CLUB
Working Papers in Linguistics

A cura di Emanuele Miola e Rosa Pugliese

Volume 5, 2021

CLUB – CIRCOLO LINGUISTICO DELL'UNIVERSITÀ DI BOLOGNA

ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA

CLUB
Working Papers in Linguistics

A cura di Emanuele Miola e Rosa Pugliese

Volume 5, 2021

CLUB – CIRCOLO LINGUISTICO DELL'UNIVERSITÀ DI BOLOGNA

ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA

Collana
CLUB WORKING PAPERS IN LINGUISTICS
(CLUB-WPL)

Comitato di Direzione
Cristiana De Santis
Nicola Grandi
Francesca Masini
Fabio Tamburini

Comitato Scientifico
Marianna Bolognesi
Claudia Borghetti
Chiara Gianollo
Elisabetta Magni
Yahis Martari
Caterina Mauri
Marco Mazzoleni
Emanuele Miola
Rosa Pugliese
Mario Vayra
Matteo Viale

Il CLUB – Circolo Linguistico dell'Università di Bologna nasce nel 2015 con l'obiettivo di riunire coloro che, all'interno dell'Alma Mater, svolgono attività di ricerca in ambito linguistico.

Il CLUB organizza ogni anno un ciclo di seminari e pubblica una selezione degli interventi nella collana CLUB WORKING PAPERS IN LINGUISTICS.

I volumi, sottoposti a una procedura di peer-review, sono pubblicati online sulla piattaforma AMS Acta dell'Università di Bologna e sono liberamente accessibili.

CC BY-NC

CLUB Working Papers in Linguistics, Volume 5, 2021
ISBN: 9788854970649

CLUB – CIRCOLO LINGUISTICO DELL'UNIVERSITÀ DI BOLOGNA
http://corpora.ficlit.unibo.it/CLUB

CLUB Working Papers in Linguistics

Volume 5
a cura di *Emanuele Miola* e *Rosa Pugliese*

Indice

Presentazione 5
Emanuele Miola e Rosa Pugliese

Trattare gli errori a scuola: fra regolarità attese e sperimentazione linguistica 9
Cecilia Andorno e Anna Boario

ParlaTO: la variazione (e il repertorio) in un corpus di parlato 23
Massimo Cerruti

Prospettive di ricerca per valutare la competenza lessicale in una lingua seconda 41
Francesca Gallina

Potenziali *marker* linguistici nello scritto di pazienti con Anoressia Nervosa: uno studio con tecniche di NLP 57
Giulia Minori

The semantics behind the reanalysis of French articles in French-based creoles 77
Laura Tramutoli

Sezione "Leggere per comprendere a partire dai dati INVALSI"

Insegnare a leggere e capire 103
Mario Ambel

Text comprehension: the role of individual differences 119
Lucia Mason e Angelica Ronconi

Qualche osservazione sociolinguistica sulle prove INVALSI 135
Emanuele Miola

Quattro caratteristiche dell'analisi logica. Considerazioni a margine delle prove INVALSI del grado 8 149
Diego Pescarini

Heri dicebamus. Competenze linguistiche prima e dopo la pandemia 165
Alberto A. Sobrero

ParlaTO: la variazione (e il repertorio) in un corpus di parlato

Massimo Cerruti
Università di Torino
massimosimone.cerruti@unito.it

Abstract
ParlaTO è un corpus di parlato spontaneo raccolto a Torino, di recente allestimento. Insieme al corpus KIP, a cui è per molti aspetti complementare, confluisce nel più ampio KIParla. È costituito da interviste semi-strutturate a informatori con caratteristiche sociodemografiche differenti e comprende prevalentemente produzioni in italiano, pur includendo anche brani in dialetto e lingue d'immigrazione. In questo contributo se ne illustrano le caratteristiche essenziali e si discutono i risultati dei primi studi condotti sui materiali di ParlaTO e KIP, collocandoli nel quadro di alcune tendenze in atto nell'italiano: la crescente circolazione di tratti regionali oltre i confini d'origine, il progressivo ridursi delle differenze linguistiche fra parlanti con collocazione sociale diversa e la presenza sempre più regolare di fenomeni propri del parlato colloquiale anche nell'uso di norma.

1. Introduzione

ParlaTO (www.corpusparlato.com) è un corpus di parlato spontaneo raccolto a Torino, che viene ad aggiungersi ai (pochi) corpora di parlato esistenti per l'italiano.[1] Il corpus consente di indagare l'italiano parlato in città, per come questo si presenta in gruppi sociali differenti e in parlanti con provenienza geografica diversa, talvolta in alternanza con dialetti e lingue d'immigrazione. Un'introduzione al corpus è fornita in Cerruti & Ballarè (2020), a cui si rimanda per la discussione dei criteri di allestimento, della metodologia di raccolta e trattamento dei dati e delle modalità di consultazione.

In questo articolo si presentano le caratteristiche essenziali del corpus, commentando alcuni brani esemplificativi delle varietà d'italiano e delle interazioni bilingui presenti in esso (cfr. paragrafo 2); si illustrano i rapporti fra ParlaTO e KIP, due corpora per molti versi complementari che confluiscono nel KIParla (cfr. paragrafo 3); e si prendono in

[1] Fra i quali, mette conto ricordare per lo meno LIP (badip.uni-graz.at/it, www.parlaritaliano.it/index.php/it/volip), CLIPS (www.clips.unina.it), C-ORAL-ROM (Cresti & Moneglia 2005) e KIP (cfr. paragrafo 3). Per una rassegna recente, comprensiva anche di corpora in via di allestimento, si veda Goria & Ciccolone (2020).

esame i risultati delle prime ricerche condotte su questi due corpora, discutendoli alla luce di alcune tendenze attuali nella variazione sociolinguistica dell'italiano (cfr. paragrafo 4). Si conclude con qualche cenno agli sviluppi futuri del ParlaTO (cfr. paragrafo 5).

2. ParlaTO

2.1 Il corpus

ParlaTO è composto essenzialmente da una serie di conversazioni registrate a Torino fra il 2018 e il 2020 per mezzo di interviste semi-strutturate. Le interazioni hanno coinvolto più di un centinaio di parlanti d'età diversa (giovani, adulti e anziani), alcuni di origine italiana (piemontese e non), altri di origine straniera, e con gradi d'istruzione e tipi di occupazione differenti. Le interviste si sono concentrate su esperienze di vita personale in città (studio, lavoro, attività nel tempo libero o in pensione, ricordi del passato, ecc.) e hanno visto all'opera più intervistatori, ciascuno dei quali quasi sempre appartenente alla rete sociale o familiare dell'intervistato. A una stessa intervista, inoltre, hanno spesso preso parte più intervistati; il che ha dato luogo di frequente a vere e proprie discussioni collettive. Condizioni come queste hanno favorito modalità di interazione tipiche del comportamento *in-group*, caratterizzate dall'uso di varietà spontanee e, in casi specifici, dall'alternanza di lingue diverse.

Il corpus è costituito da due sezioni, contenenti l'una le conversazioni con parlanti di origine italiana e l'altra le conversazioni con parlanti di origine straniera. Al momento è interrogabile online soltanto la prima sezione, che ammonta a circa 50 ore di parlato e comprende produzioni linguistiche in italiano, piemontese e altri dialetti (in particolar modo di area meridionale). L'uso dell'italiano è largamente prevalente in tutte le conversazioni, ed esclusivo nella maggior parte di esse, mentre il piemontese e gli altri dialetti compaiono soltanto occasionalmente, secondo la nota fenomenologia del discorso bilingue. La seconda sezione, attualmente in preparazione, avrà anch'essa una dimensione stimabile in circa 50 ore di parlato e sarà incentrata sull'italiano di nativi e non nativi e sulle lingue di recente immigrazione. Le Tabelle 1 e 2 riportano alcune informazioni essenziali riferite alla prima sezione del corpus.

INFORMATORI	88 parlanti
RACCOLTA DATI	Torino, 2018-2020
METODO	Interviste semi-strutturate (individuali e di gruppo)
LINGUE	Italiano, dialetto piemontese, altri dialetti
METADATI	Parlanti: classe d'età, sesso, regione di nascita, titolo di studio, occupazione, lingua materna, competenza (attiva o passiva) di dialetto/i italo-romanzo/i, competenza (attiva o passiva) di altre lingue, città e quartiere di residenza, città di nascita del padre e della madre. Interazioni: numero di partecipanti, lingue impiegate
ORE DI PARLATO	Giovani ($18 \leq x \leq 30$): 17:33:20
	Adulti ($30 < x \leq 60$): 14:49:53
	Anziani ($60 < x \leq 89$): 16:15:31
DIMENSIONI	552.461 tokens
ACCESSO	www.corpusparlato.com

Tabella 1. Prima sezione del corpus ParlaTO.

Il corpus è dotato di una serie di metadati relativi alle caratteristiche sociodemografiche degli informatori e ad aspetti specifici dell'interazione (cfr. Tabella 1). Alcuni metadati, quali ad esempio la classe d'età, il sesso, la regione di nascita, il titolo di studio e l'occupazione del parlante, sono usabili come filtri di ricerca per l'interrogazione (cfr. Tabella 2);[2] altri, quali il quartiere di residenza dell'informatore o la città di nascita del padre e della madre, sono accessibili soltanto come informazioni supplementari (contenute in tabelle Excel scaricabili dalle pagine del sito web). Oltre a questi, la seconda sezione del corpus avrà alcuni metadati riguardanti più direttamente i parlanti di origine straniera, come il tempo di permanenza e gli anni di studio in Italia.

PARLANTI	Classe d'età:	16-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, 61-65, 66-70, 71-75, 76-80, 81-85, 85-90
	Sesso:	M, F
	Regione di nascita:	Abruzzo, Basilicata, Friuli-Venezia Giulia, Lazio, Lombardia, Piemonte, Puglia, Sardegna, Sicilia, Trentino-Alto Adige, Veneto
	Titolo di studio:	*elem* (diploma di scuola elementare), *medie* (licenza media), *it* (diploma di istituto tecnico o professionale), *lic* (diploma di liceo), *laurea* (laurea triennale, magistrale e a ciclo unico), *phd* (dottorato di ricerca)
	Occupazione:	*artig* (artigiani, operai specializzati e agricoltori), *comm* (professioni qualificate nelle attività commerciali e nei servizi), *disocc* (disoccupati), *impr* (legislatori, imprenditori e alta dirigenza), *intell* (professioni intellettuali, scientifiche e di elevata specializzazione), *nonq* (professioni non qualificate), *oper* (conduttori di impianti, operai di macchinari fissi e mobili e conducenti di veicoli), *pens* (pensionati), *stud* (studenti)
INTERAZIONI	Partecipanti:	2, 3, 4, 5, 6
	Lingue:	italiano, italiano e dialetto

Tabella 2. Prima sezione del corpus ParlaTO: metadati usabili come filtri di ricerca.

Allo stato attuale, l'interrogazione del corpus può avvenire soltanto per forme; è dunque possibile estrarre singole parole, oppure insiemi e sequenze di parole, in formato KWIC e/o ottenere liste di frequenza (PoS-tagging e lemmatizzazione sono tuttavia in fase di sviluppo; cfr. Bosco et al. 2020). A mo' d'esempio, si riporta in Figura 1 il risultato di una ricerca semplice. Ogni occorrenza della forma richiesta (in questo caso, *tipo*) è inserita in una riga di contesto ed è associata al codice della conversazione in cui essa compare. Cliccando sul codice della conversazione (es. *PTA013*), appare un riquadro che fornisce una serie di informazioni; tra le altre: i link alla trascrizione dell'intera conversazione, in formato ortografico (*doc.full_conversation*) e nel 'sistema Jefferson'

[2] Per le denominazioni dei tipi di occupazione si è fatto riferimento alla classificazione delle unità professionali proposta dall'Istituto nazionale di statistica (http://professioni.istat.it), aggiungendovi tre categorie (*disoccupati*, *pensionati*, *studenti*).

(*doc.full_conversation_Jefferson*);[3] il link alla traccia audio (*annotation.audio_file*) della porzione di dialogo individuata dall'occorrenza; e i metadati usabili come filtri di ricerca, che danno conto delle caratteristiche sociodemografiche del parlante e di alcuni aspetti dell'interazione (cfr. Tabella 2).

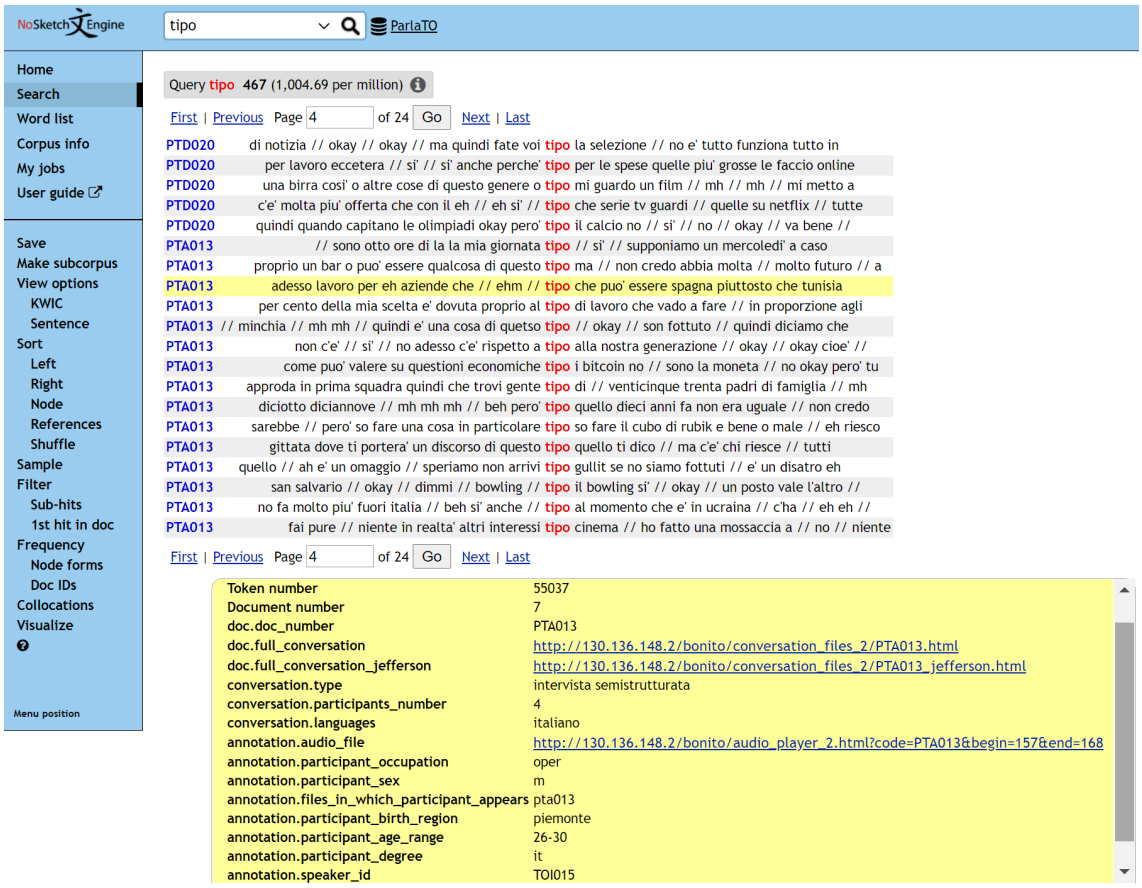

Figura 1. Un esempio di ricerca semplice.

2.2 Varietà d'italiano e discorso bilingue

Come già accennato al paragrafo 2.1, le condizioni dell'intervista hanno favorito l'uso di varietà spontanee, caratterizzate anche da tratti sub-standard. Si veda ad esempio il brano (1),[4] in cui si registrano diversi fenomeni del parlato dialogico; alcuni tipici dell'italiano colloquiale, altri più frequenti nei registri informali. Oltre all'emergere di disfluenze testuali, come autocorrezioni (es. *nella tribu- nel settore ospiti*), cambi di pianificazione (es. *c'è da dire che il terz- eh Milano*) e anacoluti (es. *il Delle Alpi non potevi vedere le partite*), si può facilmente riscontrare la presenza di segnali discorsivi e interiezioni (es.

[3] Le convenzioni usate per la trascrizione conversazionale rappresentano in realtà una versione semplificata del sistema proposto da Jefferson (2004); cfr. Cerruti & Ballarè (2020), Ballarè & Mauri (2020).

[4] Tutti i brani nell'articolo sono in trascrizione conversazionale. Si segnalano qui alcune convenzioni: *[ciao]* sovrapposizione tra parlanti, *(ciao)* sequenza di difficile comprensione, *cia-* parola interrotta, *>ciao<* pronuncia veloce, *((ride))* comportamento non verbale, *(.)* pausa breve, *:* suono prolungato, *x* sequenza incomprensibile. A ciascun partecipante è assegnato un codice alfanumerico (es. *TOI002*, *TOR001*), il terzo carattere del quale consente di distinguere gli informatori (*I*) dai raccoglitori (*R*).

cioè, *vabbò*), elementi deittici, anche in impieghi allusivi (es. *sei lì*), colloquialismi espressivi (es. *te lo giuro*), disfemismi (es. *cazzo*, *minchia*), usi non nominali di *tipo* (es. *è fatto tipo a teatro*), costruzioni negative marcate (es. *non è che si allarga*), dislocazioni a sinistra (es. *il gol non l'ho visto*) e a destra (es. *io le ho viste un paio di partite*), ecc. Alcuni di questi tratti tendono in realtà a perdere, o hanno già perso, l'originario valore sub-standard, essendo ormai in via di accoglimento, o entrati a tutti gli effetti, nell'italiano neo-standard (cfr. ad es. Berruto 2012: 67−126; 2017):

(1) TOI002: no il delle alpi non potevi vedere le partite
TOI002: io le ho viste un paio di partite quando ero bambino minchia
TOI002: cazzo era (.) no eri lontanissimo [perche' poi fra l'altro c'era:]
TOR001: [c'era la (.) pista]
TOI002: [c'era la pista ciclabile] e la ram- minchia [x]
TOI003: [e' un po' co]me l'olimpico eh io quando sono andato all'olim[pico:]
TOI002: [vai a vedere adesso]
TOR001: [mh mh]
TOI002: [minchia adesso:]
TOI002: [adesso e' (sfruttato)]
TOI003: [e' un po' come il nostro terzo] anello
TOI003: pero' c'e' da dire che il terz- eh milano
TOI003: e' fatto pro(i)o b- eh e' fatto tipo a teatro [non e' che si allarga sei li']
TOI002: [vabbo' io l'ho vista x]
TOR001: [si' non s'allarga (almeno)]
TOI003: [sei altis]simo eh vedi minchia [(vedi) delle]
TOI002: [io l'ho vis]to in terza e non la vedi bene come (.) meglio del delle alpi al secon[do]
TOI003: [no]: meglio del delle [alpi sicuro]
TOR001: [davvero? cosi' male::]
TOI002: [eh te lo giuro:] min[chia (.) io]
TOI003: [il delle alpi era male]
TOI002: [sono venuto a inter napoli]
TOI002: no un toro napoli al delle alpi da bambino in serie b
TOI002: e ero in nella tribu- nel: settore ospiti
TOI002: io il gol non l'ho visto del e:::h del toro (quello [che avevano fatto)]
TOI003: [anche all'olimpico vedi m]ale
TOI002: no no ma il delle alp- c(io)e' lo juventus stadium no
(ParlaTO, PTA003)

Nel corpus compaiono inoltre più varietà sociali e (socio-)geografiche. Nel brano (2),[5] ad esempio, sono presenti tratti diagnostici dell'italiano popolare (cfr. D'Achille 2010 e Berruto 2014: 287), come la sovraestensione di forme dell'articolo determinativo (*i stracci*) e gli scambi di ausiliare (*aveva cambiato*, con *cambiare* inaccusativo), e tratti sub-standard particolarmente frequenti in parlanti semicolti, come le reduplicazioni pronominali (*dovesse capitarmi a me*); alcuni dei quali diffusi soprattutto in varietà meridionali (come la forma plurale *melenzane*). A questi si aggiunge un certo numero di

[5] L'intervistato, TOI017, è un parlante anziano (fascia d'età: 85-90 anni) di origine pugliese con sola licenza elementare.

fenomeni in comune con l'italiano colloquiale, come la sovraestensione di *le* per *gli* (*le ho detto*), il dativo etico col pronome di prima persona (*lui m'ha abbassato la testa*) e il rafforzamento dei dimostrativi (*questo qua*); fenomeni che tendono a penetrare, o sono già penetrati, nell'italiano neo-standard (cfr. sopra). Nel brano (3), invece, si notano almeno tre tratti morfosintattici dell'italiano del Piemonte, ben appoggiati al dialetto di sostrato: uno più propriamente locale, il focalizzatore esclusivo e scalare *solo più* (cfr. Regis 2006), e due attestati anche in altre varietà regionali: la costruzione *aver male* + SN (*una ha male la gamba*) e il mancato accordo nel costrutto locativo-esistenziale-presentativo (*c'è tre o quattro campi*, *ce n'è solo più due che giocano*; cfr. Berruto & Cerruti 2015):

(2) TOI017: questo qua m'ha risposto male
TOR004: mh
TOI017: allora io l'ho preso per i stracci e le ho detto, per il braccio e le ho detto dovesse capitarmi a me
TOI017: io non, no qua dentro, fuori
TOI017: ti faccio gli occhi come melenzane, del colore della melanzana
TOI017: capito?
TOR004: si' si'
TOI017: qua non ti faccio niente ma quando ti prendo fuori
TOI017: [allora lui m'ha abbassato] la testa e se n'e' andato e non m'ha risposto piu' pero' (aveva) cambiato da quella sera
TOR004: [te le do di santa ragione]
(ParlaTO, PTB002)

(3) TOI024: quando possiamo (.) giochiamo tra di noi [tanto c'e'] (.) tre o quattro campi
TOI024: e il posto c'e'
TOI024: pero' purtroppo date le eta' eh non stan tutti bene come me
TOR005: eh cer[to]
TOI024: [ma]gari una ha male la gamba l'altra non sta bene e sono mancate del-due tre amiche sono manca[te]
TOR005: [mh mh]
TOI024: e allora adesso:
TOI024: ce n'e' solo piu' due che giocano
(ParlaTO, PTB008)

In certi casi, e il più delle volte *a latere* dell'italiano, sono impiegati anche il piemontese e altri dialetti italoromanzi. Si veda ad esempio il brano (4), in cui intervengono alcune commutazioni di codice, inter- e intra-frasali, fra italiano e piemontese. L'occorrenza di porzioni in dialetto, e più in generale in lingue diverse dall'italiano – quando questi codici appartengano al repertorio linguistico del parlante –, è segnalata da un 'cancelletto' (#) posto all'inizio dell'unità intonativa. Per la trascrizione delle parti in dialetto si è fatto ricorso a un'ortografia di servizio, basata di fatto sulla grafia dell'italiano:

(4) TOI118: m- mio padre (.) era nato in francia
TOI118: #mi am ricordu nen perche': ((*non mi ricordo perché*))
TOI118: #a l'e' mort mi l'avia ses ani ((*è morto che avevo sei anni*))

TOI118: e mamma
TOI118: veneta di noventa vicentina
TOR002: m:h m:h
TOI118: #e::h cume sun surti da:: ((*come sono uscito da*))
TOI118: #da le scole ((*dalle scuole*))
TOI118: #truva' un parent (a l'a) dime ven cun mi ven cun mi:: ((*ho trovato un parente mi ha detto vieni con me vieni con me*))
TOI118: #a fe el tule' l'idr[aulico] ((*a fare l'idraulico*))
TOR002: [mh mh]
TOI118: e cosi' ho iniziato
TOI118: a far l'idraulico (.) nei cantieri: eh

(ParlaTO, PTB026)

Le interazioni che confluiranno nella seconda sezione del corpus vedranno altresì la presenza di varietà d'italiano di parlanti di origine straniera, di prima o seconda generazione, e l'uso alternato di italiano e lingue di recente immigrazione. Alcuni esempi commentati si possono trovare in Cerruti & Ballarè (2020).

3. ParlaTO e KIP

Il corpus ParlaTO si inserisce nel più ampio KIParla (www.kiparla.it, cfr. Mauri et al. 2019), che ospita già un altro corpus linguistico di recente allestimento, denominato KIP. Quest'ultimo è costituito da circa 70 ore di parlato e comprende un insieme di interazioni raccolte a Bologna e Torino fra il 2016 e il 2019 in contesto universitario. Le interazioni hanno coinvolto studenti e professori provenienti da varie regioni d'Italia e hanno avuto luogo in situazioni comunicative di tipo diverso, differenti tra di loro principalmente per il tipo di rapporto fra i partecipanti (simmetrico vs. asimmetrico), e quindi per il grado di formalità: lezioni, esami orali, colloqui in orario di ricevimento, interviste semi-strutturate e conversazioni libere.[6] Il corpus KIP dà conto dunque di una gamma di tipi di interazione più ampia rispetto al ParlaTO, ma considera un campione di parlanti più omogeneo da un punto di vista sociodemografico (cfr. Figura 2).

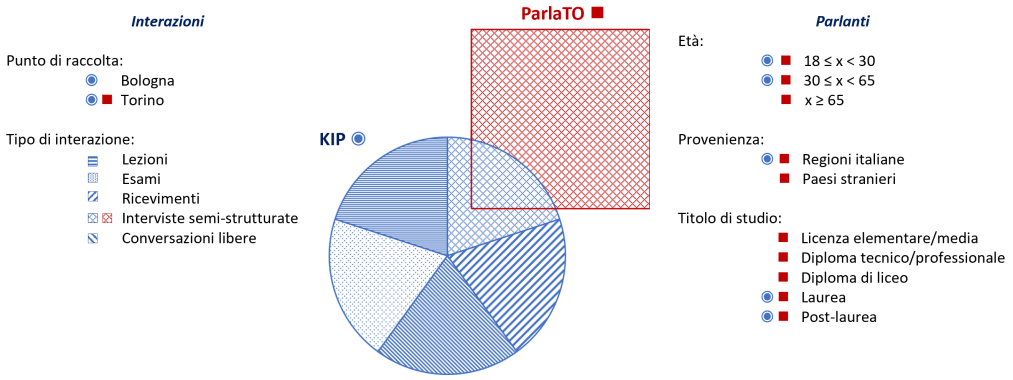

Figura 2. ParlaTO e KIP.

[6] Le interviste semi-strutturate e le conversazioni libere vedono coinvolti soltanto studenti; il rapporto fra i partecipanti è perciò da intendersi come simmetrico.

ParlaTO e KIP hanno evidentemente obiettivi diversi. Il corpus ParlaTO offre in primo luogo l'opportunità di esplorare aspetti di diversificazione diastratica dell'italiano parlato, permettendo oltretutto di condurre ricerche mirate su categorie sociali specifiche (anziani dialettofoni, giovani non universitari, parlanti di origine straniera, ecc.). Il corpus KIP, invece, offre innanzitutto la possibilità di indagare fenomeni di variazione diafasica, e in modo particolare di registro, dell'italiano nel parlato di soggetti colti (studenti e professori universitari). Nondimeno, entrambi i corpora includono produzioni di parlanti con provenienza geografica diversa, e perciò consentono di osservare almeno alcuni elementi di differenziazione diatopica dell'italiano parlato.

ParlaTO e KIP condividono inoltre uno stesso nucleo di metadati (fra i quali, la classe d'età, il sesso, la provenienza geografica e il tipo di occupazione dei parlanti) e lo stesso sistema di raccolta, archiviazione e trascrizione dei dati (cfr. Mauri et al. 2019). Al momento sono consultabili soltanto separatamente, come due corpora distinti, ma, proprio in virtù di queste caratteristiche comuni, saranno presto interrogabili anche unitamente, come un solo corpus (cfr. Cerruti & Ballarè 2020). Si avrà quindi la possibilità di indagare insieme aspetti di differenziazione geografica (KIP e ParlaTO), sociale (ParlaTO) e, limitatamente a parlanti colti, situazionale (KIP) dell'italiano parlato.

4. Primi studi

Le prime ricerche condotte sui materiali dei due corpora si sono concentrate sui comportamenti linguistici di parlanti giovani (segnatamente, di età compresa fra 18 e 30 anni). Nei paragrafi successivi si discuteranno i risultati di questi studi nel quadro di alcuni cambiamenti in atto nella variazione sociolinguistica dell'italiano, prestando particolare attenzione al progressivo ridursi della gamma di variazione geografica e sociale e all'ampliamento della sfera d'influenza di tratti del parlato colloquiale e informale.

4.1 Variazione diatopica

La riduzione generale della differenziazione geografica dell'italiano è ascrivibile in primo luogo a due dinamiche di verso contrario (cfr. Berruto 2017; 2018): da un lato, la progressiva scomparsa dei tratti diatopicamente più marcati, dall'altro la circolazione interregionale di tratti di rango tradizionalmente regionale o locale. Questa seconda dinamica è foriera, in particolare, dello sviluppo di varietà regionali 'composite' (cfr. Canepari 1983; Cerruti 2011), che tendono ovvero a costituirsi non soltanto di tratti autoctoni, riconducibili, per lo meno originariamente, al contatto con i rispettivi dialetti di sostrato, ma anche di tratti alloctoni, provenienti da aree geografiche differenti e diffusisi oltre i confini originari essenzialmente per effetto della mobilità interna della popolazione.[7]

Alcune prime ricerche condotte su porzioni dei corpora ParlaTO e KIP offrono nuovi spunti per la comprensione di questi fenomeni. Fontanot (2019), ad esempio, fornisce un chiaro riscontro empirico della penetrazione di tratti centro-meridionali nell'italiano parlato da giovani torinesi. Il lavoro pone a confronto la distribuzione di una serie di fonemi consonantici in tre insiemi di dati raccolti a Torino: ca. 150 minuti di interviste semi-

[7] Pur senza trascurare ragioni di altra natura, come ad esempio l'influenza dei modelli proposti dai mezzi di comunicazione (cfr. D'Achille 2011).

strutturate a dieci giovani lavoratori, alcuni con licenza media, altri con qualifica tecnica o professionale (ParlaTO); ca. 150 minuti di interviste semi-strutturate a dieci studenti universitari (KIP); ca. 45 minuti di lettura di frasi da parte di sedici studenti universitari (corpus CLIPS). A comparire con una certa regolarità sono soprattutto alcune varianti standard tradizionalmente alloctone per l'area settentrionale e non indotte dal modello scritto, come le geminate intrinseche (Bertinetto & Loporcaro 2005) [t:s], [d:z] e [ʎ:] (es. [dispera't:sjɔ:ne], [a'd:zjɛnda], [bi'ʎ:ɛt:o]) e il raddoppiamento fonosintattico (es. [a'v:ɔlte], [ɔk:a'pi:to], [e'n:ɔj]). Tra le varianti alloctone non standard, gode di una certa diffusione soltanto la pronuncia intensa di [b] in posizione intervocalica (anche in confine di parola, es. ['kwɛl:a 'b:ɛl:a]).

Inoltre, il confronto tra le interviste del ParlaTO e quelle del KIP pone in luce come la presenza di pronunce alloctone sia significativamente più frequente nei giovani con licenza media o qualifica tecnica o professionale, e manifesti quindi un certo grado di marcatezza diastratica. Per quanto riguarda invece gli studenti universitari, il confronto tra le interviste del KIP e le letture di frasi del CLIPS mostra come le produzioni con pronunce alloctone siano più frequenti nel contesto meno sorvegliato, l'intervista, e rivelino dunque un certo grado di marcatezza diafasica.

Si può aggiungere che la diffusione di tratti alloctoni è fortemente variabile da parlante a parlante, e perciò ampiamente soggetta a variazione interindividuale; e gli stessi comportamenti individuali sono spesso incoerenti, si caratterizzano cioè per un continuo "bricolage" (cfr. Eckert 2008) di tratti alloctoni e tratti autoctoni. Si vedano ad esempio gli enunciati (5) e (6), nei quali due diversi allofoni di /ʎ:/ intervocalica, la variante alloctona [ʎ:] e la variante autoctona [lj], coesistono nelle produzioni di uno stesso parlante (in relazione oltretutto alla pronuncia di una stessa parola):

(5) TOI005: quando lavori con la [fa'miʎ:a] e' b:ellissimo
TOR001: mh mh
TOI005: pero' devi essere una [fa'mi:lja] come la nostra
(ParlaTO, PTA005; cfr. Fontanot 2019: 39)

(6) TO041: e' completamente un altro mondo rispetto all'[i'ta:lja] a torino […]
TO041: le persone si comportano in modo totalmente diverso,
TO041: perche' non e' come in [i'taʎ:a] o nel sud europa
(KIP, TOD2005; cfr. Fontanot 2019: 57)

Altri studi vertono sulla presenza di fenomeni di italiano "composito" ai livelli del lessico e della morfosintassi. Cerruti (2018), ad esempio, si concentra sulle produzioni linguistiche di 43 studenti dell'Università di Bologna, provenienti complessivamente da dodici regioni italiane (corpus KIP), e prende in esame un certo numero di *types* morfosintattici ascrivibili ad aree geografiche specifiche, confrontandone l'occorrenza in tre diverse situazioni comunicative: conversazioni libere (ca. 4 ore), interviste semi-strutturate (ca. 4 ore), esami orali (ca. 4 ore). Particolarmente frequente è la presenza di tratti centro-meridionali nel parlato di giovani di origine settentrionale; tra i più ricorrenti: l'apocope di forme verbali, come *so'* e *anda'* in (7), e di altre parti del discorso, come *du' balle* in (8), e la sovraestensione di *stare* in funzione di copula, come *sta praticamente fuori corso* in (7). Negli esempi citati, a produrre questi tratti sono un parlante trentino (BO046) e un parlante emiliano (BO003):

(7) BO046: questa in realta' sta praticamente fuori corso, una media bassissima, eppure l'hanno presa.
BO047: il tirocinio, con l'universita', no?
BO046: mh mh
BO021: ma presa cioe' per meriti suoi, (.) non per?
BO046: no no non e' raccomandata (ne') niente
BO046: [ha fatto domanda e l'hanno presa.]
BO047: [eh, che stavo dicendo?]
BO046: oddio, e' dal giudice di pace, (.) io non ci andrei mai […]
BO046: cioe', a me, gia' non interessa il civile perche' so' cosucce
BO046: se poi devo anda' dal giudice di pace
(KIP, BOA3004)

(8) BO003: a me piaceva ugoli:no
BO002: [il conte ugolino]
BO003: [quel]lo si':, [(.) cioe' proprio quelli catti:vi paolo e france]sca du balle
BO002: [e' meraviglioso]
(KIP, BOA3001)

La Tabella 3 riporta poi il numero di occorrenze complessive dei *types* esaminati in ciascuna delle tre situazioni comunicative, distinguendo fra *tokens* alloctoni e *tokens* autoctoni (il carattere alloctono o autoctono di ogni *token* è definito in relazione alla provenienza geografica dei singoli parlanti). Si può notare non soltanto come il rapporto fra *tokens* alloctoni e *tokens* autoctoni sia maggiore nelle conversazioni libere e abbia un valore nullo negli esami orali, coerentemente con la marcatezza diafasica di alcuni tratti regionali e, più nello specifico, di produzioni composite (cfr. sopra); ma soprattutto come vi sia una differenza di rapporto (oltretutto statisticamente significativa: $F=0{,}0054$, $p<0{,}01$) tra conversazioni libere (0,191) e interviste semi-strutturate (0,064), due contesti interazionali in cui predominano funzioni comunicative diverse: nelle conversazioni libere i parlanti sono intenti soprattutto a gestire la relazione con l'interlocutore e a manifestare o costruire la propria identità sociale, mentre nelle interviste semi-strutturate i partecipanti sono focalizzati principalmente sulla trasmissione di contenuti (cfr. Biber & Conrad 2009: 44−46, 64−69 e *passim*).

	conversazioni libere	interviste semi-strutturate	esami orali
tratti alloctoni (*tokens*)	64	7	0
tratti autoctoni (*tokens*)	335	109	3
rapporto	0,191	0,064	0

Tabella 3. Tratti regionali e contesti interazionali nel corpus KIP (cfr. Cerruti 2018: 21).

Il prevalere di tratti alloctoni nelle conversazioni libere (cfr. Tabella 3) potrebbe dunque stare a indicare come le produzioni "composite" siano, almeno in certi casi, correlate a funzioni comunicative specifiche, e in particolare ai vari modi di espressione linguistica dell'identità sociale. Studi precedenti hanno del resto mostrato come l'uso di tratti alloctoni, deprivato dei potenziali connotativi propri delle varietà regionali originarie, possa assumere per i membri di un certo gruppo sociale un valore simbolico di

affermazione identitaria; si veda, a questo proposito, la diffusione del raddoppiamento fonosintattico (Boario 2017, cfr. sopra) e del suffisso di origine meridionale *-uso* (es. *incazzuso*, *stiluso*, *punkettuso*; Cortinovis & Miola 2009) in gruppi giovanili torinesi.

4.2 Variazione diastratica

Per quanto riguarda la variazione diastratica, una delle principali dinamiche in atto consiste nella riduzione progressiva delle differenze linguistiche fra parlanti con collocazione sociale diversa (cfr. Cortelazzo 2001; Sobrero 2005; Berruto 2012: 157−159), anche in relazione alla tendenziale obsolescenza delle varietà popolari (cfr. Berruto 2005; 2014). Alcuni tratti dell'italiano popolare sono ormai largamente diffusi pure nel sub-standard di parlanti colti e, benché possano essere comunque più frequenti nelle produzioni di semicolti, hanno perso in buona parte il proprio carattere diagnostico. Altri invece, tuttora marcati come "bassi" in diastratia, sono usati sempre meno di generazione in generazione e sembrano destinati a scomparire. È certamente possibile riscontrare ancora la presenza di tratti esclusivi dell'italiano popolare, specialmente in parlanti anziani; si vedano ad esempio i fenomeni individuati in (2) e l'uso del condizionale nella protasi di un'ipotetica libera in (9). Ma le differenze maggiori tra parlanti colti e semicolti tendono a essere sempre meno categoriche e sempre più di carattere graduale:

(9) TOI017: bravo m'ha detto, bravo ma
TOI017: e allora:
TOI017: se dovessi prendere una,
TOI017: una:: una camera, una stanza
TOI017: e dovresti riempirla di carbone,
TOI017: e difatti io ho detto, noi andiamo avanti con la legna li'
(ParlaTO, PTB002)

Si possono collocare in questo quadro, ad esempio, i risultati di un lavoro recente sui fenomeni di variazione osservabili nel settore delle costruzioni relative (Ballarè et al. 2019). Lo studio pone a confronto produzioni linguistiche di parlanti con collocazione sociale diversa, tra cui ca. 10 ore di interviste semi-strutturate a giovani lavoratori con licenza media o qualifica tecnico-professionale (ParlaTO) e ca. 10 ore di interviste semi-strutturate a studenti universitari (KIP). Le differenze riscontrabili fra i due gruppi di parlanti risiedono soprattutto nella diversa frequenza d'uso di costruzioni relative sub-standard (o neo-standard, cfr. paragrafo 4.3): pur presenti in entrambi i gruppi, queste costruzioni risultano più diffuse (con significatività statistica $p<0,01$ o $p<0,05$, cfr. Ballarè et al. 2019: 77−78) nei giovani con licenza media o qualifica tecnico-professionale, siano esse (seguendo la terminologia di Comrie & Kuteva 2013) del tipo *gap*, come *Rino Gaetano che ho comprato il CD* in (10), *pronoun retention*, come *uno che non gli piace il disordine* in (11), o *relative pronoun*, come *una generazione in cui il lavoro va bene* in (12):

(10) TOI001: se no:: qualsiasi genere
TOI001: da:: rino gaetano che ho comprato il cd: l'altro ieri,
TOI001: anche lou reed ho comprato il cd l'altro ie-
TOI001: pero'::

TOI001: mina: e::h
(ParlaTO, PTA001)

(11) TOI005: io non sopporto il disordine sono uno che non gli piace il disordine
TOI005: anzi son fatto cosi'
TOI005: c'e' il disordine che io non posso mettere (a posto)
TOI005: vado in palla, tientelo
(ParlaTO, PTA005)

(12) TOI012: piu' che altro[: s-] siamo una generazione in cui
TOI013: [per me si']
TOI012: il lavoro::: va bene. punto.
TOI012: perche' o prendi quello o stai a casa a dormire
(ParlaTO, PTD021)

Lo studio mostra inoltre come uno stesso processo di grammaticalizzazione possa trovarsi a uno stadio più avanzato nella varietà di un certo gruppo sociale e meno in un'altra. È il caso, in particolare, della grammaticalizzazione di *dove*, che segue grosso modo un percorso di questo tipo: a partire dall'impiego come relativizzatore di locativi, *dove* può acquisire una funzione di *concern* o *aboutness* (cf. Ballarè & Micheli 2018; Ballarè & Inglese 2021), come in (13), e in una fase successiva può sviluppare la capacità di relativizzare il soggetto di frase (cfr. Brook 2011). Dai dati riportati in Ballarè et al. (2019: 85−86), l'ampliamento funzionale di *dove* raggiunge quest'ultimo stadio soltanto nel parlato di giovani con licenza media o qualifica tecnico-professionale, come in (14). Si può constatare, del resto, come le fasi più avanzate di vari processi di grammaticalizzazione compaiano proprio nelle varietà sociali dei parlanti meno istruiti (cfr., ad esempio, Berretta 1989; Ramat 2006; Berruto & Cerruti 2015; Amenta 2020):

(13) BO013: i comunisti, sono gente tinta assa' ((legge))
BO013: tinto, vuol dire (.) cattivo eh?
BO013: vuol di- vorrebbe dire oscu::ri insomma eh?
BO013: pero', i pensieri tinti sono pensieri brutti
BO013: dove la gente tinta e' la gente cattiva e cosi' via
(KIP, BOD1001)

(14) TOI012: comunque::: no bello e' una bella zona:: lati negativi:: di venaria:: […]
TOI012: i ragazzini under sedici dove stan li' a far casino al giardinetto sotto casa
TOI012: ma quello e' normale: eh son ragazzi
(ParlaTO, PTD021)

Analogamente, è soltanto nelle produzioni di giovani non universitari che si riscontrano casi di accordo semantico nella relativizzazione del soggetto; come, ad esempio, *son di quelli che […] non guarda molto cosa paga* in (15) ed *ero il pivello che facevo queste cose qui* in (16). La relativizzazione del soggetto con accordo semantico riflette bene certi aspetti della natura linguistica delle varietà sub-standard e, più in particolare, dell'italiano popolare: una varietà caratterizzata dal prevalere della semantica sulla sintassi, dalla denotazione di significati essenziali e dall'indicazione di referenti "concreti" (cfr. Berruto

1990). Sono fenomeni come questo, dunque, che, pure nel quadro di una generale riduzione della variazione diastratica, rivelano ancora la natura "popolare" di alcune varietà sociali:

(15) TOI005: a me pia- e' sempre piaciuto andare a mangiar fuori
TOR001: mh mh mh
TOI005: son di quelli che quando va a mangiar fuori no:n non guarda:: (.) molto [cosa paga]
TOR001: [mh mh mh]
(ParlaTO, PTA005)

(16) TOI010: ho passato anche queste fasi qui nel senso che sono
TOI010: cresciuto (.) e:ro il pivello che facevo queste cose qui le facevo male
TOI010: adesso (.) mi- (.) credo di farle un po' meglio
(ParlaTO, PTD019)

4.3 Variazione diafasica

Per ciò che concerne la variazione diafasica, una delle maggiori tendenze in atto nell'italiano è l'ampliamento del raggio d'impiego dei registri bassi (Berruto 2012: 196); che porta con sé, fra l'altro, l'uso sempre più regolare di tratti del parlato colloquiale e informale anche in situazioni sorvegliate e il conseguente ingresso nella norma di tratti tradizionalmente sub-standard. Si tratta evidentemente della dinamica alla base del processo di ristandardizzazione dell'italiano, e quindi dello sviluppo di una varietà neo-standard (Berruto 2012; cfr. anche Cerruti, Crocco & Marzo 2017). Più in particolare, questo processo dà luogo alla coesistenza di due varietà standard: l'italiano neo-standard e l'italiano standard letterario; la prima più usata in certi domini, come ad esempio in vari ambiti della comunicazione mediatica, la seconda più presente in altri, come ad esempio nell'istruzione scolastica (cfr. anche Antonelli 2011: 51).

La coesistenza delle due varietà standard d'italiano è sottoposta a verifica empirica in uno studio recente (Cerruti & Vietti in stampa), condotto su una parte di materiali del corpus KIP. Lo studio prende in esame la distribuzione di 23 tratti lessicali e morfosintattici, propri alcuni dell'italiano neo-standard e altri dell'italiano standard letterario,[8] in tre tipi diversi d'interazione: conversazioni libere (ca. 4 ore), interviste semi-strutturate (ca. 4 ore), esami orali (ca. 4 ore). Il lavoro si propone, sostanzialmente, di accertare quali di quei tratti tendano a co-occorrere in maniera statisticamente significativa in ciascuna delle tre situazioni comunicative indagate. La tecnica statistica usata è l'analisi delle componenti principali (PCA, *Principal Component Analysis*), che consente di rilevare l'esistenza di rapporti di correlazione tra variabili.

[8] Fra i primi, ad esempio, si sono considerati alcuni ordini marcati dei costituenti (frasi scisse, dislocazioni a sinistra, dislocazioni a destra), costruzioni relative di tipo *gap* (cfr. paragrafo 4.2), le forme aferetiche del dimostrativo prossimale, *questo/quello* in funzione di dimostrativo neutro, *gli* come dativale sincretico, *ci* come locativo e l'uso dei pronomi personali *lui*, *lei*, *loro* con valore riflessivo. Fra i secondi, la costruzione passiva, le relative del tipo *relative pronoun* (cfr. paragrafo 4.2), le forme non aferetiche del dimostrativo prossimale, *ciò* come dimostrativo neutro, *le* come dativale femminile (singolare), *vi* come locativo e la presenza del pronome riflessivo *sé*.

L'analisi mostra che tutti i tratti dell'italiano neo-standard considerati (con una sola eccezione) correlano con il carattere informale dell'interazione, definito essenzialmente dal rapporto simmetrico fra i partecipanti. I fenomeni neo-standard tendono cioè a co-occorrere regolarmente nel corso di interviste semi-strutturate e conversazioni libere (cfr. paragrafo 3). Nel brano (17), ad esempio, tratto da una conversazione libera, si può osservare la co-occorrenza di una dislocazione a sinistra (*Grey's Anatomy la guardano tutti*), una relativa di tipo *gap* (*American Gigolò che mi pare ci sia sempre lui*) e un costrutto locativo con *ci* (*mi pare ci sia sempre lui*). I tratti neo-standard tendono poi a presentarsi insieme a fenomeni del parlato colloquiale e informale; quali ad esempio, in (17), colloquialismi espressivi (es. *tanta roba*), imprecazioni (es. *porca miseria*), e termini dal significato generale (come *fare*, es. *quello che ha fatto Derek*):

(17) BO003: l'ultimo cavaliere
BO002: non l'ho [mai visto]
BO003: [mia mamma lo guar]da se:mpre li' c'e' lui (.) giovane. col capello lungo
BO002: mi pare: di aver visto american gigolo' (.) che mi pare ci sia sem[pre lui]
BO003: [si'] >porca miseria< (.) anche pretty woman
BO002: si'. (.) >(proprio)< e'
BO003: si'. li' lui si'
BO002: no no era tanta roba [anche li']
BO003: [si' si'] madonna […]
BO003: poi anche
BO003: >non so< se lo conosci
BO003: ehm (.) patrick dempsey: quello che ha fatto derek in gre[y's anatomy] anche lui
BO002: [ah: si':]
BO002: si' [piace-]
BO003: [nei suoi momen]ti d'o:ro [era]
BO002: [pia]ce a ta::n[ti (.)] si']
BO003: [davvero]
BO003: anche lui era bello. grey's anatomy tutti
BO002: si'. grey's anatomy, la guardano tu:tti
(KIP, BOA3001)

La stessa analisi mostra inoltre che tutti i tratti dell'italiano standard letterario (anche in questo caso con una sola eccezione) correlano invece con il carattere formale della situazione comunicativa, determinato fondamentalmente dalla relazione asimmetrica fra i parlanti. I tratti dello standard letterario tendono ovvero a co-occorrere regolarmente durante gli esami orali. In (18), ad esempio, si può notare la co-occorrenza di una frase passiva (*Anselmo fu cacciato dall'abbazia*), una costruzione con un pronome relativo (*re Desiderio* […] *del quale era oppositore*) e un costrutto locativo con *vi* (*prima vi era* […] *la raccolta di libri dall'esterno*). Fenomeni come questi tendono poi a presentarsi insieme a tratti dell'italiano formale; si veda ad esempio, in (18), l'uso frequente della subordinazione frasale e la ricorrenza di termini specifici (es. *oppositore*) e dal significato astratto (es. *attività*):

(18) BO088: un episodio molto importante e:h si ebbe quando e:hm anselmo fu cacciato
BO088: eh da- dall'abbazia

BO088: eh dall'abazia, (.) ehm da da re desiderio, al quale appunto anselmo era::: era (.) del quale era oppositore diciamo,
BO088: e ehm si rifugio' a monte cassino […]
BO088: e fu appunto e:hm una volta che carlo magno conquisto' l'italia, fu permesso ad anselmo di tornare eh a mantova,
BO088: nel settecentosettantaquattro,
BO088: e::hm::e cosi' po- pote' riprendere la sua attivita', prima e:::h di creazione della biblioteca,
BO088: perche' secondo appunto l'uso eh delle biblioteche=e::h
BO088: medioev- medievali diciamo, prima e:h vi era
BO088: m:h la insomma la raccolta di libri dall'esterno
(KIP, BOC1001)

È perciò possibile identificare due *cluster* di tratti standard co-occorrenti: l'uno di tratti neo-standard, l'altro di tratti dello standard letterario. Italiano neo-standard e italiano standard letterario paiono quindi effettivamente manifestarsi come due varietà separate, autonome e coerenti, nei comportamenti linguistici dei parlanti (per lo meno nel gruppo osservato, piuttosto omogeneo e ben definito in termini di collocazione sociale, e in certe situazioni comunicative).[9] Occorre inoltre mettere in risalto che questi due *cluster* di tratti correlano, come si è detto, con il carattere informale *vs.* formale dell'interazione verbale. Ciò è rilevante perché svela l'esistenza di *pattern* sistematici nella variazione di registro, determinati non soltanto dall'alternanza fra singoli tratti ma – più significativamente – dall'alternanza fra insiemi di tratti co-occorrenti; è noto infatti come la variazione di registro sia più spesso caratterizzata da co-occorrenze invece poco sistematiche di tratti linguistici (cfr. ad es. Ghyselen & De Vogelaer 2018: 2).

Si può infine sottolineare che i tratti dell'italiano neo-standard tendono a co-occorrere in interazioni verbali di tipo diverso, nelle quali prevalgono funzioni comunicative differenti: le interviste semi-strutturate, incentrate sulla trasmissione di contenuti, e le conversazioni libere, dominate dalla relazione fra gli interlocutori e dalla costruzione di identità sociale (cfr. paragrafo 4.1). Ciò è interpretabile come un riflesso del carattere "multi-funzionale" dell'italiano neo-standard, varietà d'uso medio di una lingua con funzioni di distanza e immediatezza comunicativa (la "multi-funzionalità" è, più in generale, un carattere comune a tutte le varietà neo-standard di lingue europee; cfr. ad es. Mattheier 1997 e Coupland & Kristiansen 2011). Appare invece più circoscritto l'uso dello standard tradizionale, coerentemente con il suo essere "lingua di livello letterario" (Berruto 2012: 26).

5. Conclusioni

I primi studi condotti sui materiali di ParlaTO e KIP forniscono riscontri empirici e nuovi spunti su alcune tendenze in atto nella gamma di variazione sociolinguistica dell'italiano; in particolare, sulla crescente circolazione interregionale di fenomeni propri di aree geografiche diverse, col conseguente sviluppo di produzioni linguistiche, se non varietà di lingua, "composite"; sul progressivo venir meno delle differenze di natura categorica

[9] Può forse valer la pena chiarire che le etichette *neo-standard* e *standard letterario* non sono assegnate in seguito ai risultati dello studio, che in sé non dimostra il carattere neo-standard o standard letterario dei tratti considerati, ma sono attribuite a priori in base alla letteratura esistente.

fra gruppi sociali, dettate cioè dalla presenza di tratti esclusivi di certe varietà diastratiche; e sull'affermarsi della coesistenza di due varietà standard, l'una di impronta letteraria e l'altra derivante dalla promozione di tratti sub-standard nell'uso di norma.

La gamma delle varietà rappresentate nel ParlaTO sarà poi utilmente ampliata, come si è detto, con la realizzazione di una sezione dedicata alle produzioni di parlanti di origine straniera, che offrirà tra l'altro l'opportunità di indagare una serie di aspetti dell'italiano di non nativi. Inoltre, le possibilità di interrogazione di ParlaTO e KIP saranno potenziate sia dalla consultazione congiunta dei due corpora sia dalla ricerca per lemmi e parti del discorso. L'auspicabile allestimento di nuovi moduli del KIParla (cfr. Mauri et al. 2019) potrà infine arricchire la rappresentatività, anche in termini di varietà di lingua, dell'intera risorsa.

Bibliografia

Amenta, Luisa. 2020. *Strutture tempo-aspettuali nel siciliano e nell'italiano regionale di Sicilia*. Palermo: Centro di studi Filologici e Linguistici Siciliani.

Antonelli, Giuseppe. 2011. Lingua. In Afribo, Andrea & Zinato, Emanuele (a cura di), *Modernità italiana. Cultura, lingua e letteratura dagli anni Settanta a oggi*, 15–52. Roma: Carocci.

Ballarè, Silvia & Cerruti, Massimo & Goria, Eugenio. 2019. Variazione diastratica nel parlato di giovani: il caso delle costruzioni relative. In Moretti, Bruno & Kunz, Aline & Natale, Silvia & Krakenberger, Etna (a cura di), *Le tendenze dell'italiano contemporaneo rivisitate. Atti del LII Congresso Internazionale di Studi della Società di Linguistica Italiana (Berna, 6-8 settembre 2018)*, 75–94. Milano: Officinaventuno.

Ballarè, Silvia & Mauri, Caterina. 2020. La creazione del corpus KIParla: criteri metodologici e prospettive future. *Rivista italiana di dialettologia* 44. 53–69.

Ballarè, Silvia & Micheli, M. Silvia. 2018. Usi di *dove* nell'italiano contemporaneo: costruzioni relative e dinamiche di ristandardizzazione. *Linguistica e Filologia* 38. 29–56.

Ballarè, Silvia & Inglese, Guglielmo. 2021. The development of locative relative markers: from typology to sociolinguistics (and back). *Studies in language*. (https://www.jbe-platform.com/content/journals/10.1075/sl.20013.bal) (Consultato il 15.06.2021.)

Berretta, Monica. 1989. Tracce di coniugazione oggettiva in italiano. In Foresti, Fabio & Rizzi, Elena & Benedini, Paola (a cura di), *L'italiano fra le lingue romanze. Atti del XX Congresso Internazionale di Studi della SLI (Bologna, 25-27 settembre 1986)*, 125–150. Roma: Bulzoni.

Berruto, Gaetano. 1990. Semplificazione linguistica e varietà sub-standard. In Holtus, Günter & Radtke, Edgar (a cura di), *Sprachlicher Substandard*, Vol. III, 17–43. Tübingen: Niemeyer.

Berruto, Gaetano. 2005. Intervento alla tavola rotonda. In Lo Piparo, Franco & Ruffino, Giovanni (a cura di), *Gli italiani e la lingua*, 332–342. Palermo: Sellerio.

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

Berruto, Gaetano. 2014. Esiste ancora l'italiano popolare? Una rivisitazione, in Danler Paul & Konecny, Christine (a cura di), *Dall'architettura della lingua italiana*

all'architettura linguistica dell'Italia. Saggi in omaggio a Heidi Siller–Runggaldier, 277–290. Frankfurt am Main: Peter Lang.

Berruto, Gaetano. 2017. Dinamiche nell'architettura delle varietà dell'italiano nel ventunesimo secolo. In Caprara, Giovanni & Marangon, Giorgia (a cura di), *Italiano e dintorni. La realtà linguistica italiana: approfondimenti di didattica, variazione e traduzione*, 7–31. Frankfurt am Main: Peter Lang.

Berruto, Gaetano. 2018. Tendenze nell'italiano del Duemila e rapporti fra varietà standard e sub-standard. *AggiornaMenti* 13. 5–15.

Berruto, Gaetano & Cerruti, Massimo. 2015. Un esercizio di analisi variazionista: l'accordo verbale nel costrutto locativo-esistenziale-presentativo. In Busà, Maria Grazia & Gesuato, Sara (a cura di), *Lingue e contesti. Studi in onore di Alberto M. Mioni*, 609–620. Padova: CLEUP.

Bertinetto, Pier Marco & Loporcaro, Michele. 2005. The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of The Phonetic Association* 35. 131–151.

Biber, Douglas & Conrad, Susan. 2009. *Register, genre, and style*. Cambridge: Cambridge University Press.

Boario, Anna. 2017. Native speaker–non-native speaker interaction: More than one type of accommodation. *Sociolinguistic studies* 11(2-3-4). 291–312.

Bosco, Cristina & Ballarè, Silvia & Cerruti, Massimo & Goria, Eugenio & Mauri, Caterina. 2020. KIPoS @ EVALITA2020: Overview of the Task on KIParla Part of Speech tagging. In Basile, Valerio & Croce, Danilo & Di Maro, Maria & Passaro, Lucia C. (a cura di), *Proceedings of the 7th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2020)*. (CEUR.org, http://ceur-ws.org/Vol-2765/paper154.pdf) (Consultato il 15.06.2021.)

Brook, Marisa. 2011. One of those situations where a relative pronoun becomes a complementizer: A case of grammaticalization in progress… again. In Armstrong, Lisa (a cura di), *Proceedings of the 2011 Annual Meeting of the Canadian Linguistic Association*, 1–7. (http://homes.chass.utoronto.ca/~cla-acl/actes2011/actes2011.html) (Consultato il 15.06.2021.)

Canepari, Luciano. 1983. *Italiano standard e pronunce regionali*. Padova: CLEUP.

Cerruti, Massimo. 2011. Regional Varieties of Italian in the Linguistic Repertoire. *International Journal of the Sociology of Language* 210. 9–28.

Cerruti, Massimo. 2018. Il parlato regionale oggi: un italiano composito? *LId'O Lingua italiana d'oggi* 15. 15–31.

Cerruti, Massimo & Ballarè, Silvia. 2020. ParlaTO: corpus del parlato di Torino. *Bollettino dell'Atlante Linguistico Italiano* 44. 13–38.

Cerruti, Massimo & Crocco, Claudia & Marzo, Stefania (a cura di). 2017. *Towards a new standard: Theoretical and empirical studies on the restandardization of Italian*. Berlin-New York: Mouton de Gruyter.

Cerruti, Massimo & Vietti, Alessandro. In stampa. Identifying language varieties: Coexisting standards in spoken Italian. In Beaman, Karen & Guy, Gregory R. (a cura di), *The coherence of linguistic communities: Orderly heterogeneity and social meaning*. London: Routledge.

Comrie, Bernard & Kuteva, Tania. 2013. Relativization Strategies. In Dryer, Matthew S. & Haspelmath, Martin (a cura di), *The World Atlas of Language Structures Online*. (http://wals.info/chapter/s8) (Consultato il 15.06.2021.)

Cortelazzo, Michele A. 2001. L'italiano e le sue varietà: una situazione in movimento.

Lingua e stile 34(3). 417–430.
Cortinovis, Enrica & Miola, Emanuele. 2009. Chiamami *rappuso*: il suffisso *-uso* dai dialetti meridionali al linguaggio giovanile torinese. *Rivista italiana di dialettologia* 33. 195–218.
Coupland, Nikolas & Kristiansen, Tore. 2011. SLICE: Critical perspectives on language (de)standardization. In Kristiansen, Tore & Coupland, Nikolas (a cura di), *Standard languages and language standards in a changing Europe*, 11–35. Oslo: Novus.
Cresti, Emanuela & Moneglia, Massimo. 2005. *C-Oral-Rom. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam-Philadelphia: John Benjamins.
D'Achille, Paolo. 2010. *Italiano popolare*. In Simone, Raffaele (a cura di), *Enciclopedia dell'italiano*, Vol. I, 723–725. Roma: Istituto dell'Enciclopedia italiana.
D'Achille, Paolo. 2011. *Roma, italiano di*. In Simone, Raffaele (a cura di), *Enciclopedia dell'italiano*, Vol. II, 1262–1265. Roma: Istituto dell'Enciclopedia italiana.
Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12. 453–476.
Fontanot, Arianna. 2019. *Aspetti di italiano composito nella pronuncia di giovani torinesi: un'indagine su corpora*. Torino: Università di Torino. (Tesi di laurea magistrale.)
Goria, Eugenio & Ciccolone, Simone (a cura di). 2020. Corpora di parlato nel panorama italiano: verso l'individuazione di pratiche condivise. Sezione monografica di *Rivista italiana di dialettologia* 44. 17–169.
Ghyselen, Anne-Sophie & De Vogelaer, Gunther. 2018. Seeking Systematicity in Variation: Theoretical and Methodological Considerations on the "Variety" Concept. *Frontiers in Psychology* 9. (https://doi.org/10.3389/fpsyg.2018.00385) (Consultato il 15.06.2021.)
Jefferson, Gail. 2004. Glossary of transcript symbols with an introduction. In Lerner, Gene H. (a cura di), *Conversation Analysis: studies from the first generation*, 13–31. Amsterdam-Philadelphia: John Benjamins.
Mattheier, Klaus J. 1997. Über Destandardisierung, Umstandardisierung und Standardisierung in modernen europäischen Standardsprachen. In Mattheier, Klaus J. & Radtke, Edgar (a cura di), *Standardisierung und Destandardisierung europäischer Nationalsprachen*, 1–11. Frankfurt am Main: Peter Lang.
Mauri, Caterina & Ballarè, Silvia & Goria, Eugenio & Cerruti, Massimo & Suriano, Francesco. 2019. KIParla Corpus: A New Resource for Spoken Italian. In Bernardi, Raffaella & Navigli, Roberto & Semeraro, Giovanni (a cura di), *Proceedings of the Sixth Italian Conference on Computational Linguistics*. (http://ceur-ws.org/Vol-2481/paper45.pdf) (Consultato il 15.06.2021.)
Ramat, Paolo. 2006. Italian negation from a typological/areal point of view. In Grandi, Nicola & Iannàccaro, Gabriele (a cura di), *Zhì. Scritti in onore di Emanuele Banfi in occasione del suo 60° compleanno*, 355–370. Cesena-Roma: Caissa.
Regis, Riccardo. 2006. Breve fenomenologia di una locuzione avverbiale: il *solo più* dell'italiano regionale piemontese. *Studi di lessicografia italiana* 23. 275–289.
Sobrero, Alberto A. 2005. Come parlavamo, come parliamo. Spunti per una microdiacronia delle varietà dell'italiano. In Lo Piparo, Franco & Ruffino, Giovanni (a cura di), *Gli italiani e la lingua*, 209–220. Palermo: Sellerio.