

Il ritmo del tedesco spontaneo alla luce del modello Controllo/Compensazione: prime esplorazioni

Pier Marco Bertinetto, Chiara Bertini*, Lorenzo Guidugli°*

* Scuola Normale Superiore, Pisa

° Università di Pisa

A Ullrich Wandruszka
per i suoi settant'anni

1. Stato delle conoscenze nelle ricerche sul ritmo delle lingue naturali

Lo studio sperimentale delle propensioni ritmiche delle lingue naturali è un'impresa scientifica che si potrebbe ormai, con buone ragioni, definire “di lungo corso”. Prendendo lo storico lavoro di Classe (1939) come punto d'avvio, risulta chiaro che la vicenda, iniziata tre quarti di secolo fa, non accenna a perdere d'intensità e d'interesse. Dopo una fase di rallentamento negli anni Ottanta del secolo scorso, le ricerche sono riprese con rinnovata lena a partire dagli anni Novanta, quando nuove proposte di modellizzazione hanno cominciato a circolare.

In un recente lavoro che ne ripercorre lo sviluppo, Bertinetto & Bertini (2010) hanno messo in risalto luci ed ombre di questo filone di studi, mirando ad inquadrarlo in una prospettiva dichiaratamente epistemologica. Ai due requisiti di base dell'esplicitezza e della predittività, di per sé irrinunciabili, questi autori hanno proposto di aggiungere un terzo requisito, da valutarsi in rapporto alla specifica tematica trattata: quello della “unificazione”. Con ciò si allude al fatto che un modello autenticamente maturo del ritmo delle lingue naturali debba fondarsi su una concezione unitaria, anziché bipolare. Se agli esordi ci si poteva accontentare di una drastica dicotomia – la tradizionale contrapposizione “isocronia sillabica vs. accentuale” (*syllable-* vs.

stress-timing) – si deve ormai puntare decisamente a sviluppare una trattazione unitaria, che veda tutte le lingue naturali all'interno di un medesimo spazio, anziché dirottate su piani alternativi ed incommensurabili. L'ipotesi di fondo consiste insomma nel concepire le lingue, dal punto di vista ritmico, come fondate sui medesimi ingredienti. Tutte, infatti, presentano sillabe e prominenze di frase (comunque queste ultime si manifestino in concreto); le differenze specifiche saranno da intendersi come scarti gradualmente determinati dall'effetto congiunto delle varie proprietà prosodiche che caratterizzano ciascuna lingua, anziché come il frutto di irriducibili incompatibilità strutturali.

Si rimanda al lavoro citato per maggiori dettagli su questo tema. Basti qui ripetere le conclusioni della sommaria analisi condotta dai due autori: nessuno dei modelli fino ad allora proposti (il lavoro, pubblicato nel 2010, fu concepito nel 2008) poteva dirsi pienamente adeguato dal punto di vista epistemologico, mancando tutti ora dell'uno, ora dell'altro requisito. Ciò non toglie che l'accumulo di conoscenze realizzatosi nel corso del tempo abbia prodotto importanti risultati, che permettono di meglio valutare i problemi di fondo, nonché le divergenze di comportamento tra le lingue. Sappiamo, per esempio, che le presunte regolarità associabili alla durata tendenzialmente costante di sillabe ed intervalli interaccentuali non sono mai state concretamente riscontrate; e sappiamo anche che l'idea di attribuire le impressioni di regolarità ritmica ai meccanismi della percezione – come una sorta di “illusione ritmica” – non ha mai raggiunto lo stadio della definitiva verifica empirica (cf. per es. Lehiste 1973, Allen 1975). Sappiamo, peraltro, che le lingue manifestano comportamenti differenziati, e puntualmente misurabili, per ciò che riguarda la tendenza alla compensazione delle durate in rapporto alle unità sillabiche che precedono o seguono la sillaba “bersaglio” entro la medesima parola o il medesimo intervallo interaccentuale (il lavoro di Lindblom & Rapp 1973 può essere considerato il punto d'avvio di questo filone di studi). Sappiamo, soprattutto, che le diverse proprietà che compongono la struttura prosodica di ciascuna lingua concorrono congiuntamente a determinare l'impronta ritmica complessiva, anche se con effetti difficilmente disaggregabili quanto al contributo delle singole componenti (Bertinetto 1981, Dauer 1983). In particolare, si ritiene che i seguenti fattori – ciascuno presentato qui sotto come tratto binario, ma in realtà finemente graduabile – rivestano speciale rilevanza per quanto concerne le ripercussioni ritmiche. Si noterà di certo che, nella lista seguente, l'ordine con cui i vari fattori si susseguono non è casuale. In termini tradizionali, possiamo dire che i primi membri di ciascuna coppia puntano verso la così detta “isocronia accentuale”, mentre i secondi tendono decisamente alla “isocronia sillabica”:

- riduzione delle vocali non accentate vs. piena articolazione delle medesime;
- struttura sillabica complessa (e relativa incertezza dei confini sillabici) vs. struttura semplice (e confini sillabici ben definiti);
- relativa flessibilità vs. rigidità nella collocazione dell'accento di parola (ove tale parametro sia pertinente);

- densità vs. rarefazione degli accenti secondari di parola;
- accelerazioni ottenute prevalentemente a spese delle sillabe non accentate vs. compressione tendenzialmente proporzionale di tutte le sillabe.

Proprio partendo da questo tipo di considerazioni, Bertinetto & Bertini (2010) hanno proposto un modello biplanare del ritmo delle lingue naturali, i cui due livelli sono rispettivamente denominati:

- livello-I (FONOTATTICO), basato sull'alternanza di due oscillatori, "vocalico" e "consonantico";
- livello-II (FRASTICO), i cui due oscillatori sono chiamati "accentuale" e di "picco sillabico". Il secondo coincide, in pratica, con l'oscillatore vocalico del livello-I, il che implica una stretta coarticolazione fra i due livelli. Il primo oscillatore consiste invece nel tradizionale "intervallo interaccentuale", dove "accento" è da intendersi in termini di concreta realizzazione sintagmatica, anziché come paradigmatica manifestazione dell'accento di parola.

Come si vede, nel modello qui delineato la sillaba e l'accento non sono più considerati come principi regolatori contrapposti, destinati ad orientare in maniera radicalmente alternativa il comportamento ritmico delle lingue, ma piuttosto come proprietà prosodiche di base, necessariamente possedute da ogni lingua. I diversi orientamenti ritmici dovranno dunque dipendere dalla fine "pesatura" dei tratti prosodici, piuttosto che da stridenti difformità strutturali. Beninteso, a queste considerazioni teoriche devono poi far seguito precise indicazioni operative; e in effetti, nel lavoro citato è stato proposto l'impiego di specifici algoritmi, concepiti per catturare le peculiarità ritmiche su ciascuno dei due livelli.

Il modello di Bertinetto & Bertini (2010) mira esplicitamente a conseguire i due seguenti obiettivi: (a) rispettare i criteri epistemologici in precedenza menzionati; (b) render conto in maniera flessibile delle diverse propensioni prosodiche delle lingue, le quali possono talvolta perfino generare esiti apparentemente contraddittori. Nel prossimo paragrafo, questi due obiettivi saranno, sia pur per sommi capi, precisati. Qui ci si limiterà ad un breve cenno per quanto riguarda la possibile contraddittorietà degli esiti, un tema che ha direttamente a che fare con la concezione biplanare del modello. Si pensi, ad esempio, a lingue che presentino una forte complessità sillabica da un lato, e dall'altro lato posseggano un lessico costituito in misura prevalente da polisillabi con accento rigidamente fisso. In base alle caratteristiche prosodiche sopra elencate, da una siffatta situazione si dedurrebbero tendenze contrapposte, da cui potrebbe ingenerarsi una qualche incertezza nella valutazione soggettiva dell'impressione ritmica. Non sarà dunque un mero frutto del caso se il polacco, che esibisce appunto il profilo qui sopra descritto, è stato tradizionalmente considerato una lingua dalle propensioni ritmiche ambivalenti, e pertanto difficilmente classificabile nei termini della dicotomia tradizionale.

2. Il modello “Controllo/Compensazione” (CCI)

2.1 Principi ispiratori

Ispirandosi ai requisiti epistemologici cui si è fatto riferimento (esplicitezza, predittività, “unificazione”), Bertinetto & Bertini (2008) hanno iniziato a sviluppare un modello di analisi del ritmo denominato *Control/Compensation Index* (d’ora innanzi CCI), ulteriormente espanso in direzione biplanare da Bertinetto & Bertini (2010). Poiché nel presente contributo ci si limiterà al livello-I (fonotattico), da qui in avanti ci si restringerà a quest’ordine di considerazioni. È tuttavia necessario tener presente che ciò non esaurisce il discorso, sul quale ci si propone di tornare in altra sede, con l’analisi del comportamento del tedesco a livello-II.

L’idea di fondo della proposta avanzata da Bertinetto & Bertini (2008), in merito al livello fonotattico, è la seguente. Se le lingue si fondassero su una struttura sillabica CV costantemente rispettata, ossia sulla rigorosa alternanza di una consonante ed una vocale, la ritmicità del comportamento articolatorio sarebbe (al livello, appunto, della fonotassi di base) praticamente perfetta, fatte salve le marginali deviazioni temporali dovute alle diverse caratteristiche articolatorie dei vari foni consonantici. Sappiamo infatti che la tendenza alla ritmicità è una proprietà inerente al comportamento degli organismi complessi: lungi dall’essere una capacità da conquistare, essa è la risposta automatica, spontaneamente emergente, che tali organismi mettono in campo per risolvere le complessità inerenti alla difficile coordinazione dei differenti “articolatori” (siano essi le membra di un corpo o, come nel nostro caso, i gesti che sottostanno all’articolazione di foni). La riprova è molto semplice. Per qualunque parlante (fin dalla primissima fase della lallazione) è molto più facile pronunciare una sequenza rigorosamente ritmata del tipo *mamama*, piuttosto che una sequenza analoga in cui si dia a ciascuna ricorrenza di [m] e [a] una diversa durata. Non è ovviamente impossibile ottenere un tale risultato, ma occorre un certo allenamento, mentre la pronuncia ritmata non comporta alcuna fatica per il parlante (a parte, ovviamente, i soggetti affetti da balbuzie).

Poiché tuttavia sappiamo che le lingue – quale più quale meno – si discostano dalla rigorosa alternanza del tipo CV, ne consegue che l’appaiamento dei due oscillatori consonantico e vocalico non può essere regolato in maniera semplice, ma richiede una sapiente strategia da parte del parlante. Coloro che hanno appreso nativamente la pronuncia di una lingua dalla fonotassi complicata hanno evidentemente sviluppato la capacità di far fronte a queste difficoltà. Possiamo dunque asserire che la loro pronuncia possiede il necessario grado di flessibilità per accomodare i diversi foni, caso per caso, nel tessuto sillabico. Per esprimerci in maniera più rigorosa, possiamo riprendere la formulazione proposta da Goldstein, Chitoran & Selkirk (2007), i quali osservano come in una fonotassi semplice gli oscillatori consonantico e voca-

lico abbiano fortissima probabilità di essere perfettamente “in fase”, mentre in una fonotassi complessa si assisterà ad una marcata tendenza alla sovrapposizione dei gesti. In concreto, ciò significa due cose: (i) tendenziale sovrapposizione dei gesti, rispettivamente consonantici o vocalici, quando vi siano consonanti o vocali adiacenti (come, ad es., in una sequenza del tipo CCVVCCCV); (ii) tendenziale sovrapposizione di consonanti e vocali fra di loro. Quest’ultimo è, in particolare, il meccanismo della così detta “riduzione vocalica”, ossia della parziale co-produzione di una vocale non accentata con le consonanti adiacenti, “sotto” le quali, per così dire, essa si nasconde. La porzione di segnale riservata alla vocale può addirittura scomparire visivamente dallo spettrogramma, ma ciò non significa che la vocale sia stata *tout court* cancellata. La cancellazione, beninteso, può avvenire, com’è ben documentato sul piano diacronico, e ciò va anzi considerato come la principale causa dell’affermarsi di una fonotassi complessa. Ma nella fattispecie che qui ci interessa, il gesto vocalico viene “sommerso” dal gesto consonantico, senza che con ciò venga annullato il nucleo sillabico di cui la vocale è normalmente portatrice. È questo il caso delle così dette consonanti “sillabiche”, assai frequenti in tedesco, come nella prevalente pronuncia di *leben* [ˈlebŋ].

È importante notare, tuttavia, che il manifestarsi di una (più o meno marcata) riduzione vocalica non è automaticamente imputabile alla pressione localmente forte della complessità fonotattica. L’esempio appena considerato ne è una dimostrazione, poiché per il parlante tedesco non sarebbe affatto complicato pronunciare la suddetta parola come [ˈlebən], o addirittura [ˈleben]. Il fatto che perfino in questi casi venga invece preferita una pronuncia con fortissima riduzione vocalica ha dunque a che fare con il consolidarsi, nelle abitudini fonatorie dei parlanti, di una spiccata tendenza alla flessibilità coarticolatoria, evidentemente dovuta alla necessità di risolvere, nei casi pertinenti, le difficoltà derivanti dalla fitta successione di foni consonantici. Da ciò si deduce che la tendenza a coarticolare strettamente i foni diventa un abito costante nel parlante di certe lingue, che caratterizza il comportamento complessivo anche al di là dello stretto necessario. Può anche verificarsi, è bene dirlo per chiarezza, qualcosa di diverso: ossia, che l’esistenza di una marcata complessità fonotattica non dia adito, se non in misura fisiologica, al meccanismo della riduzione vocalica. Tale sembra essere la situazione del polacco e dell’islandese. Ci si può dunque attendere che, sul piano del comportamento ritmico, a parità di complessità fonotattica si possano osservare spiccate differenze tra le lingue in ragione della presenza vs. assenza di riduzione vocalica.

Nell’ipotesi sviluppata da Bertinetto & Bertini (2008), si ritiene che le lingue differiscano rispetto al parametro della flessibilità articolatoria, orientandosi verso i due poli contrapposti del “controllo” e della “compensazione”. Si badi, peraltro, che l’uso di due termini in reciproca opposizione non si fonda sul presupposto di una radicale difformità strutturale, ma (come sopra precisato) sulla graduazione del parametro pertinente, da identificarsi nella diversa propensione alla flessibilità articolatoria. Detto in altri termini: una

lingua fortemente “a controllo” è da intendersi come una lingua articolata in maniera relativamente rigida, ossia, con minor capacità, da parte dei parlanti, di coarticolare in maniera disinvolta i foni adiacenti. Ciò garantisce quindi il rispetto del requisito della “unificazione”.

Quanto ai requisiti di esplicitezza e predittività, Bertinetto & Bertini (2008) hanno proposto l’uso di uno specifico algoritmo, denominato *Control/Compensation Index* (CCI), che consiste in pratica nell’adattamento dell’algoritmo *Pairwise Variability Index* (PVI) proposto da Grabe & Low (2002). Quest’ultimo è una misura di tipo dinamico (a differenza per esempio di quella proposta da Ramus et al. 1999), in cui si calcola la media delle differenze fra le durate degli intervalli omogenei adiacenti (consonantici o vocalici). La modifica consiste nel dividere le durate di tali intervalli per il numero dei foni che li costituiscono. Si veda la seguente formula, in cui d indica la durata di un dato intervallo (k è un indice generico di identificazione), n esprime il numero di foni in esso compresi, e la moltiplicazione per 100 mira a ridurre il numero dei decimali:

Formula 1.

Questa misura si basa, com’è subito evidente, su una grossolana semplificazione, in quanto assume che la durata di ciascun fono contenuto in un dato intervallo (consonantico o vocalico) sia esattamente identica, il che è palesemente falso. Tuttavia, è incontestabile che ciascuno di essi contribuisce a determinare la durata complessiva dell’intervallo. Si può pertanto ragionevolmente assumere che tale durata sia funzione del numero di elementi che lo compongono, ancorché non si tratti di una relazione lineare. L’ipotesi di fondo è dunque che nelle lingue a compensazione (caratterizzate da maggior flessibilità coarticulatoria) la compressione dei singoli foni contenuti in un intervallo sia maggiore – soprattutto nelle sillabe non accentate – di quanto accade nelle lingue denominate, per l’appunto, a controllo.

Nei lavori di Bertinetto & Bertini (2008) e Bertinetto & Bertini (2010) si formulava l’iniziale predizione secondo cui, idealmente, nelle lingue strettamente a controllo, i foni consonantici e vocalici dovrebbero contribuire in maniera identica alla dinamica temporale dell’enunciato. Avendo dunque l’indice CCI delle consonanti sull’ordinata e in ascissa quello delle vocali, la proiezione dovrebbe cadere sulla bisettrice del piano cartesiano. Ma si trattava, ovviamente, di una prima, imperfetta formulazione. I risultati via via ottenuti attraverso l’applicazione dell’indice a varie lingue (nonché stili di pronuncia delle medesime) hanno permesso di correggere la prima formulazione. Sembra ora maggiormente conforme al vero asserire (come schematicamente illustrato nella figura 1) che le lingue tipicamente a compensazione debbano presentare una fluttuazione locale delle durate vocaliche più marcata rispetto a quella delle consonanti, e l’inverso si ritiene che debba accadere nelle lingue a controllo. Se infatti la forte fluttuazione nelle durate delle vocali accentate vs. non accentate costituisce un tratto saliente delle lingue a

compensazione (si pensi al citato meccanismo della riduzione vocalica), si può per converso assumere che nelle lingue a controllo la durata delle vocali tenda a subire minori variazioni.

Fig. 1.

Beninteso, ciò presuppone un ragionamento fonologicamente orientato, che porta per esempio ad individuare due unità in una vocale fonologicamente lunga (ossia, con quantità fonologicamente pertinente), a differenza di quanto accade per una vocale che subisca effetti di allungamento puramente contestuali e non fonologicamente pertinenti. L'orientamento fonologico è infatti un tratto distintivo del modello CCI, in ciò profondamente diverso dalle metriche precedenti, da cui pure l'algoritmo sopra riportato è stato ispirato. Ciò comporta una serie di conseguenze. Alcune sono di facile applicazione: per esempio, il fatto di considerare le consonanti geminate, in analogia con le vocali quantitativamente lunghe, come composte di due unità. Altre sono passibili di diversa interpretazione, a seconda di come sia impostata l'analisi dei fatti fonologici. Tra i punti delicati vanno sicuramente annoverate le decisioni da prendersi in rapporto alla scelta fra iato e dittongo, come pure al trattamento come vocale vs. consonante della parte non strettamente nucleare di un dittongo. Le analisi proposte differiscono non solo tra lingua e lingua, ma spesso anche per una medesima lingua, di cui vengono forniti profili differenti a seconda dell'analista.

2.2 Prime applicazioni

L'algoritmo CCI è stato inizialmente applicato ad una selezione di enunciati tratti dal corpus di italiano semispontaneo API (liberamente scaricabile presso <http://www.parlaritaliano.it/index.php/it/corpora/673-corpus-avip-api>). Gli enunciati erano tratti dalle produzioni di cinque coppie di locutori impegnati nel così detto "compito della mappa", in cui uno dei due interlocutori deve guidare l'altro verso una certa meta, inconsapevoli l'uno e l'altro del fatto che le mappe a loro disposizione non coincidono in ogni dettaglio. I criteri di scelta dei materiali sono qui di seguito elencati:

- enunciati che fossero privi di esitazioni, pause, forme di assenso, esclamazioni, false partenze, emissioni vocali non verbali, sequenze inintelligibili, routines discorsive (come le frasi fatte ricorrenti);
- enunciati che avessero, in trascrizione ortografica, almeno 9 sillabe e, foneticamente, almeno 8 sillabe. La differenza è motivata dal fatto che non tutte le potenziali sillabe si realizzano effettivamente nel parlato, a causa di possibili fenomeni di fusione tra vocali adiacenti, riduzione di iato, ipoarticolazione etc.

Dagli enunciati sono state peraltro eliminate:

- le parti terminali di tipo asseverativo (ad es.: ..., *no?*) o quelle introduttive di tipo esplicativo (ad es.: *cioè*,...);
- la parte finale, a partire dall'ultima vocale accentata (compresa), per non incamerare l'effetto dell'allungamento finale, particolarmente condizionante su brevi porzioni di parlato.

I risultati della computazione saranno meglio presentati nella prossima sezione, assieme ai risultati relativi al tedesco spontaneo. Essi sono comunque anticipati nella figura 2 (cf. il punto indicato come SPI), in cui sono raccolti, per opportuno confronto, i dati ottenuti da Mairano e Romano (2008) a partire da enunciati letti.¹ In via prudentemente cautelativa, va considerato che l'applicazione dei criteri fonologici operata da questi due autori potrebbe non essere stata in tutto e per tutto convergente con quella adottata dal nostro gruppo di lavoro; inoltre, va tenuto presente che, a parte il dato dell'islandese (IC) basato sulla media di dieci locutori, ogni altro punto si riferisce ad un solo locutore: due finlandesi (FI) due italiani (IT, di cui viene anche data la media), un francese (FR), un franco-canadese (Can), un inglese britannico (RP), uno statunitense (AM), due tedeschi (Ger). I risultati desumibili dagli studi di Mairano e Romano comunque sembrano confortare le attese che qualunque studioso avrebbe riguardo alle lingue considerate, con l'unica eccezione dell'islandese. Va tuttavia considerato che questa lingua, pur disponendo di una fonotassi germanica, non presenta vistosi fenomeni di riduzione vocalica, il che legittima la sua collocazione nel settore delle lingue a controllo. Si noterà infine come la media dei due locutori italiani analizzati nello stile di lettura sia maggiormente spostata verso il polo del controllo, in confronto con la media dei dieci locutori analizzati nel parlato semi-spontaneo (SPI). Questo è un risultato atteso, dato che, per ipotesi, nel parlato spontaneo possono essere messe in atto strategie coarticulatorie più marcate che nello stile di lettura. Si noterà, peraltro, che gli enunciati letti da un sottoinsieme dei locutori a suo tempo utilizzati per il parlato semi-spontaneo occupano in figura una posizione più distante dalla bisettrice (SPI/letto). Questo è un elemento in controtendenza e potenzialmente delicato. Va tuttavia considerato che esso potrebbe essere condizionato da due fattori. Innanzi tutto, pur essendo stati coinvolti sei locutori, soltanto quattro di essi hanno prodotto un numero minimamente rappresentativo di enunciati. Inoltre, si tratta in gran parte di enunciati contenuti nel corpus semispontaneo e fatti rileggere al medesimo locutore che li aveva prodotti quasi dieci anni prima. Si tratta insomma di un corpus di parlato letto *sui generis*. È possibile infatti che, nel rileggere le frasi a suo tempo prodotte, i locutori

¹ Va segnalato che i medesimi autori hanno successivamente applicato la metrica CCI, così come le altre principali metriche adottate in quest'ambito di studi, ad un numero maggiore di lingue (cf. Mairano & Romano 2010). Si è preferito riportare qui solo una parte dei dati disponibili, per non appesantire inutilmente il discorso.

siano stati indotti ad adottare un comportamento non del tutto naturale; questo dato è quindi meritevole di ulteriore riflessione.

Con queste premesse, si presenteranno ora i dati relativi al tedesco.

Fig. 2.

3. Analisi del tedesco spontaneo

3.1 Selezione dei materiali e loro trattamento

I materiali tedeschi sono stati estratti dal corpus di Kiel, elaborato presso l'*Institut für Phonetik und digitale Sprachverarbeitung* dell'Università di Kiel sotto la direzione del prof. Klaus Kohler. Il Laboratorio di Linguistica possiede, per gentile concessione degli autori, la porzione del corpus intitolata 'Lindenstrasse', che contiene conversazioni libere dove si confrontano due versioni di un medesimo filmato, viste individualmente dai due interlocutori. Gli enunciati sono stati estratti in base agli stessi criteri adottati per l'analogo corpus italiano e si riferiscono a 10 locutori (di entrambi i sessi). Il vantaggio ovvio di poter disporre di un corpus come quello in questione è rappresentato dal fatto che i vari foni sono segmentati, con indicazione della durata rispettiva. Sono stati estratti 231 enunciati, per un totale di 2761 intervalli consonantici e 2751 intervalli vocalici. Per confronto, ecco i dati dell'italiano semispontaneo: 231 enunciati, con 2444 intervalli consonantici e 2796 intervalli vocalici.

Il tedesco costituisce una sfida particolarmente impegnativa per l'applicazione del modello CCI, date le sue proprietà fonetico-fonologiche. La questione ha vari aspetti. Occorre innanzi tutto affrontare quella parte del problema che, per quanto maggiormente vistosa in talune lingue piuttosto che in altre, costituisce un fattore ineliminabile in ciascuna. Si allude qui alle ipoarticolazioni tipiche del parlato spontaneo, che consistono non di rado in cancellazioni di foni, la cui identità è peraltro facilmente ricostruibile dall'uditore sulla base della ridondanza semantico-contestuale. Nello spirito del modello CCI, orientato in senso fonologico piuttosto che fonetico, un fono accidentalmente cancellato andrà comunque contato, in quanto lo si può ritenere presente nel programma articolatorio del parlante. La sua cancellazione può essere paragonata all'espulsione di un elemento dovuta al ridursi dello "spazio" disponibile, ossia alla compressione dell'intervallo (soprattutto, in questi casi, consonantico) causata da una elocuzione accelerata. Bisogna peraltro fare attenzione a non confondere le cancellazioni irregolari ed estemporanee con le cancellazioni codificate nella particolare varietà di lingua considerata. Per fare un esempio, in italiano non andrà considerata come cancellazione irregolare la pronuncia di *sto* in luogo di *questo*; si tratta infatti di un allomorfo perfettamente ammesso, com'è dimostrato dal fatto

che esso può comparire, negli stili discorsivi appropriati, anche a velocità di elocuzione non elevata. Diverso sarebbe il caso di *finesta* per *finestra*, che di certo non appartiene ad alcuna varietà diafasica dell'italiano e può unicamente intendersi come lapsus contingente.

La varietà di tedesco parlata dai locutori del corpus è di tipo settentrionale. Per verificare la natura delle semplificazioni riscontrate, si è fatto ricorso alla consulenza della collega Marianne Hepp, che qui si ringrazia. Si sono dunque considerate cancellazioni “regolari” quelle del tipo: *bringe* → *bring'*, *habe* → *hab'*, *eine* → *'ne*, *es* → *'s* (ad esempio, *gibt's*), *Reihenfolge* → *Reih'nfolge*, *verstehen* → *versteh'n*, etc. Per converso, sono state considerate irregolari cancellazioni quali: *Abendrot* → *Amrot*, *Andenkenfoto* → *Ande[ŋ]enfoto*, *anscheinend* → *anscheinen*, *auch* → *au*, *das* → *as*, *eigentlich* → *eī[ŋ]kich*, *einmal* → *eima*, *ist* → *is/i/s*, *Szene* → *Sene*, *kommt* → *komm*, *man* → *a*, *nicht* → *nich*, *sollten* → *son*, *und* → *un* etc.

Un aspetto particolarmente delicato della fonetica-fonologia del tedesco è costituito dalle consonanti, rispettivamente, sillabiche e vocalizzate. Si allude qui a parole come *leben* o, rispettivamente, *lernbar*. Nel primo caso, come sopra notato, la vocale scompare in apparenza, lasciando come traccia la sillabicità della sonorante seguente (cf. [ˈlebŋ]); nel secondo, a scomparire è la vibrante finale di sillaba, che lascia come residuo la dittongazione della vocale precedente (una possibile pronuncia è [ˈlɛɐ̯nbə]). I due fenomeni sono, in un certo senso, simmetrici, dato che nell'uno si ha la fusione di una vocale con la consonante seguente, nell'altro di una consonante con la vocale precedente. Nell'ottica fonologica del modello CCI, non c'è peraltro perdita di fonemi, neppure nel caso delle consonanti sillabiche, le quali vanno interpretate come un episodio estremo di co-produzione fra due foni adiacenti (nella fattispecie: la sonorante sillabica e la vocale “nascosta” che ne giustifica la sillabicità). Resta tuttavia il problema di rappresentare questo fenomeno nel computo dell'algoritmo. La soluzione “fonologica” consisterà nell'assumere le seguenti rappresentazioni, dove il simbolo \emptyset in pedice indica la durata “zero” dell'elemento considerato:

- consonanti sillabiche: es. *leben* = C V C V_∅ C
- consonanti vocalizzate: es. *lernbar* = C V C_∅ C C V C_∅

Dato questo trattamento, non sarà difficile prevedere che l'impatto delle consonanti sillabiche sarà molto maggiore di quello delle consonanti vocalizzate. Queste ultime, infatti, tendono per lo più a diventare un elemento frazionario all'interno dell'intervallo consonantico di cui fanno parte, mentre con le prime si deve computare una vocale che costituisce di per sé sola un intervallo. Con le consonanti vocalizzate si ha pertanto, come unico effetto, una riduzione della durata delle unità che costituiscono l'intervallo consonantico; con le consonanti sillabiche si ha invece un intervallo vocalico di durata “zero” che, interagendo colle durate degli intervalli vocalici adiacenti, produrrà effetti molto più marcati.

A parte questi casi, in un certo senso speciali, la fonologia del tedesco presenta altri fenomeni meritevoli di attenzione:

- Occlusione glottidale: l'attacco di una vocale è spesso preceduto da un'occlusione glottidale perfettamente misurabile. Essa può comparire anche all'interno di parola per marcare il confine tra due vocali in iato. Poiché non riveste valore fonologico, la durata dell'occlusione è stata assegnata interamente alla vocale quand'essa è preceduta da una consonante, o in parti uguali alle due vocali adiacenti in caso di iato. Si potrebbe osservare che l'occlusione serve a ricostituire la naturale tendenza alla struttura CV, ma ciò ovviamente potrebbe valere solo nel caso dello iato; in ogni caso, tali foni non hanno statuto fonemico.
- Altrettanto visibile sullo spettrogramma è la porzione di aspirazione che segue le occlusive forti. Benché tali porzioni di segnale siano segmentate nel corpus di Kiel, la corrispondente durata è stata assegnata alla consonante precedente. In questo, come nel precedente caso, non è stata aggiunta alcuna unità nel computo dei fonemi entro gli intervalli di appartenenza.
- Analogamente, sono state considerate come un unico segmento le due porzioni di segnale corrispondenti ad un'affricata (cf. [pf ts tʃ]).
- Per converso le assimilazioni del tipo [nd] → [nn] sono state conteggiate come costituite da due segmenti – benché il segnale risultante non avesse le caratteristiche di un'autentica geminata – in ossequio all'intenzione fonemica del parlante.
- Il fonema /ŋ/ è stato trattato come singola unità, a dispetto della sua realizzazione ortografica <ng>.
- Lo stesso è stato fatto per le vocali lunghe, che tali sono per ragioni fonotattiche, anziché per pertinenza fonologica.
- I dittonghi sono stati trattati in diversa maniera a seconda che fossero ascendenti o discendenti. In ciò si è adottato il punto di vista del corpus Kiel, che segmenta regolarmente il legamento /j/ in attacco di dittongo e lascia invece indivisi gli altri dittonghi. È utile ricordare che per l'italiano è stato adottando un diverso trattamento (spiegato in dettaglio in Bertini & Bertinetto 2009), con la segmentazione dei legamenti /j w/ sia ascendenti che discendenti. Per facilitare tuttavia il confronto fra le due lingue, nella fig. 3, i dati dell'italiano sono riportati secondo entrambe le strategie di segmentazione dei dittonghi ("italiana" e "tedesca").

Naturalmente, ciascuna delle scelte sopra esposte è passibile di critica. Trattandosi del primo tentativo di applicazione del modello CCI ad una lingua con queste caratteristiche, non ci si può aspettare che tutti i problemi possano essere risolti in maniera semplice. D'altra parte, lo scopo di questo contributo è precisamente quello di affinare l'analisi e comprenderne gli eventuali limiti di impostazione. Come si vedrà nel seguito, si è comunque agito con doverosa prudenza: sui punti maggiormente critici, l'analisi è stata condotta secondo strategie alternative, contrapponendo un trattamento fonetico al trattamento fonologico sopra descritto. Per esempio, nel caso delle consonanti sillabiche, è stato anche adottato, per confronto, il trattamento

consistente nel trascurare del tutto la V_0 e nel collocare le due consonanti ai suoi lati entro il medesimo intervallo.

Si è del resto derogato sistematicamente dal trattamento fonologico nei rarissimi casi di inserzione indebita di consonanti, frutto di occasionale lapsus. Tali consonanti sono state conteggiate tra le unità dell'intervallo di riferimento, pur non facendo parte dell'intenzione fonematica del parlante.

3.2 Risultati

Prima di riportare i risultati dell'analisi secondo l'algoritmo CCI, occorre sottolineare un dato rilevante. La velocità media dei locutori tedeschi del corpus utilizzato è risultata particolarmente elevata. Calcolata in segmenti per secondo (e, beninteso, conteggiando esclusivamente i segmenti effettivamente prodotti, senza tener conto di quelli irregolarmente cancellati), essa è stata di 17,5 segmenti/sec. Per utile confronto, si consideri che i locutori italiani hanno denunciato una velocità media di 16,0 segmenti/sec. Ciò non è privo di conseguenze. È infatti noto da tempo (cf. per esempio Dellwo 2006) che la velocità di elocuzione produce un sensibile effetto sui risultati, qualunque sia la metrica impiegata, a meno che non si usino procedimenti di normalizzazione che ne riducano l'impatto.

La fig. 3 mostra la proiezione sul piano cartesiano del tedesco (SG) e dell'italiano (SPI/onglide), nonché le proiezioni ottenibili mediante la selezione di tre porzioni di corpus, equamente distribuite – per ciascuna lingua – quanto a numero di enunciati: dai più lenti (indice 1) ai più veloci (indice 3). L'abbreviazione "SPI/onglide" allude al fatto che i dittonghi sono stati trattati allo stesso modo dei dittonghi tedeschi, considerando i legamenti ascendenti come consonantici e quelli discendenti come vocalici. Le velocità medie in segmenti/sec delle varie selezioni sono risultate le seguenti:

- TEDESCO: velocità 1: 14,7 (\leq 16,4); velocità 2: 17,5 ($>$ 16,4; $<$ 18,5); velocità 3: 20,2 (\geq 18,5);
- ITALIANO: velocità 1: 13,4 (\leq 14,7); velocità 2: 15,7 ($>$ 14,7; $<$ 16,8); velocità 3: 18,7 (\geq 16,8).

Contrariamente a quanto ci si potrebbe aspettare – anche sulla base dei dati in fig. 2 – il tedesco spontaneo appare maggiormente spostato, rispetto all'italiano semispontaneo, verso il territorio delle lingue a controllo. È possibile che i materiali tedeschi contengano qualche disomogeneità, come mostra l'andamento parzialmente ondivago (in confronto all'italiano) delle diverse selezioni di velocità, con la velocità 2 maggiormente spostata verso la bisettrice rispetto alla velocità elevata. Ciò potrebbe dipendere da un'anomala presenza di locutori caratterizzati da un comportamento "meno prototipico", soprattutto (come in effetti si è potuto constatare nella fattispecie) fra coloro che hanno prevalentemente prodotto enunciati a velocità 2. Ma ciò non toglie che il risultato complessivo sia piuttosto sorprendente.

Fig. 3.

Va tuttavia considerato che la velocità media del corpus tedesco è superiore a quella del corpus italiano. Per rendere più solido il confronto, si è selezionato un sottoinsieme del corpus italiano (cf. “SPI/17,4” in fig. 3) caratterizzato da una velocità media di 17,4 segmenti/sec., praticamente identica a quella dell’intero corpus tedesco (17,5). Significativamente, la proiezione si è molto avvicinata a quella del tedesco. Ciò sembra indicare che il comportamento ritmico delle lingue, al crescere della velocità, tende a convergere. Per quanta prudenza si voglia usare in merito a questo dato, che si basa sull’osservazione di due sole lingue, non si può negare che tale conclusione è in sé dotata di immediata plausibilità.

La figura 3 riporta anche la proiezione del corpus italiano in cui i legamenti dei dittonghi discendenti sono trattati – diversamente dal corpus tedesco – come segmenti consonantici (SPI). Come si può constatare, questa variante non altera il risultato in maniera significativa. Ma, beninteso, il fatto che ciò accada in questo specifico caso non autorizza a ritenere che il diverso trattamento dei dittonghi generi sempre effetti marginali: dipenderà, evidentemente, dal peso (soprattutto quantitativo) che i dittonghi rivestono nella lingua considerata.

È necessario, a questo punto, interrogarsi sulle conseguenze derivanti dalle scelte “fonologiche” illustrate nel paragrafo precedente. La fig. 4 mostra, oltre alla proiezione del tedesco trattato fonologicamente (con SG = “Standard German”), altre quattro proiezioni in vario modo foneticamente orientate, a seconda che: (i) si escludano dal computo delle unità fonematiche le “deviazioni”, ossia i segmenti cancellati irregolarmente (SG-Abw, che sta per *Abweichungen*); (ii) si escludano le C \emptyset (SG-C \emptyset); (iii) si considerino le C \emptyset come elementi pienamente vocalici, considerando quindi l’intervallo vocalico precedente come costituito da due segmenti (SG-C \emptyset ,V+V); o, infine, (iv) si escludano le V \emptyset (SG-V \emptyset). Come si può constatare, le prime due modifiche producono conseguenze piuttosto marginali, mentre le ultime due creano un sensibilissimo spostamento verso la bisettrice. Dunque, non ogni intervento fonologicamente orientato produce effetti particolarmente sensibili: dipende, evidentemente, da quale aspetto della fonotassi viene toccato. Quale sia il modo migliore per catturare il comportamento ritmico di una lingua che possieda le caratteristiche fonotattiche del tedesco resta ovviamente, allo stato attuale delle analisi, *sub iudice*. Nel paragrafo finale il problema sarà brevemente discusso.

Fig. 4.

4. Conclusioni provvisorie di lavori in corso

I risultati presentati nel paragrafo precedente contengono due aspetti critici, meritevoli di approfondimento.

Il primo riguarda la scelta del modello descrittivo da adottarsi per lingue come il tedesco, ricche di fenomeni fonologici particolarmente invasivi sul piano fonotattico. Benché la nostra preferenza si indirizzi nettamente verso il trattamento qui etichettato come “fonologico”, in quanto meglio corrispondente alla concezione del modello CCI, si è avuta la precauzione di verificare, caso per caso, l’effetto delle scelte adottate. Come si è visto, l’impatto delle $V\emptyset$ (nonché del trattamento vocalico delle $C\emptyset$) produce conseguenze molto più vistose di quanto non accada con le $C\emptyset$. Ciò è interessante per due ragioni. Innanzi tutto, perché i due fenomeni ($C\emptyset$ e $V\emptyset$) sembrerebbero apparentemente dello stesso tipo. In secondo luogo, perché le $C\emptyset$ sono in numero assai più elevato delle $V\emptyset$ (441 contro 171). Ciò che maggiormente incide sui risultati non è dunque la frequenza del fenomeno considerato, ma il suo specifico impatto sulla struttura fonotattica, almeno in rapporto al modo con cui l’algoritmo CCI computa i dati. A conferma, si consideri che le deviazioni (Abw), che parimenti producono variazioni marginali, sono 387. Da ciò discende, come ulteriore spunto di riflessione teorica, la possibile opzione di predisporre algoritmi meno fortemente condizionati dalle differenze di durata fra coppie di intervalli adiacenti. Il gruppo di lavoro del Laboratorio di Linguistica della SNS ha allo studio possibili raffinamenti del metodo di computo.

Il secondo aspetto critico concerne la possibile accusa di circolarità che si potrebbe rivolgere alla nostra impostazione. Si è infatti asserito, nel capoverso precedente, che il trattamento “fonologico” è quello che meglio incarna lo spirito del modello CCI. Ma poiché esso è anche tra quelli che producono risultati maggiormente compatibili con le attese – che presupporrebbero per il tedesco una proiezione decisamente a destra della bisettrice – l’asserzione di cui sopra rischia appunto di esporsi ad una tale obiezione. Ad essa si può rispondere in due modi. Innanzi tutto, ribadendo che la preferenza per il trattamento “fonologico” è asserita indipendentemente dal comportamento della singola lingua esaminata, in quanto costituisce un tratto costitutivo del modello. In secondo luogo, osservando che di circolarità si potrebbe a buon diritto parlare se la preferenza fosse qui accordata alla proiezione generata dal trattamento $C\emptyset$, che (come si vede nella fig. 4) si colloca a maggior distanza dalla bisettrice rispetto al trattamento “fonologico”. Tale non è, tuttavia, la posizione assunta dai presenti autori.

Ovviamente, va considerato che i dati qui presentati sono fortemente condizionati dal tipo di materiali utilizzati. Il corpus di Kiel ci offre materiali spontanei prodotti a velocità alquanto elevata. Le considerazioni svolte nel paragrafo precedente, come pure le osservazioni contenute in precedenti lavori di questo gruppo di ricerca, suggeriscono che, al crescere della velocità

di elocuzione, i comportamenti ritmici tendano a convergere. Ciò dipende dal minor grado di libertà nell'articolazione dei singoli foni, in quanto soggetti a forte compressione temporale. Non deve quindi destare sorpresa il fatto che i dati qui presentati generino una proiezione che punta verso il territorio prevedibilmente assegnabile alle lingue a controllo. Significativo appare infatti il diverso comportamento dei due locutori tedeschi analizzati da Mairano & Romano e riportati in fig. 2; i quali, leggendo, debbono verosimilmente aver prodotto campioni di parlato a velocità assai più contenuta. Il nostro gruppo di ricerca si propone di estendere l'analisi ad ulteriori materiali tedeschi prodotti con stile di lettura, al fine di verificare l'attendibilità complessiva delle osservazioni riportate in questo contributo.

Non va infine dimenticato che il modello CCI poggia su un'architettura biplanare. Tutte le osservazioni contenute nel presente lavoro si riferiscono, come in precedenza dichiarato, al livello-I, quello della fonotassi. Nulla è stato qui detto in rapporto al livello-II, relativo alla distribuzione delle prominenze di frase. Anche a questo riguardo, ci si propone di ampliare l'analisi per completare il confronto fra italiano e tedesco. Come si vede, le cose da fare sono ancora molte, ma per buona fortuna incrementalmente modulabili. E la pazienza non ci manca.

Bibliografia

- ALLEN, George D. [1975]. "Speech rhythm: its relation to performance universals and articulatory timing". *Journal of Phonetics* 3: 75-86.
- BERTINETTO, Pier Marco (ed.) (1981): *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici*. Firenze: Accademia della Crusca.
- BERTINETTO, Pier Marco/BERTINI, Chiara (2008): "On modeling the rhythm of natural languages". In: BARBOSA, Plínio Almeida/MADUREIRA, Sandra/REIS, Cesar (eds.): *Proceedings of the Fourth Conference on Speech Prosody*, 427-430. University of Campinas (Brazil). ISCA Archive, <http://www.isca-speech.org/archive/sp2008>.
- BERTINETTO, Pier Marco/BERTINI, Chiara (2010): "Towards a unified predictive model of Natural Language Rhythm". In RUSSO, Michela (ed.): *Prosodic Universals. Comparative studies in rhythmic modeling and rhythm typology*. Rome: Aracne, 43-77.
- BERTINI, Chiara/BERTINETTO, Pier Marco (2009): "Prospezioni sulla struttura ritmica dell'italiano basate sul corpus semispontaneo AVIP/API". In ROMITO, Luciano/GALATÀ, Vincenzo/LIO, Rosita (eds.): *Atti del 4° Convegno Nazionale AISV 2007. La fonetica sperimentale - Metodo e Applicazioni*. Università della Calabria. Torriana (RN): EDK Editore.
- BERTINI, Chiara/TARANTO, Matias/BERTINETTO, Pier Marco (in press): "Rhythmic Index Elaborator (RIE) come strumento di indagine della struttura ritmica. Un'applicazione al pisano semi-spontaneo vs. letto." ". In: *Atti del 7° Convegno Nazionale AISV 2007. Contesto comunicativo e variabilità nella produzione e percezione della lingua*. Università del Salento, Lecce.

- CLASSE, Andre (1939): *The rhythm of english prose*. Oxford : Basil Blackwell
- DAUER, Rebecca M. (1983). "Stress-timing and syllable-timing reanalyzed". In: *Journal of Phonetics* 11, 51-62.
- DELLWO, Volker (2006). "Rhythm and speech rate: a variation coefficient for deltaC", In: KARNOWSKI, P. / SZIGETI, I. (eds.): *Language and language processing*, 231-241. Frankfurt: Peter Lang.
- GOLDSTEIN, Louis/CHITORAN, Ioana/SELKIRK, Elisabeth (2007): "Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlihyt Berber". In: TROUVAIN, Jürgen/BARRY, William J. (eds.): *Proceedings of the 16th Int. Congress of Phonetic Sciences*, 241-244. Universität Saarbrücken. Dudweiler: Pirrot GmbH.
- GRABE, Esther/LOW, Ee Ling (2002). "Durational variability in speech and the rhythm class hypothesis". In: *Papers in Laboratory Phonology* 7, 515-546. Berlin: Mouton de Gruyter.
- LEHISTE, Ilse [1973]. "Rhythmic units and syntactic units in production and perception", *Journal of the Acoustical Society of America* 54: 1228-34.
- LINDBLOM, Björn E.F./RAPP, Karin (1973): "Some Temporal Regularities of Spoken Swedish". In: *Papers of the Institute of Linguistics*, 21. University of Stockholm.
- MAIRANO, Paolo/ROMANO, Antonio (2008): *A comparison of four rhythm metrics for six languages*. Poster presented at the conference on Empirical Approaches to Speech Rhythm. University College London.
- MAIRANO, Paolo/ROMANO, Antonio (2010): "Un confronto tra diverse metriche ritmiche usando Correlaore". In: SCHMID, Stephan/SCHWARZENBACH, Michael/STUDER, Dieter (eds.): *Atti del 5° Convegno Nazionale AISV 2009. La dimensione temporale del parlato*. Universität Zürich. Torriana (RN): EDK Editore.
- RAMUS, Franck/NESPOR, Marina/MEHLER, Jacques (1999): "Correlates of linguistic rhythm in the speech signal". In: *Cognition* 73, 265-292.

$$CCI = \frac{100}{m-1} \sum_{k=1}^{m-1} \left| \frac{d_k}{n_k} - \frac{d_{k+1}}{n_{k+1}} \right|$$

Formula 1.

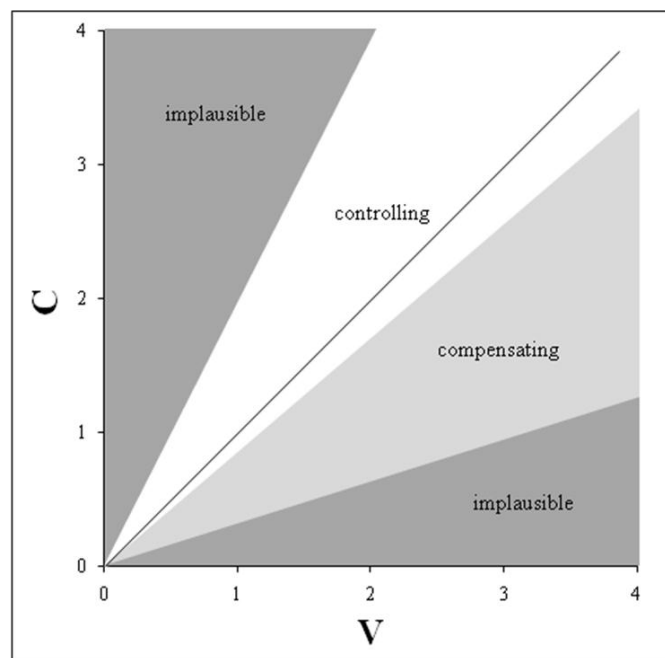


Fig. 1. Rappresentazione schematica delle tendenze ritmiche

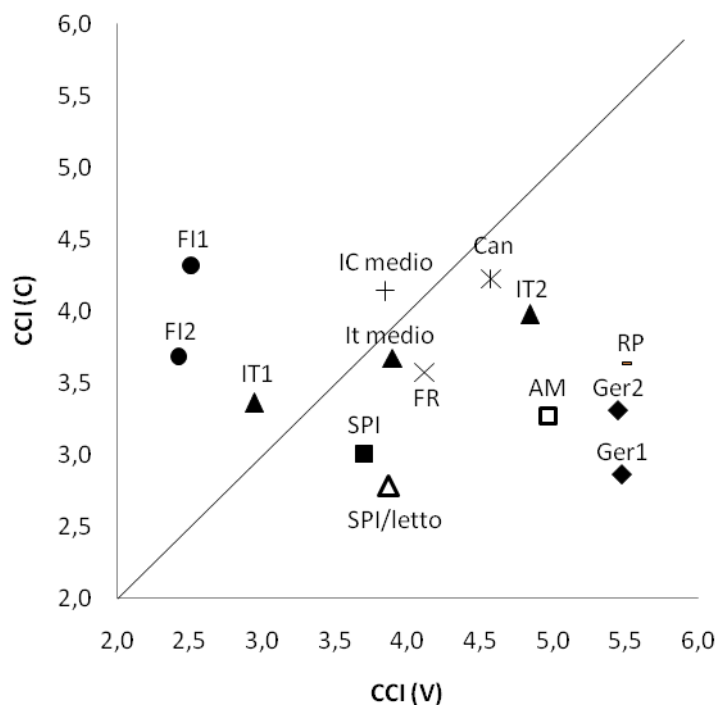


Fig. 2. Dati ricavati dalle misure sul parlato letto di Mairano & Romano 2008, con l'aggiunta dei dati riferiti al corpus di italiano pisano spontaneo (SPI) e italiano pisano letto (SPI/letto). Per le altre abbreviazioni, cf. il testo.

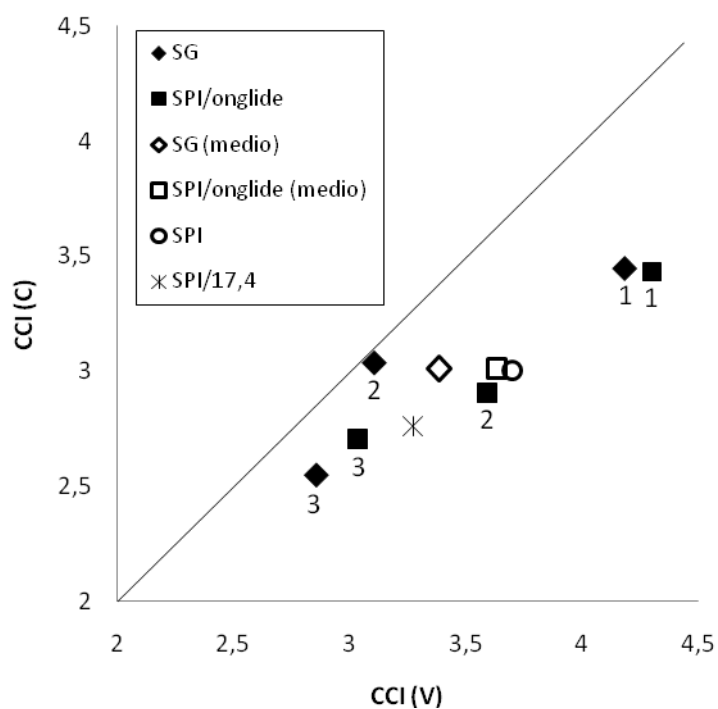


Fig. 3. Proiezioni del tedesco spontaneo (SG) e dell'italiano semispontaneo prodotto da locutori pisani (SPI/onglide). Gli indici numerici da 1 a 3 indicano sottoinsiemi dei due corpora caratterizzati da velocità crescente, mentre il rombo e il quadrato vuoti indicano il valore medio. Il cerchio vuoto (SPI) indica la proiezione dell'italiano pisano semispontaneo in cui anche i legamenti postnucleari sono trattati come segmenti consonantici. SPI/17,4 si riferisce alla porzione del corpus italiano con velocità media paragonabile a quella del corpus tedesco.

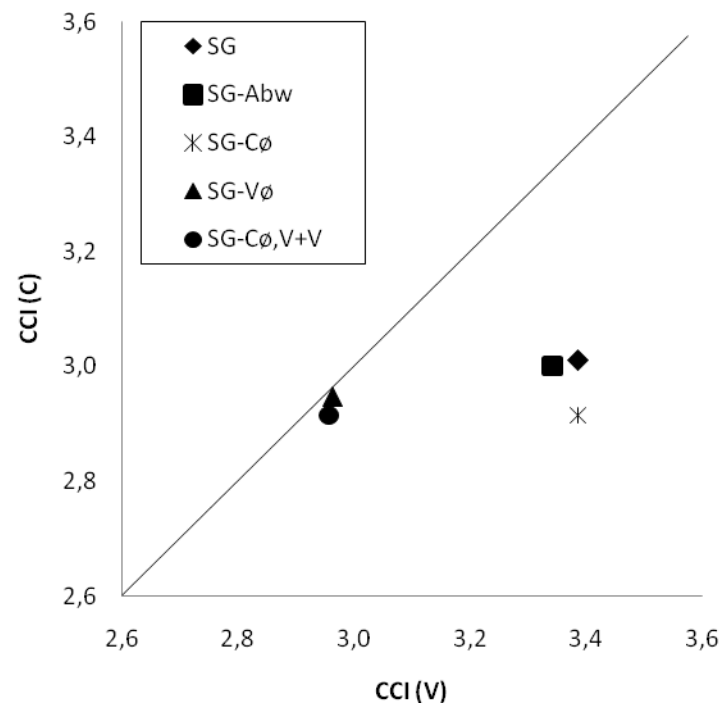


Fig. 4. Proiezioni del tedesco spontaneo: con trattamento “fonologico” (SG) e con quattro trattamenti foneticamente orientati (cf. il testo per i dettagli).