IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

C’è con soggetto plurale. La realizzazione variabile di un tratto sub-standard dell'italiano contemporaneo

This is the author's manuscript

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/150958 since 2016-06-06T16:54:46Z

Publisher:

Edizioni dell'Orso

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

17 September 2026

UNIVERSITÀ DEGLI STUDI DI TORINO

Questa è la versione dell'autore dell'articolo:

Massimo Cerruti (2014),

***C'è* con soggetto plurale.**

La realizzazione variabile di un tratto sub-standard dell'italiano contemporaneo.

In: Massimo Cerruti, Elisa Corino, Cristina Onesti (a cura di),

Lingue in contesto. Studi di linguistica e glottodidattica sulla variazione diafasica,

Dell'Orso, Alessandria, pp. 53-76.

C'è con soggetto plurale. La realizzazione variabile di un tratto sub-standard dell'italiano contemporaneo

Massimo Cerruti
Università di Torino

1. Introduzione

Nel presente contributo si affronta un fenomeno che in italiano come in altre lingue è soggetto a variazione diafasica: il mancato accordo tra una forma locativo-esistenziale-presentativa e un soggetto plurale postverbale (come, ad esempio, in *c'è quattro gatti*). Dopo un breve inquadramento generale rispetto ai meccanismi di accordo tra soggetto e verbo (§ 1.1), il fenomeno è indagato in relazione a quali fattori linguistici ed extra-linguistici siano in grado di incidere, positivamente o negativamente, sulla sua realizzazione (§ 2). Ci si avvale a questo proposito dei risultati di uno studio quantitativo condotto con il programma di analisi multivariata Goldvarb x, alcuni dei quali esposti in Berruto/Cerruti (in stampa); guardando anche, più cursoriamente, alla presenza del fenomeno nella comunicazione mediata del computer (§ 2.1).

1. 1 Generalità

Alcuni fra i tratti strutturali dell'italiano contemporaneo sensibili alla variazione diafasica sono interessati da meccanismi di accordo, specie tra soggetto e verbo. Un fenomeno assai noto, ad esempio, è la cosiddetta concordanza a senso, particolarmente frequente con collettivi, come in (1), e con partitivi plurali retti o da collettivi (2) o da quantificatori semanticamente plurali (ancorché formalmente singolari, v. 3).

(1) *gli italiani e il furto dei dati personali: la maggioranza sono ansiosi vulnerabili* (Gianni Rusconi, *Il Sole 24 Ore*, 24.11.2013)

(2) *una serie di questioni sono trattate in un unico documento* (Giacomo Galeazzi, *La Stampa*, 18.07.2010)

(3) *un paio di persone hanno anche vivacemente protestato* (Emilio Randacio, *La Repubblica*, 23.03.2012)

Il fenomeno, riconducibile alla generale preferenza del parlato per una modalità di strutturazione a base semantica piuttosto che sintattica, rientra in quel complesso di tratti già presenti in italiano antico, v. (4), e oggi caratteristici dell'italiano neo-standard (cfr. D'Achille 1990, 277ss.; Berruto 2012, 89).

(4) *la maggior parte de' giovani debbono essere, e sono, larghi e liberali e arditi e di grande isperanza* (Anonimo, *Reggimento de' principi di Egidio Romano*, 1288, L. 1, pt. 4, p. 109, Corpus OVI)

Meno frequente del precedente, e generalmente marcato come tratto del parlato e 'basso' in diafasia e diastratia, è l'assenza di accordo tra verbo e soggetto plurale postverbale, come in (5). In questi casi, concorrono a determinare il mancato accordo la posizione postverbale del soggetto (i cui effetti possono essere letti anche in termini di pianificazione del discorso; seguendo Durante 1981, 125: "un verbo in posizione iniziale non specifica in anticipo il numero e altresì il genere del nome che seguirà, sempreché tali dati non siano già noti: vale a dire la struttura della frase non ammette riferimenti a destra") e il suo carattere non topicale (v. oltre).

(5) *veniva su i bersaglieri* (Berretta 1994, 265)

Già in italiano antico "un soggetto postverbale di terza persona non provoca[va] necessariamente l'accordo con il verbo" (Salvi 2010, 557; v. anche D'Achille 1990, 277-312); ciò, in particolar modo, con i verbi inaccusativi (v. ad es. 6), la costruzione passiva e il *si* passivante (Salvi 2010, 557-558).

(6) *della cattiva e rea volontà, di cui nasce i sette vizî capitali overo mortali* (Bono Giamboni, *Trattato di virtù e di vizi*, 1292, cap. 3, p. 124, Corpus OVI)

Un caso specifico dell'assenza di accordo tra verbo e soggetto posposto è rappresentato dal tipo in (7), sul quale qui ci concentriamo; consiste cioè nella mancata concordanza tra una forma del tipo di *c'è* e un soggetto plurale postverbale. Fenomeno, questo, come i precedenti, che già compariva in italiano antico (v. ad es. 8).

(7) *c'è degli studenti che ti vengono a chiedere le cose dopo* (Cerruti 2009, 161)

(8) *tutti trovaro la porta, infra' quali ci fu molti nobili* (Anonimo, *Cronaca senese dall'anno 1202 al 1362*, 1362, 108.4; Corpus OVI)

Il sintagma nominale postverbale rappresenta evidentemente il soggetto della frase; non occupa tuttavia la posizione sintattica prevista per il soggetto dall'ordine non marcato dei costituenti (dell'italiano così come della larga maggioranza delle lingue del mondo, considerando frasi principali dichiarative). Inoltre, a differenza dei soggetti 'canonici', che sono topicali, ha carattere rematico; ovvero mostra le stesse proprietà del soggetto dei verbi inaccusativi, tipiche piuttosto dell'oggetto dei verbi transitivi (cfr. ad esempio Jezek 2010[1] e relativa bibliografia; sulle proprietà del soggetto postverbale, e del clitico *ci* in posizione soggettuale, v. anche Salvi 2001, 126; Salvi/Vanelli 2004, 65; La Fauci 2009, 140-41). La posizione postverbale e lo statuto non topicale del soggetto ne sfavoriscono quindi l'accordo con il verbo; è del resto diffusa interlinguisticamente la tendenza a non accordare il verbo con il soggetto quando quest'ultimo è poco 'canonico' (per una rassegna in prospettiva tipologica, e diacronica, delle proprietà dei soggetti non 'canonici' si veda ora Seržant, Kulikov 2013).

E' ovviamente speculare la tendenza di segno opposto, a marcare la relazione fra il verbo e un costituente che, a prescindere dalla funzione sintattica svolta, condivida le proprietà dei soggetti 'canonici'; segnalando tale relazione o sulla testa (è in questo

[1] "Semplificando, si può dire che, secondo la formulazione originaria dell'ipotesi inaccusativa, il soggetto dei verbi inaccusativi è un soggetto a livello superficiale, ma a livello sottostante è un oggetto" (Jezek 2010, 640).

senso, ad esempio, che si possono interpretare i casi più frequenti di *clitic doubling* nelle lingue romanze; cfr. Bossong 2003 e, per l'area italo-romanza, Berretta 1989; Cerruti 2008) o sulla dipendenza (si veda ad es. la marcatura differenziale dell'oggetto in casi quali *a me non mi vogliono*, in cui l'elemento dislocato, preverbale, è collocabile ai ranghi più alti delle gerarchie di animatezza e topicalità; cfr. tra gli altri Berretta 2002[1989]).

Spogliando la bibliografia sull'italiano, per una prima caratterizzazione del fenomeno in termini di marcatezza sociolinguistica, *c'è* con soggetto plurale risulta particolarmente presente nelle varietà regionali settentrionali (si vedano, tra gli altri, Grassi, Pautasso 1989, 157 per l'italiano regionale piemontese; Bianconi 1980, 142 per l'italiano regionale ticinese; Banfi 1997, 283 per l'italiano regionale lombardo), come tratto genericamente substandard (Berruto 2012, 89) o più specificamente come tratto di italiano popolare (cfr. Rovere 1977, 83; Telmon 1993, 128; v. anche Cerruti 2009, 161-162); nondimeno, il fenomeno è riscontrabile anche nell'italiano colloquiale di parlanti colti (cfr. Berretta 1994, 265)[2].

La tendenza a non concordare una forma del tipo di *c'è* con un soggetto plurale postverbale è presente, come si è già anticipato, in varie altre lingue; e in certi casi ha condotto allo sviluppo di una forma esistenziale invariabile, come il francese *il y a* (es. *il y a deux hommes qui parlent français*). Il fenomeno è stato analizzato particolarmente a fondo per l'inglese, dove troviamo enunciati come (9), incontrando speciale interesse da parte sia della linguistica generativa (v. ad es. Cornips/Corrigan 2005; Adger/Smith 2010) sia della sociolinguistica variazionista (v. ora Tagliamonte/Baayen 2012 e relativa bibliografia)[3].

(9) *there was four suitcases one inside each other* (Britain/Sudbury 2002, 218)

Proprio a partire da considerazioni relative all'inglese, si è discusso anche come il tipo *c'è* + soggetto plurale, quale caso specifico di *default singular* (Chambers 2004; v. sotto), riveli in lingue diverse tendenzialmente lo stesso grado e la stessa natura di marcatezza sociolinguistica; e, più specificamente, come possieda in lingue diverse uno stesso valore caratterizzante come marca di collocazione sociale bassa e/o di informalità. La questione assume particolare importanza per lo studio dei cosiddetti *vernacular universals* (per cui v. almeno Filppula *et alii* 2009; cfr. anche Trudgill 2011); con le parole di Chambers (2004, 132-133): "the nonstandard default singulars and the standard concord patterns are poles on a continuum. In between, there is a graded hierarchy in which concord occurs more frequently with certain types of subjects than others".

2. Analisi

Si inserisce in questo quadro l'analisi multivariata di cui si presentano i risultati (alcuni dei quali, come si è detto, sono esposti anche in Berruto/Cerruti in stampa).

---

[2] Pare invece più generalizzato il tipo *ce n'è* con soggetto plurale, anche preverbale, dove *ce n'è* "sembra quasi diventato una formula fissa invariabile" (Berruto 2012, 124).

[3] Oltre che da parte di altri paradigmi di ricerca; si vedano ad esempi gli studi di impostazione psicolinguistica di Bock/Kroch (1988) e Bock/Miller (1991).

L'analisi è stata condotta con Goldvarb X, un programma informatico che stima, a partire dalla distribuzione di frequenza di un certo fenomeno in un dato corpus, con quali probabilità fattori diversi – linguistici ed extralinguistici – incidano sul presentarsi di quel fenomeno. Il programma rimane uno dei più usati nella sociolinguistica quantitativa (cfr. Johnson 2009); va detto tuttavia che presenta problemi metodologici di non poco conto, superabili almeno in parte ricorrendo a software di analisi multivariata più recenti, quali R e Rbrul, il cui impiego negli studi di impostazione variazionista è non a caso crescente (ma su questi aspetti si veda più avanti).

Si sono esaminati tre insiemi di dati: due corpora di italiano parlato, il LIP (con esclusione dei dati di Firenze[4]) e il C-ORAL-ROM, e una raccolta di testi di italiano popolare, i materiali in Appendice a Foresti *et alii* (1983; dei quali si è operato uno spoglio manuale). In fase di codifica[5], questi tre insiemi sono stati siglati convenzionalmente con L, F e M, rispettivamente; da qui in avanti, così saranno indicati nel testo e nelle tabelle.

Goldvarb X, come i programmi di analisi multivariata in generale, analizza il rapporto di co-variazione che sussiste fra una variabile dipendente e più variabili indipendenti. La variabile dipendente è spesso a due valori, e consiste nella realizzazione o non realizzazione di un certo fenomeno (in altri termini, di una data variante di una variabile sociolinguistica). Le variabili indipendenti sono invece rappresentate dai fattori linguistici ed extralinguistici che possono influenzare la realizzazione di quel fenomeno.

Nel nostro caso, la variabile dipendente è l'accordo verbale nel costrutto di tipo *c'è*/*ci sono* con soggetto plurale postverbale. I casi senza accordo, del tipo di *c'è tante cose* (il tipo include tutti i possibili tempi verbali: *c'era*, *c'è stato*, *ci sarà*, ecc.), rappresentano le realizzazioni sub-standard; i casi con accordo, come *ci sono tante cose* (e *c'erano*, *ci sono stati*/*e*, *ci saranno*, ecc.), quelle standard. La variabile dipendente ha dunque due valori: non accordo (variante sub-standard) e accordo (variante standard); i casi di non accordo sono stati codificati con 1, quelli di accordo con 0.

Si sono considerate, poi, tre variabili indipendenti di natura linguistica e due variabili indipendenti di natura extralinguistica. Le variabili indipendenti di natura linguistica sono: (i) la struttura del sintagma nominale; (ii) l'adiacenza fra il verbo e il sintagma nominale; (iii) l'adiacenza fra il verbo e la testa del sintagma nominale.

I valori della variabile (i) tengono conto dei fattori seguenti:

- presenza della sola testa senza modificatori prenominali: nome (v. es. 9), codificato *t*; o pronome (10), codificato *o*;

  (9) *ma ridi lo stesso, non c'è problemi* (C-ORAL-ROM, conversazione in famiglia)

  (10) *in altri posti c'era quelli, dicevo* (Foresti *et alii* 1983, 33)

---

4 Si è presa questa decisione dopo aver constatato, in seguito a una prima esplorazione dell'intero corpus LIP, che le occorrenze di *c'è* + soggetto plurale si presentavano a Firenze in misura e con modalità parzialmente diverse rispetto alle altre tre città del corpus (Milano, Roma, Napoli).

5 Il programma 'legge' un *token file* costituito da stringhe di caratteri. Ogni stringa corrisponde a un'occorrenza della variabile (dipendente, si dirà più avanti) e riporta varie informazioni, ciascuna delle quali codificata con un certo carattere. Fra queste informazioni compare anche l'indicazione del corpus di riferimento, codificato appunto, nel nostro caso, con L, F o M (v. anche nota 8).

- presenza di un modificatore prenominale: partitivo (11), *p*; quantificatore (12), *q*; numerale (13), *n*; articolo (14), *d*; dimostrativo (15), *i*; aggettivo (16), *g*;

  (11) *a Palmanova c'era dei magazzini che rifornivano l'esercito* (Foresti *et alii* 1983, 56)

  (12) *dice sa c'è stato tanti fiorentini tanti anni fa* (C-ORAL-ROM, monologo in famiglia)

  (13) *c'era sessantamila austriaci che volevano sfondare* (Foresti *et alii* 1983, 69)

  (14) *se c'è le donne si viene* (C-ORAL-ROM, dialogo in pubblico)

  (15) *allora c'è quei pacchi lì colorati* (LIP, M.E.2.5.S)[6]

  (16) *con lui sì non c'è grandi problemi poi pagare paga bene* (LIP, M.B.8.81.A)

- combinazione di più modificatori prenominali (17), *c*;

  (17) *c'è tutte queste ville* (C-ORAL-ROM, monologo in famiglia)

- elenco o coordinazione di nomi o sintagmi nominali, v. (18) e (19), *e*.

  (18) *tra queste città c'è Bari Bologna Genova* (LIP, R.C.8.187.G)

  (19) *c'è il buono e il cattivo* (C-ORAL-ROM, dialogo in famiglia)

Le variabili (ii) e (iii) hanno ciascuna due valori; rispettivamente:

- adiacenza (come negli esempi da 9 a 19), *w*, e non adiacenza (20), *z*, tra verbo e sintagma nominale;

  (20) *c'era solo le corde per andar su* (Foresti *et alii* 1983, 34)

- adiacenza (come in 9, ad esempio), *x*, e non adiacenza (come in 13, fra gli altri), *y*, tra verbo e testa del sintagma nominale.

Per la scelta dei fattori linguistici, in particolare, si è fatto riferimento alle analisi multivariate condotte sulla presenza dello stesso fenomeno in altre lingue (e in particolare in inglese; per cui v. da ultimo Tagliamonte, Baayen 2012); alcuni fattori emersi come significativi in quelle indagini sono stati tuttavia esclusi dalla definizione delle variabili indipendenti appena presentate. La selezione è dipesa o dalla natura dei dati o da caratteristiche strutturali dell'italiano. Non si è considerata ad esempio la polarità dell'enunciato, che in varie lingue, quando positiva, mostra di favorire il mancato accordo[7] (v. 21), data la scarsa presenza di *c'è*/*ci sono* con polarità negativa nei nostri dati (pari al 2,5% dei contesti esaminati). Non si è tenuto conto nemmeno della distanza fra verbo e numerale, da ritenersi più significativa in lingue come

---

6 Gli esempi tratti dal LIP provengono dalla versione del corpus interrogabile all'indirizzo *http://badip.uni-graz.at*. I valori della stringa, qui ad esempio M.E.2.5.S, indicano, rispettivamente: la città di provenienza (Milano), il tipo di testo (E, cfr. *http://badip.uni-graz.at/it/tipologia-dei-testi*) e le coordinate necessarie per l'identificazione dell'enunciato e del parlante all'interno del corpus.

7 Benché, talvolta, possa incidere positivamente in alcune varietà e negativamente in altre varietà di una stessa lingua; cfr. ad es., per l'inglese, Anderwald (2002) e Tagliamonte (2009).

l'inglese (v. 22) che non in italiano, dove all'interno di un sintagma nominale si ha marcatura di numero su tutti i modificatori del nome.

(21) *there weren't always bulls, sometimes there was a few pigs, a few sheep* (Tagliamonte, Baayen 2012, 4)

(22) *there was about twelve different groups* (*ivi*, 5)

Si sono assunti invece come variabili indipendenti di natura extralinguistica gli stessi corpora di riferimento e la variabilità diafasica. I tre corpora indagati possono difatti essere intesi come rappresentativi ciascuno di una varietà di lingua: l'italiano parlato (non marcato in diatopia), L, l'italiano parlato a Firenze, F, e l'italiano popolare, M. Quanto alla variabilità diafasica, per il corpus L si sono ritenuti formali i testi indicati in De Mauro *et alii* (1993: 40-41) come di tipo C e D, e informali quelli di tipo A, B ed E; per il *corpus* F, si è mantenuta la distinzione fra testi formali e informali fornita in Cresti/Moneglia (2005: 10-11); per il corpus M, invece, si sono considerati separatamente testi scritti e testi parlati. In fase di codifica, queste tre differenziazioni sono state ricondotte a un'unica opposizione, quella fra 'alto' (codifica: A) e 'basso' (B) in diafasia (intendendo questa, per lo meno operativamente, come inclusiva della dimensione diamesica)[8].

Una prima sintesi di alcuni risultati dell'analisi è riportata in Tab. 1, nella quale si dà conto della distribuzione dei valori della variabile dipendente rispetto alle varietà di lingua considerate e all'opposizione tra 'alto' e 'basso' in diafasia. La tabella fornisce già qualche indicazione circa la diversa sensibilità della variabile a fattori extralinguistici.

		Variabile dipendente		
		1	0	*totale*
Corpora	M	75 (49.7)	76 (50.3)	151 (21.9)
	F	51 (19.6)	209 (80.4)	260 (37.7)
	L	22 (7.9)	257 (92.1)	279 (40.4)
Variabilità diafasica	B	126 (28.6)	315 (71.4)	441 (63.9)
	A	22 (8.8)	227 (91.2)	249 (36.1)
	totale	148 (21.4)	542 (78.6)	690 (100.0)

Tabella 1: Occorrenze della variabile dipendente rispetto a fattori extralinguistici

Come da attese (cfr. § 1.1.), l'assenza di accordo nel costrutto di tipo *c'è* + soggetto plurale compare prevalentemente nell'italiano popolare, dove occorre in poco meno del 50% dei casi possibili, e in produzioni linguistiche caratterizzabili come 'basse' in diafasia, nelle quali raggiunge un valore percentuale del 28,6% (di contro a un 8,8% in produzioni 'alte'). Non è invece conforme alle attese la particolare diffusione del tratto

[8] Ciascuna occorrenza della variabile dipendente è stata dunque codificata con una stringa composta da sei caratteri: il primo di questi indicava il valore assunto dalla variabile dipendente nel contesto specifico (1 nei casi di non accordo, 0 in quelli di accordo) e i successivi cinque i valori di ciascuna delle variabili indipendenti in quel dato contesto; ad esempio, un enunciato come *dice sa c'è stato tanti fiorentini tanti anni fa*, v. (12) sopra, è stato codificato come *1FBwyq* (per il valore dei caratteri, scelti convenzionalmente, si vedano gli esempi (9)-(20)). Per indicazioni metodologiche e di procedura si rimanda comunque a lavori come Tagliamonte (2006) e, in ambito italiano, Vietti (2005).

nell'italiano di Firenze, specie se paragonata alla sua attestazione nel corpus panitaliano di parlato (19,6% di occorrenze nel primo, 7,9% nel secondo); il che induce ad ascrivere al fenomeno una marcatezza diatopica di natura diversa da quella tradizionalmente attribuitagli in letteratura (v. di nuovo § 1.1.). Nel complesso, in ogni modo, prevalgono sensibilmente le realizzazioni standard, ovvero con accordo verbale (presenti nel 78,6% dei casi).

A partire dalla distribuzione di frequenza della variabile, anche rispetto a fattori linguistici (dei quali però si dà conto soltanto in Appendice, Tab. I, dove si presenta il quadro completo dei dati), il programma ha poi calcolato, attraverso il metodo statistico della regressione logistica, le diverse probabilità con le quali fattori diversi possono influenzare positivamente o negativamente il mancato accordo.

I diversi valori di probabilità – ossia i diversi 'pesi' – assegnati ai fattori, linguistici ed extralinguistici, sono riportati in Tab. 2. La Tabella restituisce anche i valori di *Input* e *Log likelihood*. Il primo rappresenta la probabilità che il fenomeno ha di realizzarsi a prescindere dalle variabili indipendenti considerate; nel nostro caso, dunque, notiamo subito che il fenomeno ha una bassa probabilità generale di realizzazione (0,146; considerando che un evento certo ha probabilità 1 e un evento impossibile probabilità 0). Il secondo, che è sempre un numero negativo, permette di valutare quanto la configurazione di variabili in questione sia adatta a descrivere la realizzazione variabile del fenomeno; più il numero è vicino allo zero e più tale configurazione è adeguata (si noterà più avanti come l'adeguatezza del modello migliori con analisi condotte separatamente sui diversi corpora).

Input 0.146		Corpora	
Log likelihood -275.367		M	0.837
Totale contesti 690		F	0.499
Realizzazioni sub-standard 148 (21.4)		L	0.293
		Range	*544*
Struttura del SN		Variabilità diafasica	
e	0.876	B	0.621
q	0.643	A	0.294
n	0.611	*Range*	*327*
i	0.562	Adiacenza V SN	
t	0.562	z	0.706
o	0.509	w	0.473
d	0.469	*Range*	*233*
g	0.386	Adiacenza V testa di SN	
c	0.395	x	[0.537]
p	0.295	y	[0.494]
Range	*581*	*Range*	*43*

Tabella 2: Esiti dell'analisi multivariata

Osservando i pesi dei diversi fattori è possibile stabilire quali di questi influenzino positivamente e quali negativamente il mancato accordo; ovvero, quali aumentino e

quali diminuiscano la probabilità generale di realizzazione del fenomeno (che come si è detto è di per sé bassa)[9].

Nel nostro caso, le due variabili indipendenti più significative sono l'una di natura linguistica, la struttura del sintagma nominale, l'altra di natura extralinguistica, i corpora di riferimento. Tra i fattori di cui tiene conto la prima, il più influente in positivo è l'elenco o la coordinazione di nomi o sintagmi nominali (*g*, v. (18) e (19) sopra), il più influente in negativo è la presenza di un partitivo (*p*, v. (11) sopra); del primo fattore può dar ragione la forma singolare di ciascuno degli elementi nominali in elenco o coordinati, del secondo, per lo meno in alcuni casi, la funzione del partitivo come forma plurale dell'articolo indeterminativo[10]. Quanto ai corpora, M mostra un effetto positivo sulla realizzazione del fenomeno, mentre L negativo[11]. Hanno poi effetto negativo il carattere formale della situazione comunicativa e positivo la non adiacenza tra verbo e sintagma nominale (*z*, v. es. (20) sopra).

Non risulta invece statisticamente significativa l'incidenza della variabile dell'adiacenza tra il verbo e la testa del sintagma nominale (per questo motivo, in Tab. 2, i pesi dei due fattori relativi sono riportati tra parentesi quadre); per contro, tale variabile assume rilevanza in lingue quali l'inglese, dove la testa può essere l'unico elemento del sintagma nominale che rechi una marca di numero esplicita, come in (23).

(23) *he showed how there was different local scientific rationalities within different fields* (Kristian Kobbenes Starheim, *Knowledge production in experimental molecular medicine*, University of Bergen, 2011, https://bora.uib.no/bitstream/handle/1956/5330/87783711.pdf)

A mo' di riepilogo, si possono rappresentare con i due enunciati forniti qui di seguito i casi 'prototipici' di realizzazione della variante sub-standard, l'uno (24), e della variante standard, l'altro (25); il primo esemplifica l'assenza di accordo nell'italiano popolare parlato, in contesti con coordinazione di elementi singolari e adiacenza tra verbo e sintagma nominale; il secondo la presenza di accordo nell'italiano parlato formale, con un sintagma nominale non adiacente al verbo e contenente un partitivo.

(24) *poi c'era il tenente e il capitano* (Foresti *et alii* 1983, 80)

(25) *ci sono comunque delle notizie che tu devi tagliare* (LIP, N.D.2.37.A)

L'analisi multivariata è stata condotta poi separatamente sui diversi corpora. I principali risultati sono esposti in Tab. 3 (per il quadro completo dei dati, invece, si

---

9 Più il peso si avvicina a 1, più il fattore corrispondente ha un effetto positivo sulla realizzazione del fenomeno; più si avvicina a 0, più ha un effetto negativo. L'influenza di fattori che mostrino pesi intorno allo 0,5 è invece da ritenersi scarsamente significativa. I valori di *Range* consistono nella differenza fra il peso più alto e quello più basso per ogni variabile indipendente; ha *Range* più alto la variabile complessivamente più influente nella realizzazione del fenomeno, *Range* più basso quella meno influente (che può essere non significativa).

10 Benché, va detto, l'elemento nominale modificato dal partitivo non condivida un tratto caratteristico dei soggetti 'canonici': la definitezza; il che, come accennato al § 1.1, potrebbe rappresentare una condizione sfavorevole per l'accordo.

11 Il che significa, ad esempio, considerando ciascun corpus come rappresentativo di una data varietà di lingua, che i fattori extralinguistici in dipendenza dai quali è più probabile che si verifichi l'assenza di accordo corrispondono a quelli che definiscono sul versante sociale l'italiano popolare (livello di istruzione basso, acquisizione imperfetta dell'italiano, ecc.).

rimanda ancora all'Appendice, Tabelle II, III e IV). Si può notare innanzitutto come la probabilità generale di non avere accordo, a prescindere dunque dall'effetto di variabili indipendenti, differisca a seconda della varietà di lingua considerata. Più in particolare, e in altri termini, la regola dell'assenza di accordo è sensibilmente più forte nell'italiano popolare (*Input* 0.496) che nell'italiano di Firenze (*Input* 0.143) o nell'italiano parlato in genere (*Input* 0.015); la probabilità generale di realizzazione di un fenomeno si può intendere infatti come "an overall indication of the strength of the rule" (Young/Bayley 1996, 270). Come anticipato sopra, poi, l'adeguatezza del modello migliora considerando separatamente i diversi corpora (si confrontino i valori di *Log likelihood* nelle Tabelle 2 e 3).

Il confronto fra i tre corpora consente inoltre di mettere in evidenza come l'incidenza statistica delle variabili indipendenti differisca da varietà a varietà. Si può osservare ad esempio come il mancato accordo sia influenzato significativamente dall'adiacenza tra il verbo e il sintagma nominale soltanto nel corpus panitaliano di parlato (dove peraltro ha un *Range* di 501); dunque, non nella varietà fiorentina né nell'italiano popolare. In L, in particolare, la posizione non adiacente di verbo e sintagma nominale (*z*, peso 0,964) aumenta quasi categoricamente la probabilità generale di realizzazione del fenomeno.

Allo stesso tempo, variabili comunque significative esercitano un'influenza differente a seconda della varietà di lingua. Confrontando L con F, emerge come la struttura del sintagma nominale e la variabilità diafasica siano più influenti nel primo corpus che nel secondo. Più nel parlato in genere che nella varietà fiorentina, insomma, l'accordo verbale nel costrutto *c'è* + soggetto plurale è sensibile alla composizione interna del sintagma nominale (in L la probabilità generale di non avere accordo aumenta quasi categoricamente con nominali in elenco o coordinati: *e*, peso 0,992; diminuisce quasi categoricamente con partitivi: *p*, peso 0,163) e all'opposizione tra formalità (peso 0,198) e informalità (peso 0,844).

	Corpora					
	L *Input* 0.015, *Log likelihood* -47.496		F *Input* 0.143, *Log likelihood* -109.114		M *Input* 0.496, *Log likelihood* -101.627	
Struttura del SN	e	0.992	e	0.870	e	[0.663]
	q	0.759	n	0.701	q	[0.633]
	n	0.757	q	0.630	d	[0.538]
	t	0.702	o	0.590	n	[0.510]
	i	0.570	i	0.558	t	[0.504]
	g	0.523	t	0.547	c	[0.447]
	o	0.520	c	0.547	o	[0.429]
	c	0.397	d	0.460	i	[0.413]
	d	0.291	g	0.387	p	[0.401]
	p	0.163	p	0.215	g	[0.275]
Range		*829*		*655*		*220*
Variabilità diafasica	B	0.844	B	0.650	B	[0.516]
	A	0.108	A	0.233	A	[0.449]
Range		*736*		*417*		*67*
Adiacenza V SN	z	0.964	z	[0.587]	z	[0.608]
	w	0.413	w	[0.486]	w	[0.490]
Range		*551*		*101*		*118*

Adiacenza V testa di SN	x	[0.585]	x	[0.519]	x	[0.519]
	y	[0.481]	y	[0.496]	y	[0.499]
Range		*104*		*23*		*20*

Tabella 3: Analisi multivariata per corpus

Guardando i dati relativi a M, invece, è da notare come in italiano popolare il mancato accordo di *c'è* non risulti correlare in modo significativo con alcun fattore, né linguistico né extralinguistico. La regressione logistica restituisce infatti come configurazione più adatta a descrivere il presentarsi del fenomeno quella che esclude effetti da parte delle variabili indipendenti; nessun fattore, cioè, si mostra in grado di variare la probabilità generale di realizzazione del fenomeno. Ovviamente, del resto, più un fenomeno è frequente più tende a realizzarsi indipendentemente da ogni fattore[12].

Nondimeno, occorre considerare che i materiali del corpus M presentano una variazione interindividuale piuttosto forte quanto ad assenza o presenza di accordo. In Fig. 1 si forniscono le percentuali di realizzazione della variante standard e di quella sub-standard per ogni informatore del corpus[13] (ciascuno dei quali identificato da una sigla sull'asse orizzontale); per alcuni parlanti (5 su 13) è categorica l'assenza di accordo, per altri (3 su 13) la presenza.

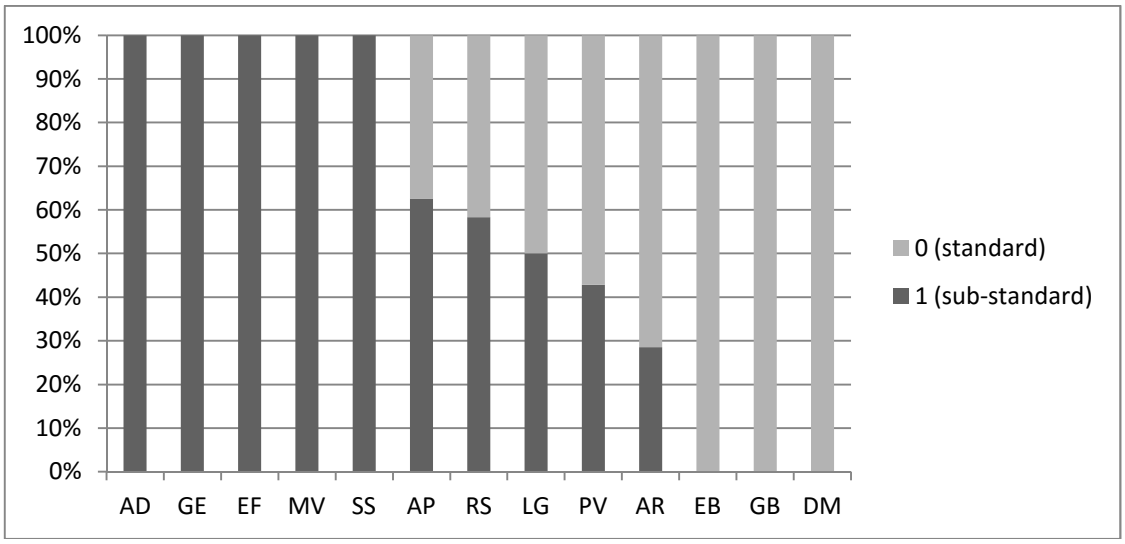

Figura 1: La variazione interindividuale nel corpus di italiano popolare

Ancorché un dato come questo possa ritenersi coerente con l'affermazione che l'assenza (o la presenza) di accordo sia sostanzialmente indipendente da fattori

12 Differenze in termini di significatività delle variabili indipendenti e di ordinamento gerarchico dei fattori possono correlare con fenomeni di contatto linguistico: nel nostro caso, possibilmente, con l'azione interferenziale del dialetto (specie nell'italiano popolare; in vari dialetti italoromanzi, infatti, la forma locativo-esistenziale-presentativa non concorda con il soggetto plurale postverbale). Cfr. anche Berruto/Cerruti (in stampa); la questione è affrontata nei suoi aspetti generali in Meyerhoff (2013).

13 Prendendo a riferimento le sole narrazioni orali; le testimonianze scritte raccolte in Foresti *et alii* (1983), sebbene relative a un numero più ampio di parlanti, sono difatti troppo brevi (spesso dell'ordine di qualche decina di righe per individuo) per poter essere tenute presenti nel computo.

linguistici ed extralinguistici, non è da escludere che proprio l'intervento della variazione interindividuale sia in qualche modo in grado di turbare la stima di significatività delle variabili indipendenti. Ciò, è chiaro, interessa potenzialmente gli esiti dell'analisi per quanto riguarda non soltanto M ma tutti e tre i corpora, e rappresenta il problema principale di un programma quale Goldvarb (così come del suo predecessore Varbrul), che non riesce a valutare i possibili effetti della variazione interindividuale.

Software di concezione più recente, come Rbrul, tengono conto invece di due tipi di effetti, denominati *fixed effects* e *random effects*. Tra i primi rientrano gli effetti prodotti da variabili indipendenti quali, nel nostro caso, struttura del sintagma nominale, adiacenza fra verbo e sintagma nominale, variabilità diafasica ecc., definite ciascuna da una serie specifica di fattori; tra i secondi, appunto, gli effetti causati dalla variazione interindividuale. Questa rappresenta non una variabile alla stregua delle precedenti, determinata da un insieme di fattori[14], ma un parametro in grado di condizionare la significatività delle variabili indipendenti. Stimarne gli effetti significa tenere conto che "some individuals might favor a linguistic outcome while others might disfavor it, over and above (or 'under and below') what their gender, age, social class, etc. would predict" (Johnson 2009, 365); in altre parole, che la significatività, o al contrario la non significatività, di una variabile indipendente possa essere in realtà dovuta al caso[15] (v. anche Gorman, Johnson 2013).

2. 1 Uno sguardo alla comunicazione mediata dal computer

Considerando la natura sub-standard e, parrebbe, specialmente 'parlata' del tipo *c'è* + soggetto plurale, può avere qualche interesse verificare l'occorrenza del fenomeno nelle produzioni formulate in un *medium* come Internet, evidentemente scritto ma caratterizzato spesso da tratti tipici del parlato. Può essere rilevante, in particolare, verificare se e come la presenza del fenomeno correli diversamente con ambiti di comunicazione mediata dal computer (da qui in avanti, CMC) differenti, distinti fra di loro proprio sul piano della concezione (*Konzeption*, nel senso di Söll 1980); ovvero, sulla base della reciproca caratterizzazione strutturale come testi più tipicamente scritti o più tipicamente parlati. È noto infatti che i vari ambiti di CMC (chat, newsgroup, forum, e-mail, blog, siti web ecc.) rappresentano una vasta gamma di tipi di testo, le cui caratteristiche linguistiche rimandano a varietà di lingua collocabili in posizioni diverse lungo il continuum scritto/parlato (oltre che formale/informale; per cui v. ad es. Cerruti, Corino, Onesti 2011).

Da una rapida ricerca condotta con Google, appoggiata a una prima interrogazione del corpus di pagine web in italiano *itTenTen10* (3.076.908.415 *tokens*, effettuata con il

---

14 Oltretutto non è sempre possibile, come in Fig. 1, estrarre da un corpus le occorrenze della variabile dipendente per ogni singolo individuo.

15 Con ciò, tuttavia, non si deve intendere che la stima di significatività delle variabili indipendenti effettuata con programmi come Rbrul sia generalmente al riparo da possibili errori: "If there is a lot of individual variation, chance can create the appearance of external effects, and Rbrul raises its standards accordingly. Misidentifying a chance effect as a real one is called a Type I error [...]. However, this […] behavior has a trade-off: in some situations, Rbrul is more likely than GoldVarb to make a Type II error by failing to identify an effect that really does exist" (Johnson 2009, 365).

programma *Sketch Engine*), pare di notare una particolare concentrazione di casi di non concordanza tra *c'è* e soggetto plurale in ambiti di CMC caratterizzati da comunicazione dialogica e sincrona, quali le chat; ambiti, cioè, in cui la comunicazione riproduce più da vicino il dialogo parlato. (Questa semplice impressione richiede tuttavia di essere verificata con un'analisi quantitativa mirata, che qui non si è condotta).

Nondimeno, non mancano occorrenze del fenomeno neppure in ambiti quali newsgroup e forum, in cui la comunicazione è dialogica ma asincrona. Se ne riportano alcuni esempi al punto (26), tratti da discussioni in rete, rappresentativi tra l'altro delle varie strutture del sintagma nominale di cui ha tenuto conto l'analisi multivariata.

(26) i) *nella coppia etero c'è un uomo e una donna nella coppia omo ci sono due uomini*; ii) *c'è tanti ragazzi che sono veramente da buttare!*; iii) *ovviamente c'è mille metodi per fare soldi su nos*; iv) *c'è quei piccoli problemi che ti ho scritto*; v) *voglio dire, c'è dell'horror, c'è cose italiane ecc.*; vi) *magari c'è quelli che apprezzano*; vii) *sì che c'è le donne che non si lamentano, ma non va bene*; viii) *anche da ragazzino si giocava nel fiume se c'era belle giornate*; ix) *c'era tutti quei balli favolosi, come il Charleston*; x) *c'è delle volte che non riesco ad arrivare*

Per una prima indicazione di natura quantitativa, si può sondare l'assenza di accordo in un corpus di messaggi di newsgroup, NUNC (Newsgroup *UseNet Corpus*, 74.695.600 tokens; per cui v. ad es. Barbera *et alii* 2007). In Tab. 4 si pongono a confronto i valori di assenza (1) e presenza (0) di accordo relativi ai tre fattori linguistici che, stando ai risultati dell'analisi multivariata, mostrano di favorire maggiormente la comparsa del tipo *c'è* + soggetto plurale (cfr. Tab. 2, § 2.); fra i quantificatori, come modificatori prenominali, si sono considerati soltanto i tre più frequenti nell'italiano parlato (*tanto, molto, alcuno*, v. LIP[16]).

		Variabile dipendente		
		1	0	*totale*
Struttura del SN	elenco o coordinazione di SN	37 (15.2)	207 (84.8)	244 (12.0)
	quantificatori *tanti*/*e*, *molti*/*e*, *alcuni*/*e*	4 (0.4)	1021 (99,6)	1025 (50.5)
	numerali	1 (0.1)	759 (99.9)	760 (37.5)
	totale	42 (2.1)	1987 (97.9)	2029 (100.0)

Tabella 4: Occorrenze della variabile dipendente nel corpus NUNC

Se raffrontiamo questi dati con quelli riportati in Appendice, Tab. I, non si può non registrare come, nonostante il numero dei contesti esaminati in NUNC (2029) sia largamente superiore a quello dei contesti considerati nell'analisi multivariata (210, quindi più di dieci volte questo), i valori di assenza di accordo nel primo caso (42, pari al 2,1% dei casi possibili) siano sensibilmente inferiori a quelli riscontrati nel secondo (65, pari al 30,9%[17]; con una differenza statisticamente significativa: chi^2=260,40 e p<0,001).

16 Escludendo i quantificatori per i quali non si dia accordo al plurale (es. *qualche*).

17 Considerando oltretutto che i tipi di contesto esaminati in NUNC rappresentano di fatto un sottoinsieme di quelli esaminati nell'analisi multivariata (che non era ristretta ai tre quantificatori più frequenti).

Un'ipotesi di lavoro, da verificare con indagini mirate, può essere dunque che la scrittura (ovvero il *medium* grafico) incida negativamente sulla realizzazione del fenomeno non soltanto negli ambiti di CMC più distanti, in termini di concezione strutturale, dal dialogo parlato, come quelli contrassegnati da comunicazione monologica e asincrona (es. blog e siti web), ma anche in ambiti meno distanti da questo, come quelli in cui la comunicazione è dialogica ancorché asincrona (es. newsgroup, forum). Di contro, la stessa scrittura parrebbe sfavorire in misura minore la realizzazione del fenomeno negli ambiti in cui la comunicazione riproduce più direttamente il dialogo parlato, ovvero dov'è dialogica e sincrona (es. chat). Il che confermerebbe la natura spiccatamente 'parlata' dell'assenza di accordo nel costrutto con *c'è*.

Più specificamente, il presentarsi del fenomeno sembrerebbe particolarmente sensibile all'opposizione tra comunicazione sincrona e asincrona; opposizione alla quale è notoriamente correlata l'esistenza di tempi di pianificazione diversi. È noto d'altra parte come alcuni tratti del parlato abbiano una presenza soltanto sporadica o marginale nella CMC. Si tratta, in particolare, di quei fenomeni riconducibili a fatti di mancata pianificazione del parlato spontaneo (cfr. Berruto 2005); fatti, come si è visto (§ 1.1), che possono entrare in gioco anche nel caso di *c'è* + soggetto plurale.

3. Conclusione

L'indagine mostra come la frequenza, e dunque la probabilità, di realizzazione del fenomeno differisca significativamente in dipendenza sia da contesti strutturali diversi sia da varietà di lingua diverse. Si osserva in particolare come alcune strutture del sintagma nominale incidano positivamente sull'assenza di accordo tra soggetto e verbo (fra tutte, l'elenco o la coordinazione tra nomi o sintagmi nominali), altre negativamente (tra queste, la presenza di un partitivo); e come la distanza fra soggetto e verbo sia favorevole alla non concordanza tra i due costituenti. (Altri fattori linguistici, quale l'adiacenza tra il verbo e la testa del sintagma nominale, risultano invece più pertinenti in lingue con altre caratteristiche strutturali). Quanto a fattori di natura extra-linguistica, se, da una parte, la prevalenza del mancato accordo nelle varietà popolari e informali non fa che confermarne il valore di marca di collocazione sociale bassa e di informalità, dall'altra la sua particolare diffusione nella varietà fiorentina permette di caratterizzare il tratto in termini nuovi quanto a marcatezza diatopica. Seppure cursorie e impressionistiche, poi, le osservazioni relative ad alcuni ambiti di comunicazione mediata dal computer sottolineano il carattere eminentemente 'parlato' del fenomeno; la scrittura parrebbe infatti sfavorirne la presenza anche in ambiti non distanti, quanto a concezione strutturale, dal dialogo parlato.

Bibliografia

Adger David; Smith Jennifer (2010): *Variation in agreement: A lexical feature-based approach*. In "Lingua" 120, pp. 1109-1134.

Anderwald, Lieselotte (2002): *Negation in non-standard British English: gaps, regularizations, and asymmetries*. London, Routledge.

Banfi, Emanuele (1997): *Le tante 'identità' lombarde*. In "Italiano e oltre" 5, pp. 280-286.

Barbera, Manuel; Corino, Elisa; Onesti, Cristina (a cura di) (2007): *Corpora e linguistica in rete*. Perugia, Guerra.

Berretta, Monica (1989): *Tracce di coniugazione oggettiva in italiano*. In Fabio Foresti; Elena Rizzi: Paola Benedini (a cura di), *L'italiano tra le lingue romanze*. SLI 20. Roma, Bulzoni, pp. 125-150.

Berretta, Monica (1994): *Il parlato italiano contemporaneo*. In L. Serianni; P. Trifone (a cura di), *Storia della lingua italiana*. Vol. II. Torino, Einaudi, pp. 239-270.

Berretta, Monica (2002[1989]): *Sulla presenza dell'accusativo preposizionale in italiano settentrionale: note tipologiche*. In Silvia Dal Negro; Bice Mortara Garavelli (a cura di), *Temi e percorsi della linguistica. Scritti scelti di Monica Berretta*. Vercelli, Mercurio, pp. 347-378.

Berruto, Gaetano (2005): *Italiano parlato e comunicazione mediata dal computer*. In Klaus Hölker; Christiane Maaß (Hrsg.), *Aspetti dell'italiano parlato. Tra lingua nazionale e varietà regionali*. Münster, LIT, pp. 109-124.

Berruto, Gaetano (2012): *Sociolinguistica dell'italiano contemporaneo. Nuova edizione*. Roma, Carocci.

Berruto, Gaetano; Cerruti, Massimo (in stampa): *Un esercizio di analisi variazionista: l'accordo verbale nel costrutto locativo-esistenziale-presentativo*. In Maria Grazia Busà; Sara Gesuato (a cura di), *Studi in onore di Alberto Mioni*. Padova, CLEUP.

Bianconi, Sandro (1980): *Lingua matrigna. Italiano e dialetto nella Svizzera italiana*. Bologna, Il Mulino.

Bock, K. J.; Kroch, A. S. (1988): *The isolability of syntactic processing*. In G. N. Carlson; M. K. Tannenhaus (eds.), *The Isolability of Syntactic Processing. Linguistic structure in language processing*. Dordrecht, Kluwer, pp. 157-196.

Bock, Karthryn; Miller, Carol A. (1991): *Broken agreement*. In "Cognitive Psychology" 23, pp. 45-93.

Bossong, Georg (2003): *Nominal and/or verbal marking of central actants*. In Giuliana Fiorentino (ed.), *Romance Objects. Transitivity in Romance Languages*. Berlin-New York, Mouton de Gruyter, pp. 17-47.

Britain, David; Sudbury, Andrea (2002): *There's sheep and there's penguins; convergence, 'drift' and 'slant' in New Zealand and Falkland Island English*. In Mari C. Jones; Edith Esch (eds.), *Language change. The interplay of internal, external and extra-linguistic factors*. Berlin-New York, Mouton de Gruyter, pp. 211-240.

Cerruti, Massimo (2008): *Condizioni e indizi di coniugazione oggettiva: i dialetti italiani settentrionali tra le lingue romanze*. In "Rivista Italiana di Dialettologia" 32, pp. 13-38.

Cerruti, Massimo (2009): *Strutture dell'italiano regionale. Morfosintassi di una varietà diatopica in prospettiva sociolinguistica.* Frankfurt am Main, Lang.
Cerruti, Massimo; Corino, Elisa; Onesti, Cristina (a cura di) (2011): *Formale e informale. La variazione di registro nella comunicazione elettronica*. Roma, Carocci.
Chambers, Jack K. (2004): *Dynamic typology and vernacular universals*. In Bernd Kortmann (ed.), *Dialectology Meets Typology: Dialect Grammar from a Cross-Linguistic Perspective*. Berlin-New York, Mouton de Gruyter, pp. 128-145.
Cornips, Leonie; Corrigan, Karen P. (eds.) (2005): *Syntax and Variation. Reconciling the Biological and the Social*. Amsterdam-Philadelphia, John Benjamins.
Cresti, Emanuela; Moneglia, Massimo (2005): *C-Oral-Rom. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam-Philadelphia, John Benjamins.
D'Achille, Paolo (1990): *Sintassi del parlato e tradizione scritta della lingua italiana. Analisi di testi dalle origini al secolo XVIII*. Roma, Bonacci.
De Mauro, Tullio; Mancini, Federico; Vedovelli, Massimo; Voghera, Miriam (1993): *Lessico di frequenza dell'italiano parlato*. Milano, Etaslibri.
Durante, Marcello (1981): *Dal latino all'italiano moderno. Saggio di storia linguistica e culturale*. Bologna, Zanichelli.
Filppula, Markku; Klemola, Juhani; Paulasto, Heli (eds.) (2009): *Vernacular Universals and Language Contacts. Evidence from Varieties of English and Beyond*. London, Routledge.
Foresti, Fabio; Morisi, Paola; Resca, Maria (a cura di) (1983): *Era come a mietere. Testimonianze orali e scritte di soldati sulla Grande guerra con immagini inedite*. Comune di San Giovanni in Persiceto.
Gorman, Kyle; Johnson, Daniel E. (2013): *Quantitative analysis*. In Robert Bayley; Richard Cameron; Ceil Lucas (eds.), *The Oxford Handbook of Sociolinguistics*. Oxford, Oxford University Press, pp. 214-240.
Grassi, Corrado; Pautasso, Mariella (1989): *Prima roba il parlare… Lingue e dialetti dell'emigrazione biellese*. Milano, Electa.
Jezek, Elisabetta (2010), *Verbi Inaccusativi*. In Raffaele Simone (dir.), *Enciclopedia dell'italiano*. Roma, Istituto della Enciclopedia italiana Giovanni Treccani, pp. 640-641.
Johnson, Daniel E. (2009): *Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis*. In "Language and Linguistics Compass" 3/1, pp. 359-383.
La Fauci, Nunzio (2009): *Compendio di sintassi italiana*. Bologna, il Mulino.
Meyerhoff, Miriam (2013): *Syntactic variation and change: The variationist framework and language contact*. In Isabelle Léglise; Claudine Chamoreau (eds.), *The Interplay of Variation and Change in Contact Settings*. Amsterdam-Philadelphia, Benjamins, pp. 23-51.
Rovere, Giovanni (1977): *Testi di italiano popolare. Autobiografie di lavoratori e figli di lavoratori emigrati*. Roma, Centro Studi Emigrazione.
Salvi, Giampaolo (2001): *La frase semplice*. In Lorenzo Renzi; Giampaolo Salvi; Anna Cardinaletti (a cura di), *Grande grammatica italiana di consultazione*. Vol. I. Bologna, Il Mulino, pp. 37-127.
Salvi, Giampaolo (2010): *L'accordo*. In Giampaolo Salvi; Lorenzo Renzi (a cura di), *Grammatica dell'italiano antico*. Vol. I. Bologna, Il Mulino, pp. 547-568.

Salvi, Giampaolo; Vanelli, Laura (2004): *Nuova grammatica italiana*. Bologna, il Mulino.

Seržant, Ilja A.; Kulikov, Leonid (eds.) (2013): *The Diachronic Typology of Non-Canonical Subjects*. Amsterdam, John Benjamins.

Söll, Ludwig (1980): *Gesprochenes und geschriebenes Französisch*. Berlin, Schmidt.

Tagliamonte, Sali A. (2006): *Analysing linguistic variation*. Cambridge, Cambridge University Press.

Tagliamonte, Sali A. (2009): *There Was Universals, Then There Weren't: A Comparative Sociolinguistic Perspective on 'Default Singulars'*. In Filppula, Markku; Klemola, Juhani; Paulasto, Heli (eds.), *Vernacular Universals and Language Contacts. Evidence from Varieties of English and Beyond*. London, Routledge, pp. 103-129.

Tagliamonte, Sali A.; Baayen, R. Harald (2012): *Models, forests and trees of York English:* Was/were *variation as a case study for statistical practice*. In "Language Variation and Change" 24/2, pp. 135-178.

Telmon, Tullio (1993): *Varietà regionali*. In Alberto A. Sobrero (a cura di), *Introduzione all'italiano contemporaneo*. Vol. II. Roma-Bari, Laterza, pp. 93-149.

Trudgill, Peter (2011): *Sociolinguistic typology. Social determinants of linguistic complexities*. Oxford, Oxford University Press.

Vietti, Alessandro (2005): *Approcci quantitativi all'analisi della variazione linguistica: il caso di GOLDVARB 2001*. In "Linguistica e filologia" 20, pp. 31-69.

Young, Richard; Bayley, R. (1996): *VARBRUL analysis for second language acquisition research*. In Bayley, Robert; Preston, Dennis (eds.), *Second language acquisition and linguistic variation*. Amsterdam-Philadelphia, Benjamins, pp. 253-306.

Appendice

Input 0.146
Log likelihood -275.367

	Peso	1	0	*Totale*
Corpora (Varietà di lingua)				
M	0.837	75 (49.7)	76 (50.3)	151 (21.9)
F	0.499	51 (19.6)	209 (80.4)	260 (37.7)
L	0.293	22 (7.9)	257 (92.1)	279 (40.4)
Range	*544*			
Contesti linguistici				
e	0.876	12 (63.2)	7 (36.8)	19 (2.8)
q	0.643	34 (27.0)	92 (73.0)	126 (18.3)
n	0.611	19 (29.2)	46 (70.8)	65 (9.4)
i	0.562	8 (26.7)	22 (73.3)	30 (4.3)
t	0.562	16 (16.0)	84 (84.0)	100 (14.5)
o	0.509	5 (27.8)	13 (72.2)	18 (2.6)
d	0.469	29 (24.4)	90 (75.6)	119 (17.2)
g	0.386	4 (13.3)	26 (86.7)	30 (4.3)
c	0.395	3 (9.1)	30 (90.9)	33 (4.8)
p	0.295	18 (12.0)	132 (88.0)	150 (21.7)
Range	*581*			
Variazione diafasica				
B	0.621	126 (28.6)	315 (71.4)	441 (63.9)
A	0.294	22 (8.8)	227 (91.2)	249 (36.1)
Range	*327*			
Adiacenza V SN				
z (no)	0.706	24 (32.0)	51 (68.0)	75 (10.9)
w (sì)	0.473	124 (20.2)	491 (79.8)	615 (89.1)
Range	*233*			
Adiacenza V Testa				
x (sì)	[0.537]	16 (16.0)	84 (84.0)	100 (14.5)
y (no)	[0.494]	132 (22.4)	458 (77.6)	590 (85.5)
	43			

Tabella I: Analisi multivariata: corpora L, F, M

Input 0.015
Log likelihood -47.496

	Peso	1	0	*Totale*
Contesti linguistici				
e	0.992	3 (60.0)	2 (40.0)	5 (1.8)
q	0.759	7 (14.9)	40 (85.1)	47 (16.8)
n	0.757	2 (8.3)	22 (91.7)	24 (8.6)
t	0.702	4 (7.8)	47 (92.2)	51 (18.3)
i	0.570	1 (20.0)	4 (80.0)	5 (1.8)
g	0.523	1 (11.1)	8 (88.9)	9 (3.2)
o	0.520	1 (14.3)	6 (85.7)	7 (2.5)
c	0.397	1 (4.0)	24 (96.0)	25 (9.0)
d	0.291	1 (2.3)	42 (97.7)	43 (15.4)
p	0.163	1 (1.6)	62 (98.4)	63 (22.6)
Range	*829*			
Variazione diafasica				
B	0.844	21 (13.5)	134 (86.5)	155 (55.6)
A	0.108	1 (0.8)	123 (99.2)	124 (44.4)
Range	*736*			
Adiacenza V SN				
z (no)	0.964	8 (29.6)	19 (70.4)	27 (9.7)
w (sì)	0.413	14 (5.6)	238 (94.4)	252 (90.3)
Range	*551*			
Adiacenza V Testa				
x (sì)	[0.585]	4 (7.8)	47 (92.2)	51 (18.3)
y (no)	[0.481]	18 (7.9)	210 (92.1)	228 (81.7)
	104			

Tabella II: Analisi multivariata: corpus L

Input 0.143
Log likelihood -109.114

	Peso	1	0	*Totale*
Contesti linguistici				
e	0.870	5 (62.5)	3 (37.5)	8 (3.1)
n	0.701	7 (33.3)	14 (66.7)	21 (8.1)
q	0.630	15 (25.0)	45 (75.0)	60 (23.1)
o	0.590	1 (25.0)	3 (75.0)	4 (1.5)
i	0.558	5 (25.0)	15 (75.0)	20 (7.7)
t	0.547	8 (19.5)	33 (80.5)	41 (15.8)
c	0.547	1 (16.7)	5 (83.3)	6 (2.3)
d	0.460	4 (12.9)	27 (87.1)	31 (11.9)
g	0.387	2 (11.8)	15 (88.2)	17 (6.5)
p	0.215	3 (5.8)	49 (94.2)	52 (20.0)
Range	*655*			
Variazione diafasica				
B	0.650	46 (26.9)	125 (73.1)	171 (65.8)
A	0.233	5 (5.6)	84 (94.4)	89 (34.2)
Range	*417*			
Adiacenza V SN				
z (no)	[0.587]	8 (22.9)	27 (77.1)	35 (13.5)
w (sì)	[0.486]	43 (19.1)	182 (80.9)	225 (86.5)
Range	*101*			
Adiacenza V Testa				
x (sì)	[0.519]	8 (19.5)	33 (80.5)	41 (15.8)
y (no)	[0.496]	43 (19.6)	176 (80.4)	219 (84.2)
	23			

Tabella III: Analisi multivariata: corpus F

Input 0.496
Log likelihood -101.627

	Peso	1	0	*Totale*
Contesti linguistici				
e	[0.663]	4 (66.7)	2 (33.3)	6 (4.0)
q	[0.633]	12 (63.2)	7 (36.8)	19 (12.6)
d	[0.538]	24 (53.3)	21 (46.7)	45 (29.8)
n	[0.510]	10 (50.0)	10 (50.0)	20 (13.2)
t	[0.504]	4 (50.0)	4 (50.0)	8 (5.3)
c	[0.447]	1 (50.0)	1 (50.0)	2 (1.3)
o	[0.429]	3 (42.9)	4 (57.1)	7 (4.6)
i	[0.413]	2 (40.0)	3 (60.0)	5 (3.3)
p	[0.401]	14 (40.0)	21 (60.0)	35 (23.2)
g	[0.275]	1 (25.0)	3 (75.0)	4 (2.6)
Range	*220*			
Variazione diafasica				
B	[0.516]	59 (51.3)	56 (48.7)	115 (76.2)
A	[0.449]	16 (44.4)	20 (55.6)	36 (23.8)
Range	*67*			
Adiacenza V SN				
z (no)	[0.608]	8 (61.5)	5 (38.5)	13 (8.6)
w (sì)	[0.490]	67 (48.6)	71 (51.4)	138 (91.4)
Range	*118*			
Adiacenza V Testa				
x (sì)	[0.519]	4 (50.0)	4 (50.0)	8 (5.3)
y (no)	[0.499]	71 (49.7)	72 (50.3)	143 (94.7)
	20			

Tabella IV: Analisi multivariata: corpus M