



PROSPEZIONI SULLA STRUTTURA RITMICA DELL'ITALIANO BASATE SUL CORPUS SEMISPONTANEO AVIP/API

Chiara Bertini & Pier Marco Bertinetto

Comunicazione presentata al 4° Convegno AISV, Università della Calabria, 2007)

RIASSUNTO

In questo lavoro, presentiamo una ricerca condotta sui dati contenuti nella banca-dati AVIP/API, con specifico riguardo alla variante pisana dell'italiano. Trattandosi di materiali semispontanei, si è dovuto preliminarmente affrontare una serie di problemi per ciò che riguarda la selezione dei materiali, al fine di basare le osservazioni su porzioni di enunciato sufficientemente rappresentative.

Le analisi hanno fornito indicazioni interessanti e non sempre scontate. Se ne fornisce qui un primo elenco:

- la porzione finale dell'enunciato, dall'ultima sillaba accentata in poi, comporta una sensibile alterazione dell'andamento ritmico;
- un effetto relativamente modesto è invece legato alle momentanee sovrapposizioni dei due locutori;
- non si è notata alcuna differenza significativa per quanto riguarda il contrasto tra enunciati dichiarativi vs. interrogativi;
- è emerso un sostanziale contrasto fra "Giver" e "Follower", fortemente correlato, peraltro, con la diversa velocità di elocuzione (decisamente più lenti i "Givers");
- la velocità di elocuzione ha tangibili conseguenze sulla configurazione ritmica.

Nel condurre l'analisi, si sono sperimentati diversi algoritmi di calcolo, tra cui uno ideato dagli autori, denominato "Control/Compensation Index", il cui comportamento appare decisamente promettente.

1. INTRODUZIONE

Gli studi sull'indole ritmica delle lingue naturali hanno ricevuto nuovo impulso in questi ultimi anni, a partire dalla proposta di Ramus (Ramus, 1999), presto ripresa da altri studiosi e segnatamente da Low (Low *et alii*, 2000) e Grabe & Low (Grabe & Low, 2002). Non sono ovviamente mancate le applicazioni all'italiano e ai suoi dialetti (Russo & Barry, 2004; Schmid, 2004; Romano *et alii*, 2006). La discussione si è soprattutto concentrata sulla ricerca di un algoritmo capace di rappresentare in maniera efficace le proprietà ritmiche delle lingue naturali, traducendo in termini concreti l'intuizione di Pike (Pike, 1947) circa la polarizzazione tra lingue iso-accentuali e iso-sillabiche (syllable- vs. stress-timed). Benché gli studi successivi abbiano ampiamente chiarito che l'interpretazione letterale di tale dicotomia non poteva trovare conferma, è tuttora opinione diffusa che l'intuizione di partenza sia sostanzialmente corretta. Non sono del resto mancati i tentativi di reinterpretarla in chiave nuova, correlando un insieme di comportamenti fonetici con un insieme di proprietà fonologiche (Bertinetto, 1981; Dauer, 1983; Bertinetto, 1989). Tra i fattori rilevanti, sono stati individuati almeno i seguenti:

- riduzione vocalica vs. piena articolazione nelle sillabe atone;
- struttura sillabica complessa (con relativa incertezza dei confini sillabici) vs. semplice (con confini ben definiti);
- relativa flessibilità vs. rigidità nella collocazione dell'accento lessicale;
- densità di accenti secondari vs. relativa rarefazione;
- accelerazioni di velocità di eloquio ottenute (per lo più) mediante la compressione delle sillabe atone vs. compressione relativamente proporzionale tra sillabe toniche ed atone.

È stata successivamente avanzata l'ipotesi che questo fascio di tratti potesse trovare un punto di convergenza nel contrasto fra lingue "a controllo" e lingue "a compensazione" (Bertinetto & Vékás, 1991).¹ L'idea di base, già accennata in Bertinetto & Fowler (Bertinetto & Fowler, 1989), è la seguente. Si assume che le lingue "a controllo" siano basate su una struttura fonotattica semplice, che induce il parlante ad assumere un'impostazione articolatoria relativamente rigida, mentre il contrario avviene nelle lingue "a compensazione", fondate su una fonotassi complessa ed un'impostazione articolatoria più flessibile. Se l'idea è corretta, essa dovrebbe comportare una serie di conseguenze sul piano dei comportamenti articolatori e prosodici. Si dovrebbero in sostanza verificare le seguenti condizioni:

- Sul piano articolatorio, le lingue "a compensazione" dovrebbero comportare meccanismi di coarticolazione più spinti rispetto a quelli che agiscono nelle lingue "a controllo", il che si traduce in una maggior tendenza alla co-produzione dei gesti articolatori, con specifico riguardo alla sovrapposizione dei gesti consonantici e vocalici.

- Sul piano prosodico, le durate dei foni dovrebbero essere diversamente regolate: in particolare, nelle lingue "a controllo" i foni dovrebbero manifestare una certa dose di incomprimibilità anche al crescere della velocità di elocuzione. Ciò non significa, beninteso, che tali lingue non possano tollerare accelerazioni di eloquio, bensì che il grado di sovrapposizione tra gesti articolatori adiacenti non possa superare un certo limite, soprattutto per quanto riguarda i foni vocalici (si veda il punto precedente). Di conseguenza, questi ultimi dovrebbero tendere a conservare una corposa "visibilità" anche in sillaba atona. Per contro, nelle lingue "a compensazione" le vocali atone dovrebbero poter tollerare livelli anche molto elevati di coarticolazione, fino a dare l'impressione – nei casi estremi – di scomparire dallo spettrogramma in quanto porzione autonoma di segnale. Non va peraltro dimenticato che le porzioni indipendentemente misurabili dei foni, così come esse ci appaiono sullo spettrogramma, non sono che la parte visibile di un meccanismo estremamente complesso, fatto di continue sovrapposizioni tra gesti articolatori.

Se queste ipotesi colgono nel segno, ne consegue che l'impostazione adottata nelle recenti ricerche sul tema dell'isocronia contiene un forte elemento di criticità. Tanto nel modello di Ramus, quanto nel modello di Grabe & Low (nonché nelle varianti cui essi hanno dato luogo), viene analizzato il comportamento degli intervalli vocalici e consonantici senza riguardo ai segmenti che li compongono. Ciò impedisce di distinguere tra situazioni fonotattiche assai diverse, che potrebbero concretamente manifestarsi nelle lingue naturali. Per esempio: un medesimo intervallo potrebbe essere occupato da due consonanti oppure da quattro (e se abbiamo a che fare con il polacco, fino a non meno di sei). Trattare tale intervallo consonantico come un tutto indifferenziato significa perdere informazioni importanti, le quali potrebbero raccontarci molte cose circa le strategie articolatorie adottate dai parlanti e, di conseguenza, circa i riflessi che tali strategie possono avere sui comportamenti prosodici. Lo stesso dicasi per gli intervalli vocalici: una sequenza di vocali in iato non è la stessa cosa di una sequenza di vocali in sinalefe. Nel secondo caso, i differenti gesti articolatori sono fusi entro un medesimo nucleo sillabico, mentre in caso di iato i nuclei sillabici restano distinti. Se dunque ammettiamo che tutto ciò possa contare nel

¹ La terminologia è ispirata a Hoqvist (Hoqvist, 1983), anche se reinterpretata in accezione nuova.

dispiegarsi del ritmo delle lingue naturali, è importante trattare diversamente queste due situazioni.

Tenendo conto di ciò, è stato proposto dagli autori di questo lavoro il modello illustrato in Bertinetto & Bertini (2008), denominato ‘Control/Compensation Index’ (CCI). Esso si basa sul seguente algoritmo, dove m sta per ‘numero degli intervalli’ (vocalici o consonantici), d per ‘durata’, n per ‘numero di segmenti entro gli intervalli pertinenti’:

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

L’idea di prendere in considerazione il numero dei segmenti che costituiscono gli intervalli non è del tutto nuova. Essa è stata messa in pratica da Rouas & Farinas (Rouas & Farinas, 2004), sia pure limitatamente agli intervalli consonantici. Vi sono tuttavia importanti differenze tra la loro impostazione e quella del modello CCI, su cui ritorneremo nel § 3.

Ovviamente, la bontà di un metodo concepito per misurare le tendenze ritmiche delle lingue naturali dipende crucialmente da due elementi:

- (i) la capacità di far emergere la correlazione tra gli orientamenti ritmici e il diverso assetto articolatorio adottato dalle lingue in rapporto alla propria struttura fonotattica;
- (ii) la capacità di generare i raggruppamenti corretti delle lingue indiziate di appartenere prototipicamente all’uno o all’altro tipo ideale, comunque li si voglia denominare (la dicotomia “controllo vs. compensazione” ci appare, peraltro, assai più persuasiva di quella tradizionale, “syllable- vs. stress-timing”, notoriamente priva di riscontri effettivi).

Quanto al secondo aspetto, risultati molto incoraggianti possono già essere citati a partire da uno studio di Mairano & Romano (in preparazione), in cui sono state confrontate, attraverso l’applicazione dell’algoritmo CCI, otto diverse lingue, studiate a partire dalla lettura di brevi testi. Gli esiti della computazione hanno prodotto una separazione delle lingue analizzate più coerente rispetto a quella prodotta dagli altri modelli correntemente impiegati. Ciò dimostra che i presupposti su cui si fonda la nostra proposta sono alquanto solidi. La direzione intrapresa merita dunque di essere proseguita.

I dati che qui verranno illustrati, tuttavia, si riferiscono tuttavia unicamente all’analisi condotta su materiali semi-spontanei di italiano pisano, estratti dal corpus AVIP/API.² Ci proponiamo beninteso di estendere in futuro l’analisi, confrontando i risultati ottenuti per l’italiano con quelli di altre lingue, studiate anch’esse a partire da materiali semi-spontanei, rigorosamente omogenei a quelli del nostro corpus. È infatti ben noto che lo stile di elocuzione (tipicamente: parlato letto vs. parlato spontaneo, diverse velocità di elocuzione etc.) produce effetti macroscopici a livello di comportamenti ritmici. Lo scopo di questa comunicazione non è comunque quello di confrontare l’italiano con altre lingue, bensì di verificare le diverse tendenze che possono emergere all’interno di un medesimo tipo di testo, confrontando locutori diversi e diversi ruoli discorsivi, sulla base di un corpus debitamente ampio. Quest’ultimo fatto permetterà, tra l’altro, di rispondere ad un’obiezione che si può muovere alle ricerche condotte nel filone inaugurato dall’avvento del modello di Ramus: ossia, la ridotta ampiezza della base di dati su cui si fondano molte delle classificazioni proposte. Il numero di intervalli vocalici e consonantici considerati è infatti spesso piuttosto ristretto, con ovvie conseguenze sulla rappresentatività statistica, come mostra il seguente prospetto.

Barbosa, 2006	36 frasi, 1 locutore
Dellwo, 2004	80 sillabe, 12 locutori
Low & Grabe, 2000	10 frasi, 10 locutori
Mairano & Romano, 2006	15 frasi, 516 segmenti
Ramus <i>et alii</i> , 1999	5 frasi, 4 locutori

² <http://www.cirass.unina.it/ricerca/studi%20parlato/raccolta%20corpora/api/api.htm>

Rouas & Farinas, 2004	15 frasi, 10 locutori
-----------------------	-----------------------

Tabella 1: Ampiezza della base di dati per ogni lingua studiata in alcuni lavori di ricerca.

Occorre infine sottolineare che con il modello CCI non presumiamo affatto di cogliere in maniera esauriente le proprietà del ritmo delle lingue naturali. Il problema è assai più complesso, e a nostro avviso va affrontato su due livelli distinti e non necessariamente convergenti, come argomentiamo altrove (Bertinetto & Bertini, in preparazione), ossia:

(i) Livello-I, in cui si dispiega l'effetto della struttura fonotattica e che si fonda sull'accoppiamento di due oscillatori (vocalico e consonantico), secondo la suggestiva proposta di Goldstein (Goldstein *et alii*, 2007).

(ii) Livello-II, anch'esso imperniato sull'azione di due oscillatori accoppiati (accentuale e sillabico), secondo la proposta di O'Dell & Nieminen (O'Dell & Nieminen, 1999).

Da quanto detto, appare evidente che il modello CCI agisce unicamente al livello-I, ed è dunque a questo che ci limiteremo in questo lavoro. Trattandosi peraltro del livello in cui si dispiega l'effetto della fonotassi, esso costituisce una componente basilare ed inaggirabile del meccanismo ritmico delle lingue, pur senza esaurirne la complessità.

Per comodità del lettore, forniamo qui un breve riassunto dell'articolazione di questo lavoro. Nel § 2 sono riportati i criteri secondo cui sono stati selezionati i materiali a partire dal corpus di riferimento. Nel § 3 sono succintamente riassunte le principali differenze emerse nel confronto tra il modello CCI ed i modelli di Ramus (Ramus, 1999), Grabe & Low (Grabe & Low, 2002), Dellwo (Dellwo, 2004), Rouas & Farinas (Rouas & Farinas, 2004). Il § 4 riporta l'esito delle analisi condotte sulle diverse variabili considerate nel nostro corpus: ruolo del locutore ("Giver" vs. "Follower"), velocità di elocuzione, tipo sintattico-intonativo (enunciati dichiarativi vs. interrogativi), sesso del locutore. Il § 5, infine, riassumerà le nostre conclusioni.

2. CRITERI DI SELEZIONE DEI MATERIALI

2.1 Materiali

La ricerca qui illustrata è stata condotta analizzando la variante pisana dell'italiano, contenuta nella banca-dati AVIP/API. Abbiamo preso in considerazione 5 coppie di parlanti (3 femmine e 7 maschi). Trattandosi di testi semispontanei elicitati con la tecnica del "map-task", è stato necessario stabilire preliminarmente dei criteri, per selezionare il materiale in modo tale da ottenere porzioni di parlato sufficientemente estese e allo stesso tempo uniformi in termini di "continuità" di eloquio. Ciò è indispensabile per estrarre indicazioni attendibili circa le tendenze ritmiche.

Per la selezione delle porzioni di enunciato utilizzabili abbiamo applicato nell'ordine i seguenti criteri:

I) sono state scelte enunciati privi di esitazioni, pause, forme di assenso, esclamazioni, false partenze, fenomeni vocali non verbali, sequenze inintelligibili, routines discorsive (come le frasi fatte ricorrenti);

II) sono state ulteriormente eliminate le parti terminali di enunciato di tipo asseverativo (ad esempio: ..., *no?*), così come quelle introduttive (ad esempio: *cioè*,...);

III) sono state selezionate sequenze che, in trascrizione ortografica, avessero almeno 9 sillabe (≥ 9 sillabe) e, foneticamente, almeno 8 sillabe (≥ 8 sillabe). La differenza è motivata dal fatto che non tutte le sillabe potenziali si realizzano effettivamente nel parlato, a causa di possibili fenomeni di fusione tra vocali adiacenti, riduzione di iato, ipoarticolazione etc.

Un problema particolarmente delicato è costituito dalla misurazione delle componenti segmentali dei dittonghi. In italiano, secondo l'opinione prevalente benché non unanime dei fonologi, i legamenti che compongono i dittonghi ascendenti e discendenti vanno considerati fonemi consonantici approssimanti (Bertinetto & Loporcaro, 2005). È dunque essenziale, per il funzionamento di un algoritmo come CCI calcolare esattamente le durate

del legamento e della vocale nucleare.³ Purtroppo, come ben sa chiunque si sia cimentato nella materia, la divisione tra legamento e vocale è impresa difficile e delicata; a maggior ragione in un corpus di parlato spontaneo, con fasi di elocuzione accelerata o addirittura ipoarticolate. È stato dunque necessario adottare una strategia operativa, con tutti i rischi del caso. Dopo una serie di analisi pilota, si è optato per la seguente soluzione:

(a) per i maschi, la durata del legamento è stata fatta coincidere convenzionalmente con le prime o, a seconda dei casi, con le ultime due vibrazioni laringali del dittongo;

(b) per le femmine, tale durata è stata fissata in 4 vibrazioni, tenuto conto della diversa frequenza intrinseca della voce femminile;

(c) quando l'intervallo corrispondente alle 2 o 4 pulsazioni superava un terzo della durata complessiva del dittongo (il che può accadere nel parlato veloce), si è convenzionalmente attribuita al legamento una durata pari ad 1/3 della durata complessiva.

La procedura adottata è sicuramente discutibile, ma ha almeno il vantaggio di non assegnare una durata fissa ad ogni legamento, dato che la durata delle pulsazioni laringali può variare da istante a istante. In tal modo, si è scongiurato il rischio di introdurre un livellamento artificiale della varianza.

Anche nel caso delle geminate si è adottato un criterio a base fonologica. L'intervallo consonantico corrispondente ad una geminata è stato pertanto interpretato come composto da due segmenti fonemati, la cui durata è stata convenzionalmente considerata pari alla metà dell'intero intervallo. Gli intervalli vocalici, per parte loro, consistevano per lo più di un singolo elemento, eccezion fatta per le sinalefi; le vocali in iato, in quanto riferibili a diversi nuclei sillabici, sono infatti state assegnate ad intervalli vocalici distinti. Beninteso, se in italiano vi fosse quantità vocalica pertinente (come in finnico o in giapponese), gli intervalli vocalici potrebbero anche constare di due elementi, in analogia a quanto accade con le geminate (e se avessimo a che fare con una lingua come l'estone, con vocali e consonanti superlunghe, dovremmo ammettere la possibilità di dividere addirittura per tre ogni fono vocalico o consonantico della lunghezza appropriata).

Le porzioni di enunciato selezionate per il nostro corpus sono state suddivise in enunciati interrogativi e dichiarativi. Tuttavia, gli enunciati in cui l'elemento interrogativo era unicamente riferibile all'espressione conclusiva preliminarmente eliminata (vedi il secondo criterio di selezione), sono stati etichettati come dichiarativi.

Per il calcolo della velocità di eloquio (in sillabe per secondo) è stata valutata la divisione sillabica della trascrizione fonetica, tenendo dunque conto delle sillabe effettivamente realizzate e considerando le vocali in iato come nuclei sillabici a sé stanti.

2.2 Verifiche preliminari

In via preliminare, sono state effettuate alcune verifiche necessarie ad interpretare correttamente i risultati. In particolare, si è considerato il possibile effetto derivante da: (i) la sovrapposizione di voci dei due locutori; (ii) le diverse categorie sintattico-intonative degli enunciati.

2.2.1 Analisi delle sequenze con sovrapposizione di voci

Le modalità con cui è stato registrato il parlato semi-spontaneo analizzato (i due parlanti disponevano di microfoni unidirezionali, ma si trovavano nel medesimo ambiente) possono comportare casi di sovrapposizione di voci, con possibili alterazioni della fluenza, magari impercettibili a prima vista.

Sono state dunque confrontate tre diverse selezioni dei materiali del corpus, così caratterizzate:

(a) eliminazione di tutti gli enunciati contenenti, anche se in minima parte, delle sovrapposizioni di voci;

³ Ma lo stesso varrebbe, in realtà, per qualsiasi altro modello, a meno che non si voglia salomonicamente ignorare il problema della natura consonantica dei legamenti, come in effetti ci risulta che sia stato fatto nelle ricerche finora prodotte.

(b) eliminazione delle sole porzioni di enunciato contenenti sovrapposizioni (quando però la sovrapposizione si colloca nella parte centrale, l'intero enunciato deve essere comunque eliminato, data la discontinuità che si verrebbe a creare giustapponendo parte iniziale e finale);

(c) come nel caso precedente, ma conservando gli enunciati in cui la sovrapposizione sia limitata alla parte finale, nell'ipotesi che in tal caso la distorsione ritmica sia di entità limitata.

I risultati, calcolati secondo l'indice CCI, sono riportati nella figura 1, avendo in ascissa il valore degli intervalli vocalici e in ordinata quello degli intervalli consonantici. La prossimità dei punti, anche in rapporto al "braccio" delle varianze, mostra che non ci sono significative differenze tra i tre tipi di elaborazione; ciò è confermato anche dalle indagini statistiche. È stato pertanto scelto di eliminare dalle successive analisi solo le porzioni di frase in cui si ha sovrapposizione, in modo da poter conservare il maggior numero possibile dei dati a disposizione.

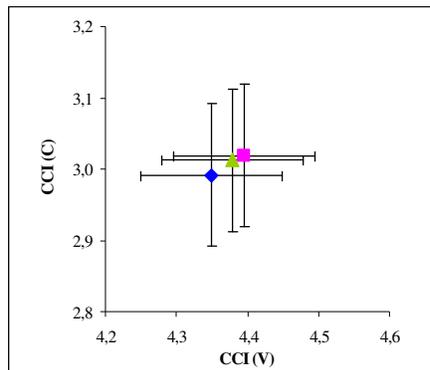


Figura 1: Confronto dell'indice CCI per diversi tipi di elaborazione del corpus, in rapporto alle sovrapposizioni di voci. Le "braccia" indicano la deviazione standard della media. Quadrato rosa: caso (a); rombo blu: caso (b); triangolo verde: caso (c). Per una puntuale definizione dei tre tipi di selezione, si veda il testo.

2.2.2 Confronto tra diverse categorie di enunciato

Successivamente, si è considerato l'effetto congiunto della sovrapposizione di voci e dell'allungamento di fine enunciato. Sono stati eseguiti i calcoli relativi a tre insiemi di enunciati, così denominati;

- enunciati senza la porzione finale: per ogni enunciato, è stata eliminata la porzione che va dall'ultima sillaba accentata in poi;
- enunciati senza sovrapposizione di parlanti: da ogni enunciato è stata eliminata la porzione di frase con sovrapposizione (come nel caso (b) descritto nel paragrafo 2.2.1);
- enunciati senza ultima accentata né sovrapposizione di voci (in pratica, sommando le due condizioni precedenti).

Questi tre insiemi sono stati confrontati con il corpus inizialmente selezionato, indicato come "intero". Il calcolo dell'indice CCI riportato in figura 2 evidenzia, come atteso, il forte effetto della porzione di eloquio che segue l'ultima sillaba accentata, in netto contrasto con l'effetto prodotto dalla sovrapposizione di voci. La non significativa discrepanza tra gli enunciati del corpus "intero" e quelli privati delle sovrapposizioni conferma che le alterazioni causate dalla mescolanza di voci non hanno prodotto conseguenze rilevanti.

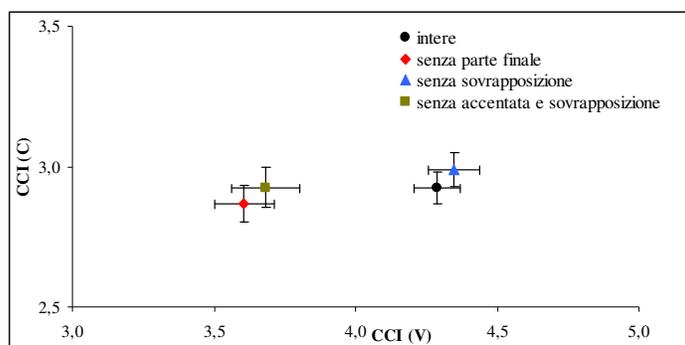


Figura 2: Confronto dell'indice CCI tra quattro diverse selezioni dei materiali: pallino nero: corpus "intero"; rombo rosso: enunciati privi della parte finale (dall'ultima sillaba accentata in poi); triangolo celeste: enunciati privi di sovrapposizione di voci; quadrato verde: somma delle due ultime condizioni.

In seguito ai risultati ottenuti in queste indagini preliminari, si è ritenuto di poter lavorare su tutti gli enunciati inizialmente selezionati, avendo unicamente cura di trascurare le porzioni finali degli stessi (dall'ultima sillaba accentata in poi). Ciò non significa, peraltro, che noi non attribuiamo alcun senso ritmico alla parte terminale degli enunciati, bensì che tale parte merita di essere analizzata indipendentemente, in quanto presenta caratteristiche proprie, con andamenti radicalmente diversi da quelli che riguardano il corpo dell'enunciato. A ciò si aggiunga che le caratteristiche della chiusa degli enunciati differiscono da lingua a lingua in maniera spesso capricciosamente idiosincratica, secondo le convenzioni adottate dalle diverse comunità di parlanti. Sappiamo, per esempio, che non tutte le lingue (o varietà di lingue) seguono la tendenza al rallentamento finale.

La tabella 2 riporta la quantità di dati utilizzati nelle analisi che seguiranno.

FONEMI		8410
INTERVALLI	vocalici	2991
	consonantici	2995
	Follower	1731
	Giver	4255
	dichiarative	5342
	interrogative	644
	maschio	4500
	femmina	1486
	TOTALI	5986
	N° medio di segmenti per intervalli vocalici	
N° medio di segmenti per intervalli consonantici		1.4
Durata media intervalli vocalici		79.2 ms
Durata media intervalli consonantici		78.8 ms
SOGGETTI		10

Tabella 2: Riepilogo dei dati analizzati nel corpus "consolidato", ossia depurato degli elementi che possono introdurre significative deformazioni nell'andamento ritmico.

3. CONFRONTO TRA MODELLI

Va osservato che la formula CCI non prende in considerazione la nozione di 'sillaba'. In Bertinetto & Bertini (in stampa) abbiamo fornito le ragioni teoriche di questa scelta, su cui

qui non ritorniamo. In ogni caso, la nostra scelta coincide con quella adottata in altri ben noti modelli, come il modello PVI o quello di Ramus.

Il modello CCI assomiglia per certi aspetti a quello proposto da Rouas & Farinas (Rouas & Farinas, 2004), ma con importanti differenze. R&F introducono la nozione di ‘semi-sillaba’, definita come la sequenza che va da una vocale (esclusa) alla successiva (compresa). La parte consonantica di tale sequenza viene poi divisa per il numero di segmenti che la compongono, mentre gli intervalli vocalici sono mantenuti indivisi. Nel modello CCI, per contro, gli intervalli vocalici possono anche consistere di più segmenti (quando vi sia sinalefe o quando si abbia a che fare con una vocale di quantità fonologica lunga), anche se in molti casi essi consistono di un unico elemento. Inoltre, mentre il modello di R&F rientra nel novero degli algoritmi che si potrebbero definire “statici”, in quanto prescindono dall’effettiva successione dei segmenti che compongono la catena fonica, il modello CCI appartiene alla categoria degli algoritmi “dinamici”, il cui capostipite è il modello PVI di Grabe & Low. Ciò comporta una visione nettamente distinta delle cose, che potremmo metaforicamente riassumere così: nei modelli dinamici, si tiene conto delle variazioni locali nella durata degli elementi misurati, come se avessimo a che fare con una successione di fotogrammi; nei modelli statici si fornisce invece, per così dire, un fotogramma unico e riassuntivo.

In sostanza, CCI presenta le seguenti caratteristiche:

i) sfrutta – a differenza del modello di R&F – le potenzialità dinamiche del modello PVI, a nostro avviso superiore al modello ramusiano nella sua capacità di seguire a passo a passo le effettive oscillazioni nella durata degli elementi adiacenti;

ii) prende in debita considerazione – a differenza del modello PVI – il livello di complessità fonotattica, riflesso nel numero di elementi che compongono gli intervalli (e non solo quelli consonantici, come in R&F). Il fattore fonotattico è, in definitiva, alla radice del contrasto fra lingue ‘a controllo’ e ‘a compensazione’.

La figura 3 delinea la situazione ideale, quale si può predire sulla base di quanto ci è noto circa il comportamento delle lingue. Le lingue che chiamiamo “a controllo” dovrebbero cadere nella parte bianca del grafico, in quanto le fluttuazioni di durata delle vocali e delle consonanti tendono ad essere della medesima entità. Le lingue ‘a compensazione’ dovrebbero invece cadere nella parte grigia inferiore, perché le fluttuazioni vocaliche sono complessivamente maggiori di quelle consonantiche, a causa del maggior grado di coarticolazione (o sovrapposizione articolatoria) cui le vocali vanno soggette. La parte grigia superiore definisce, per contro, un’area in cui non ci si aspetta di trovare alcuna lingua: è infatti poco plausibile che le fluttuazioni di durata (ed i conseguenti livelli di co-articolazione/co-produzione) siano decisamente maggiori per le consonanti che non per le vocali.

È opportuno precisare che solo un modello come CCI permette di dare un’interpretazione altrettanto rigorosa dello spazio cartesiano, imperniato sulla retta bisettrice del piano. Negli altri modelli finora proposti, non potendosi sapere a quanti segmenti corrisponde ciascun intervallo, questo importante punto di riferimento viene a mancare, anche se la posizione in figura permette comunque di valutare quale, tra le due componenti (vocalica e consonantica, ma solo in riferimento ad intervalli indifferenziati), presenta complessivamente le oscillazioni maggiori.

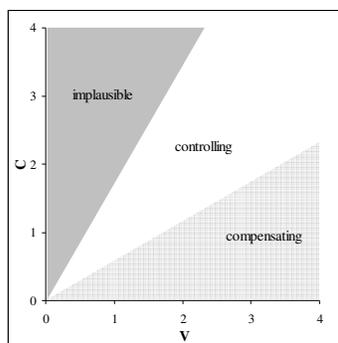


Figura 3: Rappresentazione schematica dei due prototipi ritmici fondamentali, secondo l'ipotesi "controllo/compensazione"..

La figura 4 mostra le posizioni ottenute a partire dai dati del nostro corpus per i seguenti modelli:

- CCI
- Ramus in due delle sue possibili versioni, con ΔV o $\%V$ come misura vocalica;
- PVI in due diverse versioni, con dati grezzi (rPVI) o normalizzati (nPVI; con normalizzazione riferita alle vocali);
- Varco in due delle sue possibili versioni, con VarcoV o $\%V$ come misura vocalica;
- R&F, calcolato peraltro nella versione in cui, come misura vocalica, si ha la media della durata degli intervalli (D_v).

Nella versione ΔV , il modello ramusiano calcola la deviazione standard per gli intervalli sia vocalici sia consonantici. Nella versione $\%V$, la misura vocalica è invece data dalla percentuale di intervalli vocalici rispetto alla durata complessiva della sequenza. Quanto al modello PVI, esso si basa sulla somma delle differenze di durata tra intervalli adiacenti. Esso funziona praticamente come indicato dalla formula (1) sopra riportata, ma con la cruciale assenza della variabile n , che indica il numero di segmenti compresi in ciascun intervallo. Nella sua variante nPVI, ogni differenza tra intervalli consecutivi è normalizzata rispetto alla loro durata media, al fine di eliminare la variabilità dovuta alle variazioni della velocità di eloquio all'interno della sequenza considerata. Quanto al modello Varco, esso si basa sul rapporto tra la deviazione standard e il valor medio degli intervalli adiacenti di ogni porzione di enunciato. Infine, il modello di Rouas & Farinas (R&F) calcola la media della durata degli intervalli sia vocalici sia consonantici in ogni sequenza.

Per poter collocare su un medesimo grafico i risultati ottenuti dai diversi algoritmi, i valori vocalici dei due modelli ramusiani e di nPVI sono stati divisi per un fattore 10. Gli ovali della figura 4 indicano la dispersione spaziale dei 10 soggetti considerati nel corpus; per avere un'idea della quale (limitatamente all'applicazione del modello CCI) si può consultare la figura 5, da cui risulta peraltro che la distribuzione spaziale dei locutori è tutto sommato abbastanza omogenea, con una maggior fluttuazione vocalica piuttosto che consonantica. L'analisi qualitativa mostra la collocazione particolarmente periferica dei locutori D2G e BF rispetto agli altri. Peraltro, il valore elevato della deviazione standard dei suddetti locutori lascia intendere che la responsabilità della loro devianza sia soprattutto attribuibile al minor numero di dati presi in considerazione. Le analisi statistiche evidenziano inoltre significative differenze tra svariate coppie di locutori, con contrasti più o meno forti a seconda che vocali e consonanti siano considerate separatamente o congiuntamente. La variabilità di questi risultati conferma l'idea già precedentemente avanzata, anche nelle ricerche sul ritmo, secondo cui lo studio di un solo locutore o di un numero ridotto di locutori alimenta ragionevoli dubbi circa la fondatezza delle conclusioni che si possono trarre, data la scarsa rappresentatività statistica dei dati.

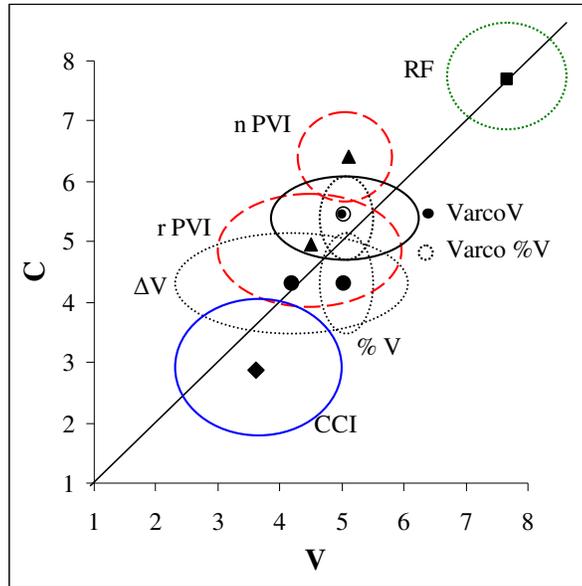


Figura 4: Rappresentazione spaziale dei modelli CCI, PVI, Ramus, Varco e R&F nelle possibili versioni. Gli ovali rappresentano la dispersione spaziale dei 10 soggetti considerati.

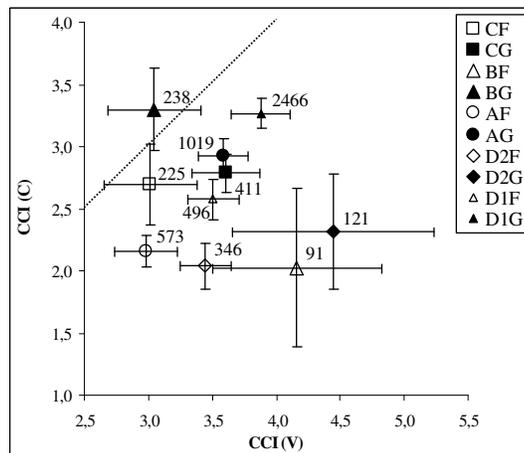


Figura 5: Confronto tra i 10 locutori secondo il metodo CCI, con indicazione del numero di intervalli analizzati per ognuno.

Pur con le cautele sopra segnalate quanto all'interpretazione della bisettrice, si possono avanzare alcune considerazioni in merito alla collocazione rispettiva dei punti generati dai diversi algoritmi. I modelli ΔV , $rPVI$ e $R\&F$, che generano un punto collocato in pratica sulla bisettrice, suggeriscono che vocali e consonanti abbiano, nel complesso, un comportamento simile. Per contro, con $nPVI$ e Varco (anche se in forma più lieve) l'oscillazione delle consonanti appare più forte di quella delle vocali: ma occorre dire che, nel primo modello, le misure di vocali e consonanti non sono rigorosamente omogenee, anche perché esso è specificamente concepito per ridurre la varianza tra le vocali. I modelli $\%V$ e CCI indicano invece la tendenza opposta, con fluttuazioni leggermente maggiori per le vocali che per le consonanti. Sulla base di quanto sappiamo dell'effettivo comportamento

dell'italiano, quest'ultima situazione appare più consona ai dati reali. In italiano ci sono in effetti variazioni di durata relativamente apprezzabili tra vocali toniche ed atone, anche se esse sono assai più contenute di quanto spesso asserito (Bertinetto, 1981; Dell'Aglio *et alii*, 2002; Bertinetto & Loporcaro, 2005). Le variazioni sul piano della durata consonantica sono invece più contenute, e su di esse non può certo influire l'opposizione fonologica tra scempie e geminate, dato che queste ultime sono state considerate come una sequenza di due segmenti, in accordo con la maggior parte dei modelli fonologici. In sostanza, pur senza raggiungere la condizione delle lingue "a compensazione", l'italiano sembra parzialmente orientato in tale direzione, dato che non costituisce certo un esempio estremo di semplicità fonotattica.

Tenendo conto dei risultati prodotti dai vari modelli confrontati, ci sentiamo autorizzati ad asserire che, pur da questi risultati preliminari, il modello CCI sembra offrire i risultati più attendibili. In effetti, il nostro modello:

- è apparso non meno sensibile dei suoi migliori competitori nel rendere le fini differenze tra soggetti diversi (cf. figura 5);
- sembra offrire i risultati più affidabili in termini di posizionamento sul piano cartesiano.

Ovviamente, lo stadio successivo della verifica dovrà vertere sul confronto fra lingue diverse, rispetto alle quali il modello dovrà produrre raggruppamenti ragionevolmente interpretabili, sia nei rapporti reciproci tra le diverse lingue, sia soprattutto in relazione alle predizioni della teoria (cf. la figura 3). Il nostro prossimo compito consisterà dunque nel confrontare gli esiti ottenuti sull'italiano con quelli ottenibili su materiali analoghi, tratti da altre lingue. Stiamo attualmente lavorando su due corpora di dialoghi semispontanei, prodotti da parlanti tedeschi e brasiliani. Quanto al confronto con i materiali letti, ci proponiamo di studiare il comportamento dei nostri dieci parlanti nella lettura di frasi, a suo tempo registrate nella stessa occasione in cui è stato prodotto il corpus dialogato. Ci conforta non poco, comunque, l'esito ottenuto dal già citato lavoro di Mairano & Romano (in preparazione), che hanno confrontato otto lingue (tra cui la nostra) utilizzando l'algoritmo CCI, riscontrando comportamenti in linea con le nostre predizioni. In ogni caso, si può sin d'ora affermare che la copertura empirica offerta dai nostri materiali, in termini di stabilità statistica, è molto più ragguardevole di quella normalmente constatabile in questo genere di lavori, date le dimensioni del campione testato.

4. CONFRONTO TRA VARIABILI

Uno degli scopi della nostra ricerca consisteva nel mettere a fuoco alcune variabili discorsive caratterizzanti il nostro corpus, per valutarne l'impatto in relazione all'andamento ritmico. Come nei casi già visti, i risultati sono rappresentati dalla proiezione di un punto sul piano cartesiano, avente come coordinate il valore dell'indice CCI calcolato sulle vocali e sulle consonanti. I risultati verranno riportati sia graficamente sia in senso statistico, attraverso (M)ANOVA, t-test e test post-hoc. Possiamo fin d'ora asserire che i valori numerici estraibili dal nostro corpus possiedono una considerevole stabilità statistica, dato il numero di intervalli presi in esame.

La prima variabile considerata riguarda il contrasto tra enunciati dichiarativi e interrogativi, in rapporto all'eventuale influenza esercitata sulle tendenze ritmiche da questi due tipi sintattico-intonativi. La figura 6 mostra che non esiste una differenza significativa (statisticamente: $p > 0.05$) per quanto riguarda tale contrasto, nonostante i punti sul piano siano separati.

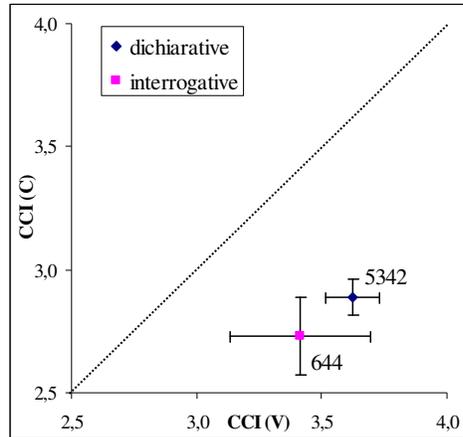


Figura 6: Confronto tra enunciati dichiarativi e interrogativi, con indicazione del numero di intervalli presi in considerazione.

Dalla medesima figura emerge, peraltro, anche il forte divario numerico tra i dati analizzati per i due gruppi a confronto, il che potrebbe aver influito sull'esito statistico. Per ovviare a ciò, è stata creata una selezione di frasi dichiarative, tratte in percentuali uguali dai dieci locutori (cf. la figura 7). In particolare, sono stati scelti enunciati tali che il valor medio della velocità di eloquio fosse analogo a quello degli enunciati interrogativi, e ciò al fine di neutralizzare la non trascurabile influenza del fattore velocità (come apparirà meglio in seguito). Sono state fatte più selezioni seguendo il criterio sopra indicato, ma in ciascun caso il risultato è rimasto invariato.

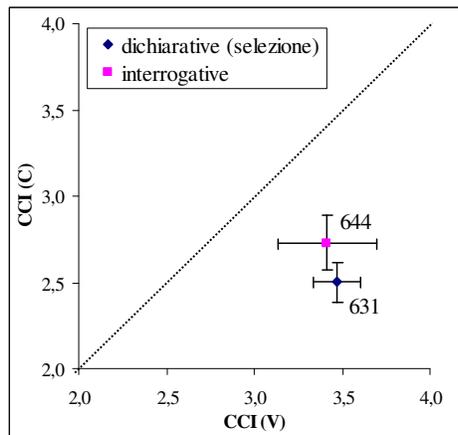


Figura 7: Confronto tra enunciati interrogativi e una selezione di enunciati dichiarativi, con numero di intervalli paragonabile e velocità di eloquio bilanciata.

Successivamente, abbiamo esaminato il contrasto tra ruolo di “Giver” e ruolo di “Follower” nei dialoghi del nostro corpus. Le registrazioni facenti parte della banca-dati AVIP/API sono state infatti realizzate con la tecnica del “map-task”, che obbliga i due locutori a svolgere un ruolo diverso nella conversazione: “Giver”, colui che dà le istruzioni sul percorso da seguire, e “Follower”, colui che deve mettere in pratica le istruzioni. Questo aspetto influisce inevitabilmente sulle strategie discorsive dei locutori e potrebbe di conseguenza influire anche sulle loro inclinazioni ritmiche. In effetti, dai nostri risultati

emerge un sostanziale contrasto tra i due ruoli (cf. figura 8), confermato dalla differenza statisticamente significativa ($F(1,234)$, $p < 0.01$).

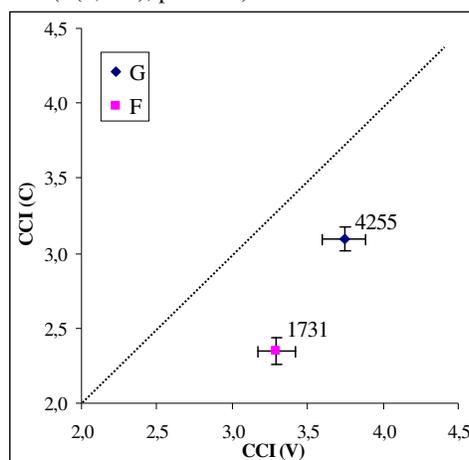


Figura 8: Confronto tra i due ruoli discorsivi (Giver vs. Follower), con relativo numero di intervalli analizzati.

Anche in questo caso, il divario tra il numero di intervalli analizzati per le due categorie a confronto potrebbe far strada all'idea che il risultato non sia veritiero. Come nel caso precedente, sono state dunque create diverse selezioni, in modo da rendere paragonabile la quantità di dati, ma il risultato è rimasto invariato. Si noti peraltro (e su ciò ritorneremo) che in questo caso non è stato possibile equilibrare la velocità di elocuzione, data la netta discrepanza che sussiste a tal riguardo tra i due ruoli discorsivi considerati.

Appare dunque evidente che tali ruoli influiscono sul comportamento ritmico dei parlanti. E poiché, nel nostro corpus di riferimento, ogni parlante assumeva un unico ruolo, è anche possibile che le differenze riscontrate abbiano a che fare con tendenze ritmiche individuali. Peraltro, osservando la figura 5 (dove i simboli pieni indicano i "Givers" e quelli vuoti i "Followers"), appare evidente come, al di là della diversità tra i singoli locutori, vi sia una ripartizione spaziale ben riconoscibile tra i due ruoli discorsivi.

Esiste tuttavia un'altra possibilità di interpretazione. Il contrasto tra i due ruoli appare infatti fortemente correlato con la velocità di elocuzione, dato che i "Givers" sono in media notevolmente più lenti dei "Followers" (6.4 vs. 7.5 sill/s). Abbiamo pertanto analizzato l'impatto del fattore velocità sulla configurazione ritmica, impatto già osservato in diversi studi sull'argomento (si vedano, per esempio, i lavori di Dellwo).

Ogni enunciato è stata catalogato a seconda della velocità di eloquio (lenta, media e veloce). Le tre classi, con valori medi equamente distribuiti, sono state così definite:

- v1: al di sotto di 6.99 sill/s (numero di intervalli vocalici e consonantici: 1301, 1308)
- v2: da 7 sill/s a 8.1 sill/s (numero di intervalli vocalici e consonantici: 660, 663)
- v3: al di sopra di 8.1 sill/s (numero di intervalli vocalici e consonantici: 514, 507)

Il grafico di figura 9 conferma quanto atteso: la diversa velocità di elocuzione ha tangibili conseguenze sull'andamento ritmico. All'aumentare della velocità, si nota una contrazione della fluttuazione nelle durate vocaliche e consonantiche, tutto sommato proporzionale anche se leggermente maggiore nel primo caso. Una prima indagine statistica indica significative differenze tra le velocità ($F(2, 233)$, $p < 0.01$); in particolare il test Games-Howell indica, tanto per le vocali quanto per le consonanti, una significativa ($p < 0.01$) differenza tra velocità bassa e medio-alta, ma non tra alta e media.

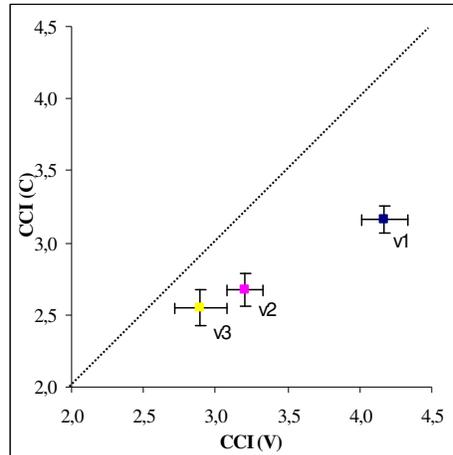


Figura 9: Effetto della velocità di eloquio (v1 = lenta, v2 = media, v3 = veloce).

Abbiamo inoltre fatto alcune prove per verificare quanto tale risultato dipendesse dalla scelta dei confini tra le classi di velocità. L'analisi ha mostrato che i risultati non sono sensibili a piccole variazioni di tali confini; per questo motivo, è stata mantenuta la sopraindicata suddivisione dello spazio delle velocità, fondata su una distribuzione equa dei rispettivi valori medi, malgrado il diverso numero di dati riferibili a ciascuna categoria. Ulteriori indagini, volte a creare nuovi e più fini raggruppamenti, sono attualmente in corso.

Alla luce di questi risultati, abbiamo rianalizzato il confronto tra "Follower" e "Giver", constatando che il suddetto contrasto è fortemente correlato con la diversa velocità di elocuzione (vedi figura 10). Le analisi statistiche hanno inoltre mostrato che il fattore "velocità" influisce maggiormente, ossia spiega una maggior quota di varianza, rispetto al fattore "ruolo discorsivo".

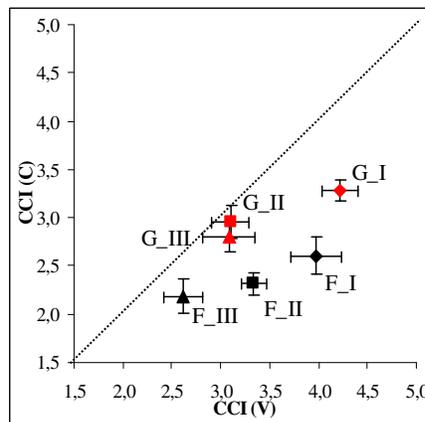


Figura 10: Confronto tra i due ruoli discorsivi (G vs. F) ripartiti secondo le velocità di elocuzione (I-III).

Sebbene risulti evidente che la variabile "velocità" è un fattore di ordine superiore rispetto alla variabile "ruolo discorsivo", resta da capire qual è l'effettivo influsso di quest'ultimo. Per dare una risposta a questa domanda, ci riserviamo di analizzare il

comportamento di locutori che, nel corso del “map task”, si scambiano di ruolo, come accade nei dialoghi del corpus CLIPS.⁴

Infine, abbiamo messo a fattore la variabile “sesso” (maschi vs. femmine), mediando sui valori riferibili ai due insiemi di locutori. Come previsto, le velocità medie differiscono: 6.3 vs. 8.0 sill/s per, rispettivamente, voce maschile e femminile. Peraltro, nonostante il grafico di figura 11 mostri come vi sia una certa differenza tra i due sessi, l’analisi statistica non rileva alcuna significatività ($p > 0.05$).

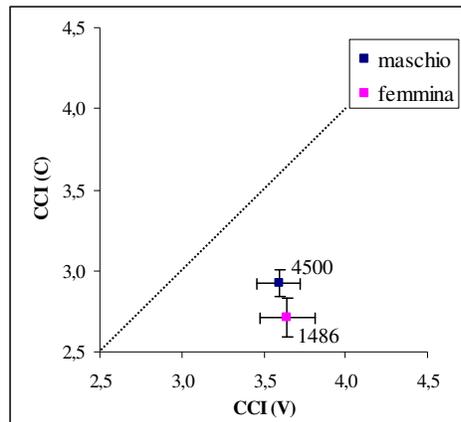


Figura 11: Confronto tra parlanti maschi e femmine, con relativo numero di intervalli analizzati.

Poiché esiste un divario notevole tra il numero di dati riferibili ai due insiemi, sono state create, in analogia con quanto fatto in precedenza, varie selezioni tra le produzioni maschili, tenendo conto del contributo di tutti i locutori e cercando per quanto possibile di equilibrare la velocità di eloquio. Il risultato è tuttavia rimasto statisticamente invariato.

5. CONCLUSIONI

Lo scopo di questo lavoro era duplice:

- (a) presentare un nuovo strumento per l’analisi degli andamenti ritmici del parlato;
- (b) applicarlo allo studio di alcune cruciali variabili discorsive individuabili nel corpus di riferimento, basato sulla varietà pisana dell’italiano registrata in condizioni semi-spontanee.

Quanto al primo punto, il confronto effettuato con altri algoritmi indica che il modello CCI sembra presentare alcuni indubbi vantaggi (cf. § 3). Ulteriori considerazioni in proposito, basate su accurati confronti statistici, si possono trovare in altri nostri studi (Bertinetto & Bertini 2008; Bertinetto & Bertini, in stampa).

L’analisi delle variabili discorsive ha mostrato (cf. § 4) che alcuni fattori influiscono fortemente sulla configurazione ritmica, mentre altri sembrano esercitare effetti piuttosto contenuti. In particolare, è emersa la forte influenza della velocità di eloquio e del ruolo del locutore (in quest’ordine gerarchico di importanza), a differenza di variabili come il tipo sintattico-intonativo (enunciati dichiarativi vs. interrogativi) o il sesso del parlante.

Questa ricerca ribadisce anche che l’utilizzo di un corpus debitamente ampio è assolutamente indispensabile per conferire sufficiente solidità statistica alle osservazioni. Ciò emerge in particolare dalla stabilità dei nostri esiti in tutti i casi in cui si è provato ad applicare il calcolo a selezioni diverse di enunciati, allo scopo di bilanciare le variabili in esame. Il che suggerisce, per converso, di considerare con il dovuto beneficio di inventario

⁴ www.clips.unina.it

molti dati presenti in letteratura circa l'andamento ritmico delle lingue, quand'essi siano basati su un numero troppo limitato di dati o addirittura su un solo locutore.

6. BIBLIOGRAFIA

- Barbosa, P. (2006), *Incursões em torno do ritmo da fala*, Campinas: Pontes.
- Bertinetto, P.M. (1981), *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici*, Firenze: Accademia della Crusca.
- Bertinetto, P.M. (1989), Reflections on the dichotomy 'stress- vs. syllable-timing', *Revue de Phonétique Appliquée*, 91/93. 99-130.
- Bertinetto, P.M. & Bertini, C. (2008), On modeling the rhythm of natural languages: the 'Control/Compensation Index'. *Speech Prosody 2008*, Campinas, Brasil, May 6-9.
- Bertinetto, P.M. & Bertini, C. (in stampa), Modelización del ritmo y estructura silábica, con aplicación al italiano, in Sanchez Miret, F. (cur.), *Romanística sin complejos: Homenaje a Carmen Pensado*, Bern: Lang.
- Bertinetto, P.M. & Bertini, C. (in preparazione), Towards a unified predictive model of natural language rhythm, comunicazione presentata al convegno *Empirical Approaches to Language Rhythm*, Londra, Inghilterra, 28 marzo 2008.
- Bertinetto, P.M. & Fowler, C.A. (1989), On sensitivity to durational modifications in Italian and English, *Rivista di Linguistica*, 1. 69-94.
- Bertinetto, P.M. & Loporcaro, M. (2005), The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome, *Journal of the International Phonetics Association* 35: 131-151.
- Bertinetto, P.M. & Vékás, D. (1991), Controllo vs. compensazione: sui due tipi di isocronia, in *L'interfaccia tra fonologia e fonetica*, Magno Caldognetto, E. & Benincà, P. (eds.), Padova: Unipress.
- Dauer, R.M. (1983), Stress-timing and syllable-timing reanalyzed, *Journal of Phonetics*, 11. 51-62.
- Dellwo, V. (2004), Rhythm and speech rate: A variation coefficient for ΔC , in *Language and language-processing*, Karnowski, P. & Szigeti, I. (ed.), Frankfurt am Main: Peter Lang. 231-241.
- Fowler, C.A. (1977), *Timing Control in Speech Production*, Indiana University Linguistics Club.
- Grabe, E. & Low, E.L. (2002), Durational variability in speech and the rhythm class hypothesis, *Papers in Laboratory Phonology* 7, 515-546.
- Hoeqvist, C.E. Jr (1983), The perceptual centre and rhythm categories, *Language and Speech*, 26, 4: 367-376.
- Low, E.L., Grabe, E. & Nolan, F. (2000), Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English, *Language and Speech*, 43 (4), 377-401.
- Mairano, P. & Romano, A. (2006), Lingue isosillabiche e isoaccentuali: misurazioni strumentali su campioni di italiano, francese, inglese e tedesco, in *Atti del III Convegno Nazionale AISV Scienze vocali e del linguaggio*, Trento, Italia, novembre- dicembre 28-1, CD-rom.
- Mairano, P. & Romano, A. (in preparazione), A comparison of four rhythm metrics for six languages, *Workshop on Empirical Approaches to Speech Rhythm*, Londra, Inghilterra, 28 marzo 2008.

O'Dell, M. & Nieminen, T. (1999), Coupled oscillator model of speech rhythm, *Proceedings 14° Int. Cong. Phon. Sciences*, San Francisco, U.S.A., 2: 1075-1078.

Pike (1947), Pike, K.L. [1945], *The Intonation of American English*, Ann Arbor.

Ramus, F., Nespors, M. & Mehler, J. (1999), Correlates of linguistic rhythm in the speech signal, *Cognition*, 73. 265-292.

Rouas, J.L. & Farinas, J. (2004), Comparaison de méthodes de caractérisation du rythme des langues, *Workshop MIDL (Modélisation pour l'Identification Des Langues)*, Paris, France, novembre 29-30, 45-50.

Russo, M. & Barry, W.J. (2004), Interaction between segmental structure and rhythm. A look at Italian Dialects and Regional Standard Italian, *Folia Linguistica*, 38, 3-4, 277-296.

Schmid, S. (2004), Une approche phonétique de l'isochronie dans quelques dialectes italo-romans, in *Nouveaux départs en phonologie. Les conceptions sub- et suprasegmentales*, Meisenburg T. & Selig M. (curr.), Tübingen : Narr, 109-124.