

LA VARIABILITA' PROSODICA: DALLA SILLABA AL CONTENUTO INFORMATIVO

Rodolfo Delmonte
Sezione di Linguistica
Dipartimento Studi Asia Orientale
Università Ca' Foscari - Venezia

1. Introduzione

L'importanza della Prosodia nella creazione di Modelli Predittivi per la Sintesi e per il Riconoscimento del Parlato è ormai un fatto acquisito a differenza di quanto si pensava dieci anni fa quando il riconoscimento si stava affermando sulla base degli HMMs e la sintesi era fondata sui modelli moltiplicativi a liste.

Malgrado ciò, la costruzione di Modelli adeguati permane problematica e non è ancora chiaro quale metodologia sia la più proficua. Il problema posto dalla Prosodia è molto semplice da definire ma molto difficile da risolvere. Ormai più nessuno sostiene che la Variabilità Prosodica sia impossibile da modellizzare in quanto intrinsecamente imprevedibile e non significativa linguisticamente.

Questi problemi sono stati esaurientemente affrontati di recente in riferimento alla Sintesi, da van Santen nel suo Keynote Speech a Rodi in occasione di EuroSpeech'97 mettendo a confronto due ipotesi, e sostenendo però una sola delle due, quella a lui più cara, cioè quella segmentale. Le due ipotesi a confronto sono anche quelle che animano la Tavola Rotonda del GFS'98. Se cioè la creazione di un Modello Prosodico in senso esteso (Timinig, Phrasing, F0) possa essere definita sulla base di regole simboliche che vengono successivamente mappate su unità linguistiche appropriate. A loro volta queste unità linguistiche individuano dei segmenti nell'onda sonora - unità fonetico-acustiche - con correlati acustici ben definiti. I segmenti acustici sottostanno a processi di ottimizzazione matematico-statistica sulla base di analisi di dati di training opportunamente selezionati.

Questo primo Modello si potrebbe definire Modello Top-Down o Deduttivo, ed è quello che noi svilupperemo nella nostra analisi sovrasegmentale. Un Modello Bottom-Up o Induttivo sarebbe invece quello sottostante al CART e ai modelli moltiplicativi. L'input in entrambi i casi è formata da uno Spazio Fattoriale. Nei modelli a CART ogni fattore viene successivamente spezzato per Livelli in modo dicotomico per formare un albero di scelta. Anche nei modelli moltiplicativi si applicano i Livelli a ciascun Fattore classificatorio. L'individuazione dei Fattori è invece fondata linguisticamente, in particolare sulla capacità presunta o verificata che ciascun fattore possa codificare delle caratteristiche linguistiche rilevanti a cui corrispondano unità fonetico-acustiche misurabili in maniera consistente ed omogenea.

Benché ci siano stati progressi nella segmentazione automatica, i confini di foni segmentati automaticamente tendono ancora a contenere errori grossolani e peggio ancora, errori sistematici, e quindi non possono essere usati per studi di timing accurati. van Santen analizza il problema di modellare le durate in diversi lavori presentati ad Eurospeech, riconosce la difficoltà insita nel tentativo di quantificare tutte le interazioni tra segmenti

consonantici e vocalici e ne individua la causa nel gran numero di fattori che intervengono a modificare o perturbare il modello. Il numero di combinazioni possibili varia a seconda del numero di fattori considerati, i quali variano in dipendenza della lingua.

Il metodo proposto prevede l'individuazione di tratti fonetico-acustici significativi e la loro organizzazione in un albero di ricerca in modo che ogni nodo terminale si suddivida lo spazio in sotto-classi omogenee. In fondo ad entrambi i Modelli sta la convinzione che la loro creazione serva per prevedere fenomeni linguistici e unità fonetico-acustiche corrispondenti in materiali linguistici organizzati in corpora, nuovi, mai visti prima.

Come abbiamo detto prima, van Santen propende per un Modello predittivo di tipo Bottom-Up che nella definizione dei Fattori e dei Livelli utilizzi tutta la conoscenza linguistica disponibile in campo fonetico-acustico per una data lingua. Il resto di questo articolo è organizzato come segue: nella prossima sezione approfondiremo il problema generale della scarsità (sparsity/sparseness) dei dati di training per la creazione di un modello statistico, in relazione a un Modello Bottom-Up di tipo segmentale; successivamente proporremo una ipotesi per un Modello di Timing fondato sulla struttura sillabica - cioè Top-Down di tipo sovrasegmentale, e quindi a supporto dell'ipotesi presenteremo dei nostri dati presi dall'italiano e dall'inglese.

2. Modelli di Timing e “Sparsity” dei dati

Come sostiene van Santen più volte, “Sparsity of the training corpus remains a central problem in duration data analysis” (ibid. 2652). Inoltre è significativo il fatto che l'individuazione dei fattori rilevanti, delle distinzioni da fare per ciascun fattore, nonché della determinazione della struttura delle sottoclassi “is done with exploratory data analysis”, oppure che “it is known from the literature”, o ancora che “sub-class structure is based on common sense”(ibid. 2652). L'analisi dell'omogeneità delle durate nelle sottoclassi è basata a sua volta su tecniche di adattamento del modello sulla base di test statistici standard. Ritornerebbe quindi il problema della “sparsity” dei dati al quale si tenta di dare soluzione con una tecnica di correzione “piecewise multiplicative”, oppure “on understanding of the effects of sparsity in the data on parameter estimate reliability, and on phonetics knowledge”.

Più in generale, van Santen fa notare come migliorare la prosodia richieda un tipo di modellizzazione speciale che oppone la complessità combinatoria di una lingua con modelli “content-specific” che siano matematicamente trattabili ed abbiano buone proprietà statistiche. Il grande successo delle analisi statistiche nell' ASR ha incoraggiato un numero di ricercatori ad investigare il campo della predizione delle durate segmentali, CART e HMMs. Grandi database di parlato annotato sono state create per questo scopo: tutti questi approcci soffrono del problema della scarsità (sparsity) dei dati esemplificatori e si cerca di sviluppare tecniche per compensare questa carenza di dati, come ad esempio proposto da van Santen.

La contesa o querelle è riferita in particolare ai modelli di Timing per i quali si ritorna a problematiche ben note nel caso di sistemi di riconoscimento e presenti quando una decina di anni fa si doveva affrontare il training degli HMMs. Si tratta della cosiddetta “coverage issue”. Per quanto riguarda la modellizzazione prosodica il problema non è di facile soluzione - o forse non è risolvibile - utilizzando modelli Bottom-Up, in quanto la variabilità prosodica sembra essere un ostacolo intrinseco e insormontabile. A questo riguardo van Santen fa notare,

“... our analyses of the coverage issue, strongly suggest that, although half a billion [acoustic units] seems like a large number, the combinatorial possibilities of the language at large are so vast that the prosodically annotated phone sequences contained in even ten hours of speech are infinitesimal by comparison.”(ibid.23)

Il suo corpus di riferimento è di 10 ore di parlato etichettato, e contiene 36.000 foni da 100-ms in media (cioè 3600 foni all'ora, 600 foni al minuto, 10 foni al secondo), quindi 650 milioni di sequenze o unità foniche - cioè 18 mila combinazioni per quella determinata lingua.

van Santen sostiene che i modelli di Timing basati su unità segmentali sono utilizzabili anche se soffrono del problema della “sparsity” dei dati. Il modello a unità sillabiche, o sovrasegmentale, è a suo parere inutile in quanto non serve partire dalle durate sillabiche e far discendere da queste quelle segmentali perché “it appears that the duration of a syllable depends on the exact segments it contains and their intrinsic durations”(ibid.,27). In particolare, benché la grandezza degli effetti di un fattore in una sotto-classe vari a seconda dei casi interessati, la direzione degli effetti (allungamento, accorciamento o nessuna variazione della durata) è la stessa. L'ipotesi viene chiamata “Directional Invariance”: questa ipotesi è stata da van Santen verificata su più lingue visto che la Bell Lucent Technologies ha costruito una sintesi multilingue che prevede, tra le altre il Cinese, l'Italiano, il Tedesco, il Russo e il Francese oltre all'inglese americano.

Nel proseguo di questo articolo si sosterrà che questa ipotesi possa essere applicata solo a lingue che hanno un range di variabilità nel Timing molto ristretto, e una coincidenza fonetica stretta tra classi di durata diverse, come avviene per le lingue citate, ad esclusione dell'Italiano, e di lingue simili, nei confronti della quale questa ipotesi non dovrebbe poter essere verificata.

3. Modelli Top-Down e Struttura Sillabica

In linea con il modello Top-Down noi assumiamo che una descrizione fonologica ricca della lingua sia sufficiente per descrivere molti degli effetti significativi che governano la durata dei segmenti fonici. In particolare, sosteniamo che la struttura più appropriata per la predizione delle durate sia la sillaba piuttosto che il fonema o il fono singolo. La durata dei singoli foni all'interno della sillaba può semplicemente essere prevista una volta che venga derivata quella della sillaba. I motivi della variabilità delle durate segmentali sono una conseguenza derivabile dalle strutture ad alto livello linguistico piuttosto che dagli effetti contestuali locali in stretta adiacenza.

Da qui la necessità di arrivare alla sillaba come interfaccia utile per codificare la variabilità prosodica del timing, è come detto sopra, dipendente dalla lingua, a nostro parere. Negli studi sull'italiano e sulle lingue a ritmo isosillabico, viene rilevata una maggiore rigidità a livello segmentale a fronte di una maggiore elasticità a livello sovrasegmentale.

L'idea fondamentale espressa nei lavori sul ritmo dell'italiano e dell'inglese da vari autori (vedi lavori riportati nella Bibliografia di Bertinetto, Fava), è che il ritmo isosillabico sia tale in quanto determinato da un lato dalla necessità di preservare le caratteristiche articolatorie dei suoni e realizzare quindi i target formantici delle vocali e delle consonanti, dall'altro dal fatto che la durata è il più importante correlato acustico dell'accento di parola. Per quanto riguarda il primo fattore vale la pena ricordare che non si trovano in lingue isosillabiche fenomeni come la palatalizzazione di occlusive o di fricative quando si incontrano con una semiconsonante palatale *(vediamo => vegiamo, cartesiano ≠

cartefano), in posizione dentale e non alveolare. Sono invece possibili fenomeni di elisione vocalica, ma solo ai confini di parola e nel rispetto di precise restrizioni fonologiche ed eufoniche.

La rigidità coarticolatoria impedisce il mantenimento di una durata costante dei suoni consonantici che viene invece sottoposta alle necessità determinate dall'alternanza di sillabe toniche e atone, riflessa nel timing della sillaba in primo luogo dalla vocale che ne costituisce per regola il nucleo, ma anche rispecchiata dal suo onset. La rigidità articolatoria a livello sillabico ha come contrappunto una notevole libertà/variabilità nel timing dei suoni consonantici costituenti il sistema fonologico della lingua di riferimento, avendo come limite intrinseco quello della comprimibilità articolatoria propria di ogni fono consonantico. Al contrario il ritmo isoaccentuale si caratterizzerebbe per una maggiore libertà articolatoria determinata dalla maggior ricchezza vocalica e dalla maggior libertà coarticolatoria, nonché dalla possibilità di specificare l'alternanza di toniche e atone per mezzo della qualità vocalica, in primo luogo e della presenza di un pitch accent a livello di parola. La maggior ricchezza vocalica sarebbe quindi in funzione della maggior rigidità articolatoria, da qui la possibilità di parlare di indipendenza vocalica e di durate vocaliche intrinseche per l'inglese come riportato da più autori - al riguardo vedi van Santen(1992). Questa nozione comunque viene confermata da quella di altri autori sulla relativa rigidità durazionale delle consonanti, come riportato dallo stesso van Santen e da Crystal & House (1988) nel loro studio segmentale dell'inglese. I dati dell'italiano in nostro possesso confermano quanto già noto e cioè che le durate segmentali sono sottoposte alle restrizioni di struttura sillabica da un lato, e dall'altro alla necessità di realizzare l'alternanza di toniche ed atone che da sola garantisce la individuabilità semantica della parola a livello lessicale.

Infatti, vale la pena ricordare che il numero di omofoni dell'italiano è molto elevato come risulta da un nostro studio preliminare compiuto sui dati del database denominato IWL e presentato in un precedente incontro(Delmonte et al.,1995). Vi sono classi di parole definibili sistematicamente dal punto di vista della loro caratteristica morfologica, sintattica e a volte semantica che hanno una doppia, o anche tripla in alcuni casi, realizzazione fonetica. Direi quindi che più che parlare di restrizioni a livello articolatorio imposte dal sistema fonologico della lingua, si possa addirittura parlare dalle restrizioni primarie imposte dall'atto comunicativo: cioè che quanto viene enunciato sia intelleggibile e comprensibile dall'interlocutore. In questo caso, quindi, sarebbe in primo luogo il sistema lessicale della lingua, oltre a quello pragmatico, quello semantico e per ultimo quello sintattico, a richiedere una realizzazione prosodica del timing accurata, in presenza di un inventario fonemico a livello vocalico relativamente povero ma di una struttura fonotattica molto ricca che sottometta le durate dei singoli suoni a notevole variabilità in funzione della struttura fonologica primaria di interfacciamento con i livelli astratti superiori, la sillaba.

Il modello segmentale assume che siano significative e sufficienti le relazioni che intercorrono in contesti di adiacenza segmentale e proiettano i vari fattori linguistici a livelli superiori sul contesto locale. Il modello sovrasegmentale suppone invece che siano significative sole le relazioni che intervengono a livello di struttura sillabica: questo livello permette di codificare l'opposizione lessicale di parole omofone che si distinguono solo per la posizione dell'accento lessicale come in " 'altero/alt'ero", in particolare quindi codifica l'opposizione tonica/atona in maniera efficace ed efficiente. A questo scopo sarebbe quindi

necessaria l'informazione relativa alla posizione dell'accento di parola con l'indicazione della sillaba tonica.

Una seconda informazione essenziale è quella legata al numero di sillabe contenute nella parola: questo secondo parametro influenza il livello di elasticità e di comprimibilità da applicare alla struttura della parola nel suo insieme. Volendo fare un primo inventario dei "core factors" potremmo quindi elencare i seguenti:

- Fattori relativi alla identità fonica: segmento corrente, segmento/i precedente/i, segmento/i seguente/i;

- Fattori relativi all'accento: grado di prominenza nel discorso, accento lessicale;

- Fattori posizionali: segmenti nella sillaba, sillaba nella parola, parola nel gruppo fonologico, gruppo fonologico nell'enunciato;

Dieci fattori a livello sillabico, elencati per grado di rilevanza sono stati riportati per l'inglese da Campbell (ibid,1083) per un modello delle durate a livello sillabico:

a. numero di segmenti nella sillaba - sette livelli; b. indice di giuntura - quattro livelli; c. natura della rima - aperta/chiusa; d. distinzione parole funzione/contenuto; e. natura del picco - quattro classi; f. indice di accento - quattro livelli; g. tipo di piede - con la testa oppure no; h. numero di sillabe nel piede - sei livelli; i. posizione nella parola - quattro classi; k. posizione del gruppo fonologico nell'enunciato - quattro classi. Ci sarebbero in tutto 39 differenti tipi, alcuni dei quali verranno da noi ripresi più avanti.

4. Ipotesi fondate linguisticamente per un Modello Sillabico del Timing

In pratica si propone la sillaba come unità linguistica capace di fornire un miglior mapping tra il segnale acustico e la realtà fonica dei singoli fonemi. I vincoli fonetico-acustici imposti alla procedura di analisi sono molto semplici:

- la durata massima e minima di una sillaba della lingua in analisi viene derivata semplicemente dalla statistica indipendente dal contesto associata a ciascuna sillaba.

I vincoli fonologici sono legati all'indicazione delle seguenti informazioni prosodiche:

- l'accento primario e secondario e mancanza di accento; la posizione della sillaba nella parola; la posizione della parola all'interno del Gruppo Intonativo; la eventuale enfasi o salienza.

La durata di ciascuna sillaba viene predetta in funzione dei vincoli fonetici e fonologici:

$D_s = F(SI, L, S, P, E)$

A sua volta, la struttura sillaba è cruciale nei modelli predittivi in quanto permette di applicare restrizioni linguistiche al processo di parsing, alcune delle quali sono le seguenti:

a. tutte le parole legali di una lingua devono poter essere segmentate in sillabe legali della lingua stessa:

- l'inventario delle sillabe legali è dipendente dalla lingua;

- non tutte le lingue permettono lo stesso tipo di strutture sillabiche sia in termini di combinazioni CV che in termini di gruppi consonantici;

- i nessi consonantici in centro parola possono essere analizzati come esempi di:

- un gruppo legale di fine parola seguito da un gruppo legale di inizio parola;

- un gruppo legale di inizio parola preceduto da un gruppo legale di fine parola;

- le lingue si differenziano per il tipo di sillaba che può costituire un onset legale di inizio parola;

- le lingue si differenziano per il tipo di sillaba che può costituire un onset legale di fine parola;

Come risultato, è quindi cruciale stabilire il tipo di possibili combinazioni di suoni CVC, i.e. la fonotattica di una certa lingua; in aggiunta, da uno studio quantitativo del vocabolario è importante trovare le più frequenti posizioni di ciascuna sillaba all'interno della parola.

b. le sillabe accentate possono aiutare a rendere l'accesso lessicale più efficiente.

In questa prospettiva, l'accesso lessicale non viene più eseguito come una ricerca in una lista: piuttosto come una ricerca in un grafo in cui le parole sono state segmentate sulla base della struttura sillabica, numero di sillabe nella parola e tipo di sillaba accentata.

Se la variabilità prosodica fosse predicibile a livello segmentale, essa dovrebbe risultare "totalmente" dal contesto locale di adiacenza, cioè non dovrebbero esserci relazioni significative rilevabili a livello di sillaba o di parola che non siano anche rilevabili a livello segmentale. Come verificheremo dai dati in nostro possesso, in Italiano, ma non in Inglese, ci sono effetti di compensazione extra/intrasillabici che non si possono spiegare solo dal contesto di adiacenza segmentale.

La variabilità viene definita in termini di Compensazione di durata a livello sillabico. Per verificare la nostra ipotesi sperimentale abbiamo utilizzato una parola polisillabica "industriali" concentrando la nostra attenzione sul nesso consonantico "str" contenuto nella sillaba "stria". Abbiamo scelto una parola pentasillabica per cercare di isolare gli effetti che potrebbero venire dalle parole adiacenti, da quelli che sono in atto all'interno della parola stessa, evitando che i primi possano influire sull'andamento delle durate delle sue sillabe. In questo modo siamo abbastanza sicuri che le variazioni di durata della sillaba "stria" è ascrivibile soltanto all'andamento delle durate della parola che la contiene. Abbiamo poi utilizzato una definizione di sillaba per il segmento fonico "stria" anche se questo potrebbe essere fonologicamente segmentabile in vari altri modi, come ad esempio, "s-tria", oppure "stri-a" o ancora "s-tri-a": la decisione di trattare "stria" come segmento unico è venuta dopo aver accertato a livello fonetico la consistenza di pronuncia dello stesso segmento all'interno della parola in questione, che nel nostro caso, essendo stata letta da un parlante foneticamente competente come Luciano Canepari, offre una notevole garanzia di coerenza fonetica. La parola "industriali" è stata inserita nel contesto di 11 frasi da noi precedentemente studiate e riportate negli atti del GFS'97, svoltosi a Pisa¹. Nelle Tabelle sottostanti mettiamo in relazione la durata della "s", cioè del suono di onset sillabico e di inizio nesso consonantico della sillaba tonica, con la durata del suo nucleo, "ia"; inoltre

¹ Le frasi sono le seguenti:

- (A) Chi hanno detto che hanno intenzione di aiutare gli industriali?
- (B) Hanno detto che avrebbero aiutato i terremotati gli industriali?
- (C) Devono pagare gli industriali i decimali se vogliono la continuazione delle trattative.
- (D) Devono pagare i decimali gli industriali se vogliono continuare le trattative.
- (E) Gli industriali devono pagarli i decimali se vogliono che le trattative continuino.
- (F) Sono i decimali che gli industriali non vogliono pagare
- (G) Adesso anche gli industriali sosterranno che bisogna intervenire sulla politica dei redditi.
- (H) Gli industriali devono pagare i decimali se vogliono che le trattative continuino.
- (I) I decimali gli industriali devono pagare se vogliono che le trattative continuino.
- (L) I decimali gli industriali devono pagarli se vogliono continuare le trattative.
- (M) Adesso gli industriali sosterranno che bisogna intervenire sulla politica dei redditi.

mettiamo in relazione questi dati con quelli della durata globale della sillaba tonica, e con la durata delle sillabe atone complessive. In Tab.1, ogni occorrenza viene classificata a parte da una serie di parametri linguistici relativi alla funzione grammaticale del costituente sintattico contenente la parola nell'enunciato in analisi, alla posizione della parola nella frase, al tipo di enunciato da un punto di vista informativo sintattico-semantic.

Tab. 1 Abbinamento sigla e graduatoria frasi per carico informativo associato alla parola "industriali"

Si-gla	Tipo Enunciato	Posizione nell'Enunciato	Funzione Grammaticale	Rank
A	Interrog.-Princip-Completiva	Prepausa Fine enunciato	Soggetto Estraposto	A-10
B	InterrogativaPrincipale	Prepausa Fine enunciato	Soggetto Estraposto	B-11
C	Interrog.Princip-Secondaria	Fine costituente SV	Soggetto Posposto	C-2
D	Principale - Secondaria	Prepausa Fine principale	Soggetto Estraposto	D-9
E	Princip- Second / Oggetto Disloc.	Inizio principale/ 2°GI	Soggetto	E-8
F	Frase Cleft o Scissa	Inizio principale/ 1°GI	Soggetto Predicaz.	F-3
G	Frase Principale/ Completiva	Inizio principale/ 1°GI	Soggetto focalizzato	G-1
H	Principale - Secondaria	Inizio princip/1°GI-Ass.	Soggetto	H-6
I	Principale - Secondaria / Oggetto Focaliz.	Inizio principale/ 2°Gr.Int.	Soggetto	I-7
L	Principale - Secondaria /LCD	Inizio principale/ 2°GI	Soggetto	L-4
M	Principale/ Completiva	Inizio principale/ 1°GI	Soggetto	M-5

Se mettiamo in relazione le variazioni di durata della sillaba tonica con l'uso comunicativo e informativo della stessa all'interno dei vari enunciati, dovremmo poter verificare che la durata globale della parola dovrebbe aumentare in relazione diretta all'aumento della sua carica informativa. Dai nostri dati è possibile verificare questa prima ipotesi solo parzialmente: infatti, i valori più elevati della durata si hanno nell'enunciato C in cui la parola "industriali" costituisce un proprio Gruppo Intonativo con intonazione parentetica, costituendo una ripresa del costituente soggetto lasciato inespresso precedentemente. Invece nel caso della frase G, che contiene l'unico caso di costituente focalizzato quantificato, la parola e la tonica hanno la durata inferiore in assoluto ed è l'andamento di F0 a giocare un ruolo di messa in rilievo(vedi Delmonte, 1987). Come commenteremo nel dettaglio più in basso, ci sono però altri elementi fonetici che vengono sfruttati per far risaltare il contenuto informativo e che non sono immediatamente rilevabili. In concomitanza con la maggior durata complessiva si ha anche la maggior durata della coppia di vocali toniche IA nonché la maggior durata del cluster consonantico che ne costituisce l'onset. Complessivamente quindi in coincidenza della maggior durata della parola, anche la sillaba tonica nel suo insieme e nelle componenti interne ha durata maggiore confermando l'ipotesi formulata anche da E.Farnetani che l'unità ritmica di riferimento sia la parola, la quale funge da interfaccia con i livelli linguistici più astratti, concettuali e informativi, e da questo livello si scenda fino al livello segmentale passando attraverso quello sillabico.

Tab. 2 Valori durazionali assoluti della parola “industriali” in relazione alla graduatoria informativa

Sigla Frasi	Dur. Tot. Parola	Dur. Sill. Tonica	Dur. Tot. Atone	Dur. Clust.	Dur. IA	Dur.S	Dur.T	Dur.R
A-10	720	400	320	160	240	60	60	40
B-11	740	380	360	160	220	80	40	40
C-2	800	480	320	200	280	120	40	40
D-9	740	400	340	160	240	60	40	60
E-8	600	320	280	160	160	100	20	40
F-3	660	380	280	160	220	120	20	20
G-1	600	280	320	140	140	80	20	40
H-6	660	340	320	160	180	80	40	40
I-7	660	360	300	160	200	100	20	40
L-4	720	400	320	160	240	100	20	40
M-5	700	420	280	160	260	120	20	20
Dev.St.	61.55	53.26	25.73	14.01	42.72	22.40	13.75	10.79

Abbiamo quindi cercato di verificare come agiscano le spinte alla compensazione o al controllo all'interno della parola, variando la posizione e il suo contenuto informativo. Nel caso della frase G1, bisogna supporre che venga realizzata una diversa strategia da parte del parlante: come è possibile notare dalla Tab.3 sottostante, non esiste nessuna relazione tra la durata della tonica e il contenuto informativo e questo sembrerebbe contraddire il criterio di validazione fonetica del contenuto informativo. In realtà, un elemento altrettanto rilevante lo si ricava notando come la durata delle sillabe atone, nella frase G-1, sia superiore a quella della tonica, caso questo unico tra tutte le frasi. Facendo il rapporto tra la durata delle Atone e quella della Tonica si vede quindi che questo risulta essere squilibrato rispetto allo stesso valore nelle altre frasi, essendo l'unico a superare l'1%. Inoltre, le durate all'interno della sillaba tonica tra il cluster e il nucleo vocalico si eguagliano. Globalmente si può notare come a parte i casi unici delle frasi G-1 e C-2, nelle rimanenti frasi, la relazione tra il contenuto informativo e il rapporto Atone/Toniche sia invertito, e la sillaba tonica tende ad avere durata superiore a quella delle atone nelle frasi con minore contenuto informativo. Anche nei casi B-11 e D-9 entrano in gioco degli effetti secondari a dispetto della posizione prepausale del costituente in cui è contenuta la parola “industriali”, in funzione di Soggetto Estraposto, cioè come “Afterthought” e quindi come informazione già nota, che dovrebbero giocare a favore di una maggiore durata delle atone come avviene nella frase C.

Se guardiamo quindi il rapporto tra la durata totale della parola e la durata della sillaba tonica - in colonna 2, vediamo che solo in un caso si ha che la durata della sillaba tonica superi la metà della Durata Totale, nella frase G. In tutti gli altri casi, la durata della tonica varia in funzione inversa alla durata totale della parola ed è in buona relazione con l'andamento del rapporto Tonica/Atona. La frase G si dimostra quella più deviante per tutti

gli indici che abbiamo considerato salvo l'ultimo che riporta la proporzione tra la durata della S e quella complessiva del cluster STR. La durata della sillaba tonica varia in funzione inversa alla durata delle atone, indicando chiaramente un primo livello alto di compensazione nella unità linguistica PAROLA;

Tab.3 Graduatoria Frasi sulla Base del Rapporto Atone/Toniche

FraSi	Atone/Ton	DTonica/DTot	DClust/IA	DurS/Clust
G-1	1,14	0,46	1	0,57
B-11	0,94	0,514	0,88	0,50
H-6	0,94	0,515	0,72	0,50
E-8	0,87	0,53	1	0,62
D-9	0,85	0,541	0,66	0,37
I-7	0,83	0,545	0,80	0,62
L-4	0,80	0,55	0,66	0,62
A-10	0,80	0,55	0,66	0,37
F-3	0,73	0,57	0,72	0,75
C-2	0,66	0,60	0,71	0,60
M-5	0,66	0,60	0,61	0,75

5. Verifica Sperimentale delle Ipotesi

Dall'analisi dei dati comparativi abbiamo quindi cercato di evidenziare nelle ultime due colonne di Tab3, l'andamento e la variazione della durata della fricativa sorda /s/ all'interno del suo cluster consonantico |STR|, il quale a sua volta costituisce l'onset della sillaba tonica |STR|A|. Le ipotesi da verificare sono le seguenti:

- **IPOTESI SEGMENTALE:**

le variazioni di durata della /s/ in funzione del suo contesto locale, la /t/, sono in forte correlazione, hanno un andamento fortemente concorde con le altre variabili di livello astratto più elevato;

- **IPOTESI SOVRASEGMENTALI:**

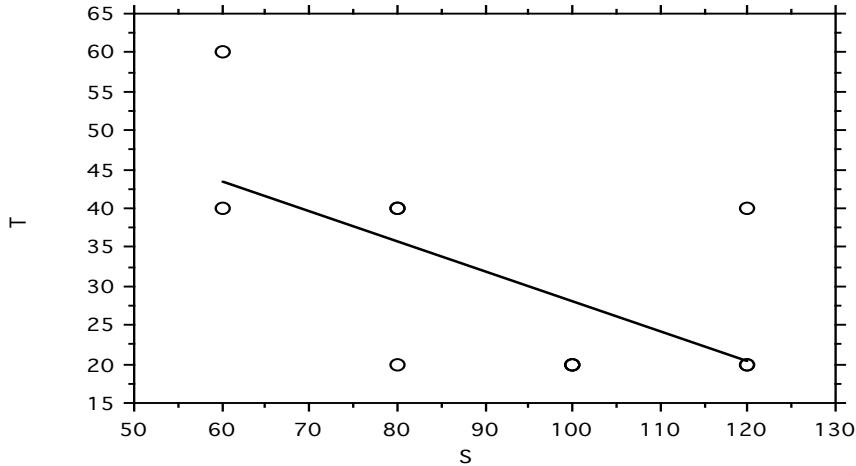
- ❖ le variazioni di durata sono in relazione al tipo di cluster e alla sua posizione all'interno della sillaba;
- ❖ le variazioni di durata sono in relazione a quelle della sillaba nel suo insieme;
- ❖ sono in funzione del rapporto esistente tra le variazioni della sillaba tonica e quelle delle sillabe atone;
- ❖ sono in funzione del ruolo sintattico-semantico riflesso dalla struttura fonologica: posizione della parola e costituente fonologico in cui è contenuta;
- ❖ sono in funzione del contenuto informativo.

Dall'analisi de dati abbiamo verificato le seguenti relazioni:

1. La durata della /s/ varia in relazione diretta alla durata del cluster di onset ed è in buona concordanza con la durata della sillaba in cui è contenuta ma non è riflessa, prevedibile dal suo contesto locale;
2. Difatti la durata della /s/ varia in relazione inversa alla durata della /t/ e della /r/;
3. Cioè all'interno del cluster di onset ci sono compensazioni, che potremmo definire di microlivello;
4. La durata del cluster di onset varia in relazione diretta con quella del nucleo sillabico;

Ma la ratio nucleo sillabico/cluster di onset varia in relazione inversa alla durata della

Tab. 5 Regression Plot
Dependent: T



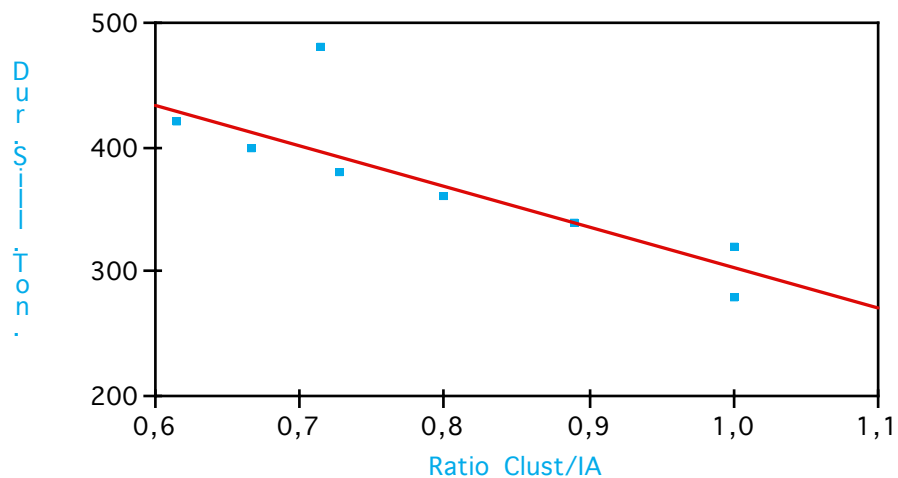
sillaba tonica nel suo insieme, cioè la variazione di durata della sillaba tonica mostra un livello di compensazione tra l'onset e il rhyme: al diminuire della durata della sillaba tonica il rapporto non resta fisso ma si ristrutturava a favore del cluster, che in quanto consonantico impone una maggior rigidità. Questo è mostrato dalla Tab. 6.

Nella Tab. 5 abbiamo riportato la linea di regressione dei dati di durata relativi alla "S" in rapporto a quelli della "T"; la significatività è alta essendo P-Value<0.0018.

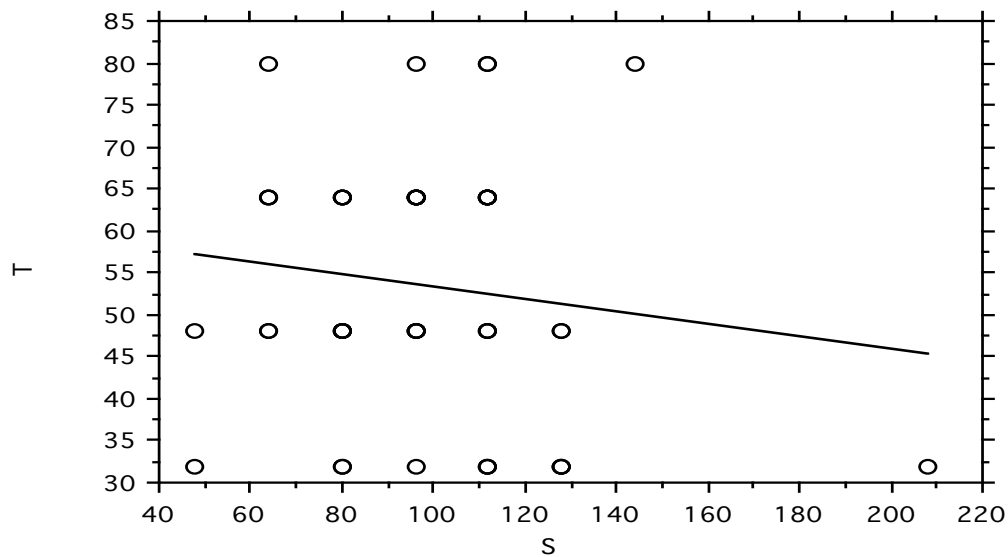
Per finire, abbiamo voluto verificare le nostre ipotesi su dati dell'inglese in nostro possesso e presentati nel lavoro di Bacalu & Delmonte, contenuto in questi Atti: abbiamo estratto tutte le occorrenze delle parole contenenti il cluster consonantico STR in relazione alla corrispondente parola italiana "industriali" e abbiamo trovato 52 occorrenze di "Industry/Industries/Industry's" e 24 occorrenze di "Industrial/Industrials". Abbiamo poi proceduto a classificare i dati su base informazionale. Non abbiamo però trovato casi di strutture non canoniche o focalizzate. I dati mostrano come le ipotesi di van Santen della "Directional Invariability" siano verificate per la lingua inglese, ad eccezione del rapporto "S" nella sillaba "DUS" e "T" nella successiva sillaba "TRY/IES". Infatti, anche nella lingua inglese sembra valere la stessa relazione inversa e quindi varrebbe un criterio di compensazione che però non potremmo definire di natura sillabica dato che i due suoni in questione appartengono a due sillabe separate. Riportiamo anche in questo caso la linea di regressione e i dati statistici relativi.

Anche per i dati dell'inglese la significatività è alta essendo il P-value<0.0001. È interessante notare come non sia possibile verificare lo stesso fenomeno nelle parole Industrial/Industrials che come struttura sillabica si avvicinano di più alla parola corrispondente italiana "Industriali": l'andamento delle durate di "S" e "T" in queste parole rispetta l'ipotesi della Directional Invariability.

Tab. 6 Lina di regressione negativa per Durata Sillabe Toniche vs. Rapporto Cluster Consonantico / Durata Nucleo sillabico



Tab.7 Regression Plot for words Industry/ies



6. Bibliografia

- [1] Bertinetto P.M.(1981), Ancora sull'italiano come lingua ad isocronia sillabica, *Scritti linguistici in onore di Giovan Battista Bellegrini*, Pisa, Pacini.
- [2] Breen A.P. (1995), A Simple Method of Predicting the Duration of Syllables, *Eurospeech'95*, 595-598.
- [3] Campbell W., S.Isard (1991), Segment durations in a syllable frame, in *Journal of Phonetics* 19, 37-47.
- [4] Campbell W. (1993), Predicting Segmental Durations for Accomodation within a Syllable-Level Timing Framework, *Eurospeech '93*, 1081-1085.
- [5] Delmonte R.(1987), The Realization of Semantic Focus and Language Modeling, in *Proceedings of the XIth ICPHS*, Vol.2, 24.1.
- [6] Delmonte R. F. Stiffoni, (1995), SIWL - Il Database Parlato della lingua Italiana, *Giornate Gruppo di Fonetica Sperimentale*, Trento, 99-116.
- [7] Farnetani E., S.Kori(1982), Lexical stress in spoken sentences: A study on duration and vowel formant pattern, *Quaderni del CSRF*, I, Padova, Progetto, 106-133.
- [8] Farnetani E., S.Kori(1984), Effects of syllable and word structure on segmental durations in spoken Italian, *Quaderni del CSRF*, III, Padova, Progetto, 143-187.
- [9] Farnetani E., S.Kori(1986), Effects of syllable and word structure on segmental durations in spoken Italian, *Speech Communication*, Vol.5, No.1, pp.17-34.
- [10] van Santen J., C.Shih, B.Moebius, E.Tzoukermann, M.Tanenblatt (1997), Multi-Lingual Duration Modeling, in *Eurospeech'97*, Rhodos, Vol.3, 2651-2654.
- [11] van Santen J. (1997), Prosodic Modeling in Text-to-Speech Synthesis, in *Eurospeech'97*, Rhodos, 19-23.
- [12] Vekas D., Bertinetto P.M.(1991), Controllo vs. Compensazione: sui due tipi di isocronia. in E.M.Caldognetto, P.Benincà(a cura di), *L'interfaccia tra fonologia e fonetica*, Padova, 155-162.