in two Italian dialects: a perspective on the Null subject Parameter. *Journal of Italian Linguistics* 8(2): 1-24.
- Brandi, L. & Cordin, P. 1989. Two Italian dialects and the null subject parameter. In O. Jaeggli & K. Safir (eds), *The Null Subject Parameter*. Dordrecht: Kluwer, 111-142.
- Cardinaletti, A. & Munaro N. (eds). 2009. *Italiano, italiani regionali e dialetti*. Milano: Franco Angeli.
- Cardoso, G. 2015. O pronome de segunda pessoa no florentino. Uma análise diacrônica baseada em corpora. BA Thesis, Belo Horizonte: UFMG.
- Cresti, E. 2000. *Corpus di italiano parlato*. Firenze: Accademia della Crusca.
- Cresti, E. 2010. Clitici e relazioni anaforiche nell'articolazione dell'informazione: una ricerca corpus-driven nel parlato italiano (C-ORAL-ROM). In M. Iliescu, H. Siller-Runggaldier & P. Danler (eds), *XXVe CILPR Congrès International de Linguistique et de Philologie Romanes. Innsbruck, 3 – 8 septembre 2007*. Berlin/New York: De Gruyter.
- Cresti, E. & Moneglia, M. (eds). 2005. *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: Benjamins.
- Ferrari, L.A. 2015. Aspectos prosódicos e sintáticos dos pronomes clíticos em Português do Brasil e no vernáculo florentino. PhD diss., Belo Horizonte: UFMG.
- Firenzuoli, V. & Signorini, S. 2003. L'unità informativa di topic: correlati intonativi. In G. Marotta (eds), *La coarticolazione: Atti delle XIII Giornate di Studio del Gruppo di Fonetica Sperimentale*. Pisa: ETS, 177-184.
- Galli De' Paratesi, N. 1984. *Lingua toscana in bocca ambrosiana. Tendenze verso l'italiano standard: un'inchiesta sociolinguistica*. Bologna: il Mulino.
- Giannelli, L. 2000. Toscana. In A. Zamboni (ed.), *Profilo dei Dialetti Italiani*. Ospitaletto: Pacini Editore.
- Kayne, R. 1989. Null Subjects and Clitic Climbing. In O. Jaeggli & K. Safir (eds), *The Null Subject Parameter*. Dordrecht: Kluwer, 239-261.
- Kayne, R. 1991. Romance Clitics, Verb Movement, and PRO. *Linguistic Inquiry* 22: 647-686.
- Manzini, M.R. & Savoia, L.M. 2005. *I dialetti italiani e romanci. Morfosintassi generativa*. Alessandria: Edizioni dell'Orso.
- Martin, P. 2005. WinPitch Corpus: A text-to-speech analysis and alignment tool. In E. Cresti, & M. Moneglia (eds), *C-ORAL-ROM: Integrated reference corpora for spoken Romance languages*. Amsterdam: John Benjamins, 40-51.
- Moneglia, M., Binazzi, N., Cella, R., Scarano, A., Panunzi, A. & Fabbri, M. 2008. L'incidenza del lessico fiorentino nella lingua d'uso a Firenze. Un confronto tra il corpus Stammerjohann del 1965 e un corpus di parlato contemporaneo. In E. Cresti (eds), *Prospettive nello studio del lessico Italiano. Atti del IX Congresso internazionale della Società di Linguistica e Filologia Italiana*. Firenze: Firenze University Press, 99-106.
- Moneglia, M. & Raso, T. 2014. Notes on Language into Act Theory (L-Act). In T. Raso & H. Mello (eds), *Spoken Corpora and Linguistic Studies*. Amsterdam/Philadelphia: John Benjamins, 468-494.
- Palermo, M. 1997. *L'espressione del pronome personale soggetto nella storia dell'italiano*. Roma: Bulzoni.
- Panunzi, A. 2010. *La variazione semantica del verbo essere nell'italiano parlato. Uno studio su corpus*. Firenze: Firenze University Press.

- Ramos, J. 1997. O uso das formas você, ocê e cê no Dialeto Mineiro. In D. Hora (ed.), *Diversidade Lingüística no Brasil*. João Pessoa: Ideia, 43-60.
- Renzi, L. & Vanelli, L. 1983. I pronomi soggetto in alcune varietà romanze. In *Scritti linguistici in onore di G.B. Pellegrini*. Pisa: Pacini.
- Renzi, L. 1983. Fiorentino e italiano: storia dei pronomi personali soggetto. In F. Albano Leoni (ed.), *Italia linguistica. Idee, storia, strutture*. Bologna: il Mulino, 223-239.
- Renzi, L. 1992. I pronomi soggetto in due varietà substandard: fiorentino e français avancé. *Zeitschrift für romanische Philologie* 108(1-2): 72-98.
- Rizzi, L. 1982. Negation, wh-movement and the Null Subject Parameter. In *Scritti linguistici in onore di G.B. Pellegrini*. Pisa: Pacini.
- Rizzi, L. 1986a. Null objects in Italian and the theory of pro. *Linguistic Inquiry* 17: 501-557.
- Rizzi, L. 1986b. On the status of subject clitics in Romance. In O. Jaeggli & C. Silva-Corvalan (eds), *Studies in Romance linguistics*. Dordrecht: Foris, 391-419.
- Scarano, A. (ed.). 2005. *Corpus di confronto diacronico LABLITA*. <http://lablita.dit.unifi.it/corpora/descriptions/corpusconfronto> (accessed November 2, 2012).
- Scarano, A. & Moneglia, M. 2008. Il Corpus Stammerjohann. Il primo corpus di italiano parlato in rete nella base dati di LABLITA. In M. Pettorino (ed.), *Atti del Congresso Internazionale "La comunicazione Parlata", Napoli 23-25 Febbraio 2006*. Napoli: Liguori editore.
- Scarano, A. & Signorini, S. 2005. Corpus linguistics and diachronic variability. A study on Italian spoken language corpora from the 1960s until nowadays. In J. Kabatek, C.D. Pusch & W. Raible (eds), *Korpora und diachrone Sprachwissenschaft / Romance Corpus Linguistics II: Corpora and Diachronic Linguistics*. Tübingen: Narr, 191-202.
- Signorini, S. 2003. Il topic: Criteri di identificazione e correlati morfosintattici in un corpus di italiano parlato. In F. Albano Leoni, F. Cutugno, M. Pettorino & R. Savy (eds), *Atti del convegno nazionale Il parlato italiano*. Firenze: Cesati, 227-238.
- Signorini, S. 2005. Topic e soggetto in corpora di italiano parlato spontaneo. PhD diss., Università di Firenze.
- Signorini, S. & Tucci, I. 2005. Il restauro e l'archiviazione elettronica del primo corpus di italiano parlato: il corpus Stammerjohann. In A. De Dominicis, L. Mori & M. Stefani (eds), *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, 119-126.
- Sobrero, A. 1997. *Italianization of the dialects*. In M. Maiden & M. Parry (eds), *The dialects of Italy*. London/New York: Routledge, 412-418.
- Stammerjohann, H. 1970. Strukturen der Rede. Beobachtungen an der Umgangssprache von Florenz. *Studi di Filologia Italiana* XXVIII: 295-397.
- Vitral, L. 1996. A forma cê e a noção de gramaticalização. *Revista de Estudos da Linguagem* 5(4).
- Vitral, L. & Ramos, J. 2006. *Gramaticalização: uma abordagem formal*. Belo Horizonte: Faculdade de Letras FALE/UFMG.