

# DUE REALTÀ LINGUISTICHE URBANE A CONFRONTO (ROMA E MILANO): QUALI PARAMETRI PROSODICI PER UN MODELLO PLAUSIBILE?

Elena Sardelli  
Università di Pisa  
[elenasardelli@infinito.it](mailto:elenasardelli@infinito.it)

## 1. SOMMARIO

L'esigenza di un algoritmo in grado di riconoscere automaticamente alcune caratteristiche fonetico-fonologiche della prosodia è di forte attualità ed interesse. I suoi impieghi andrebbero a facilitare il compito dei ricercatori in ambito forense, educativo, di trattamento automatico del linguaggio e di riproduzione vocale automatica. Il campo di ricerca si presenta particolarmente vasto, in quanto connesso alle problematiche note inerenti il riconoscimento automatico dei confini sillabici. Allo scopo di aggiungere un nuovo tassello per la realizzazione di metodologie sempre più specifiche, il fine che si intende perseguire con questa indagine è la produzione di un modello in grado di identificare automaticamente parlati afferenti a varietà diverse di italiano regionale. A questo scopo sono stati analizzati dialoghi di parlato semi-spontaneo (*corpus* CLIPS)<sup>1</sup> di parlanti maschi di Roma e Milano e sono stati riconosciuti dei "turni" e "clausole intonative" compiute. La trascrizione e l'annotazione della curva della frequenza fondamentale viene effettuata parallelamente attraverso due modalità distinte, poiché distinti sono i fini perseguiti: da una parte si identificano le sillabe prominenti (secondo la trascrizione ToBI tradizionale) e si annotano le caratteristiche dei *Pitch Accents*, dall'altra si riconoscono i parametri acustici in grado di marcare variazioni diatoniche. La prima, ovvero la notazione fonologica, permette di confrontare i dati e i risultati in oggetto con quanto emerso da indagini precedenti relative ad altre varietà italiane; l'analisi puramente fonetica dei fenomeni prosodici, permette invece di reperire le caratteristiche spettrali che possono essere adoperate per la costruzione di un modello automatico di riconoscimento.

In questa sede vengono presentati i risultati provenienti dall'analisi degli enunciati in forma interrogativa, per i quali si individuano i parametri idonei allo scopo.

## 2. OBIETTIVI E METODI

Compito primario di questo intervento è l'individuazione dei parametri acustici afferenti al livello prosodico della lingua italiana che possono essere utilizzati per la costituzione di un modello automatico di riconoscimento diatonico del parlante. Durante l'analisi e l'interpretazione del *corpus* analizzato, si è proceduto al trattamento del segnale sia in ambito strettamente fonetico che fonologico, adottando l'approccio della fonologia autosegmentale e la corrispondente trascrizione ToBI. L'indagine circa le caratteristiche funzionali, all'interno delle singole varietà locali, permette di ampliare la progressiva mappatura dei contorni melodici prodotti dai parlanti italiani e di confrontare i risultati con quanto già emerso da precedenti indagini.

Uno studio di questo genere si immette in un progetto di prospettive ed obiettivi più ampi, in quanto pone alcuni quesiti cruciali circa il rapporto, spesso controverso, tra

---

<sup>1</sup> Per gentile concessione del C.I.R.A.S.S., Università degli Studi di Napoli Federico II.

dialetti soggiacenti e varietà di italiano parlato. Una tale investigazione, dagli aspetti non solo sincronici, dovrebbe successivamente cercare di capire se sia possibile e legittimo parlare di prosodia standard (o standardizzata) per l'italiano contemporaneo parlato. Questo ultimo aspetto, spesso negato dalla letteratura, sembra invece presente non solo come esito di effettive elicitazioni, ma soprattutto come un elemento dell'immaginario collettivo, presente nella mente del parlante italiano che abbia ricevuto un'educazione scolastica di livello medio-alto sul territorio. Dalle ricerche finora condotte infatti sembra che, così come ricorda Berruto (1987 e 1988) per altri aspetti, le elicitazioni dei parlanti italiani possano distendersi su un *continuum* progressivo che si sposta dal polo marcato come [+diatopico] al polo [+standardizzato] (e non standard) considerando questa caratteristica come un punto verso il quale si tende, anche se mai completamente raggiunto. A questo proposito si definirà la prosodia del tratto [+standardizzato] come neutra o modale<sup>2</sup>, in stretto riferimento a quanto comunemente attribuito ad una certa qualità vocale, non caratterizzata da specifiche aggettivazioni.

Per l'analisi del parlato italiano delle due diverse aree, ovvero Milano e Roma, sonostate analizzate le registrazioni CLIPS<sup>3</sup> (*Corpora e Lessici di Italiano Parlato e Scritto*), gentilmente messo a disposizione dall'Università degli Studi di Napoli Federico II, ancor prima della loro possibile fruizione in rete. Del *corpus* in oggetto, sono stati analizzati i dialoghi *Map-Task*<sup>4</sup> e Test delle Differenze. L'indagine si è concentrata su parlanti maschi, in totale sei per ogni varietà.

Per la realtà milanese sono state prese in esame le mappe B01, A04, A01 e B04; per il parlato di Roma sono stati analizzati i dialoghi delle mappe A02, A03, A04 e B04. I parlanti sono tutti di età giovanile, compresa tra i 18 e i 24 anni per le registrazioni di Milano e tra i 18 e i 30 anni per i parlanti romani. Essenzialmente, tutti i locutori prescelti presentano un buon grado di spontaneità, fluenza, attenzione e spigliatezza, secondo quanto annotato dalle schede *CLIPS*.

I dialoghi sono stati in parte segmentati personalmente da chi scrive, mentre per la mappa B04 romana e A04 milanese, sono state seguite le segmentazioni già presenti. La catalogazione del segnale prevede la formazione di cartelle, suddivise, dall'alto verso il basso, per area geografica, tipologia di dialogo (*Map-Task* vs Test delle differenze), parlante, tipologia di frase (si veda di seguito), numero/codice della mappa di provenienza, parola riconosciuta prominente e numero progressivo. Ad esempio: il file "1-B04Mt cuori 17" è relativo ad un contesto interrogativo di tipo *y-n question*, proveniente dalla mappa B04 di Milano, (*t* si riferisce alla specifica del parlante, trattandosi di dialogo tra due

---

<sup>2</sup> L'aggettivazione si riferisce in letteratura ad un particolare *setting* fonatorio, non esattamente neutro, ma avulso da qualsiasi attribuzione di qualità vocale. In ambito prosodico si può ipotizzare una "prosodia modale", avulsa da caratterizzazioni locali.

<sup>3</sup> Il progetto CLIPS prevede la registrazione di materiale parlato e scritto della lingua italiana, realizzata in quindici aree geografiche diverse e precisamente nelle città di: Bari, Cagliari, Bergamo, Parma, Firenze, Genova, Catanzaro, Lecce, Milano, Napoli, Perugia, Palermo, Roma, Torino e Venezia. Al parlato letto e semispontaneo, registrato tramite la tecnica di *Map-Task* e riconoscimento delle differenze, si affiancano registrazioni di parlato radiofonico, televisivo e telefonico.

<sup>4</sup> Per *Map-Task*<sup>4</sup> si intende il metodo di registrazione di parlato semi-spontaneo, grazie al quale due parlanti sono chiamati ad interagire verbalmente, come in una situazione di gioco. I due partecipanti ricevono delle mappe, sulle quali sono stati disegnati dei percorsi e degli oggetti, in parte dissimili.

parlanti maschi), dove la parola percettivamente saliente è riconosciuta in *cuori* ed ha numero progressivo 17.

Per l'individuazione delle mosse conversazionali è stato seguito in linea di massima il criterio esposto da Carletta *et al.* (1996)<sup>5</sup>. La classificazione ad undici insiemi è stata ridotta a nove classi, concentrando l'analisi su tre/quattro sottospecie, come riportato di seguito.

È stata attribuita una specifica numerazione alle diverse tipologie frasali, così da poter suddividere le varie classificazioni in schemi riassuntivi:

**1** Domande *si-no*, del tipo: *ce l'hai il dado tu?*;

**1b** Domande *coda*, che presentano una struttura intonativa di frase molto simile alle affermative e che hanno in posizione finale l'elemento interrogativo, ad esempio: *tu non ce l'hai il dado, vero?* (all'interno di questa categoria si ritrovano anche i contesti dove la coda è sottintesa e dove la struttura prosodica della frase rimane costante);

**1x** Domande *sì-no*, a forte espressione di incredulità, come: *non c'hai il dado?!!!!!*;

**2** Domande *wh-*, ad esempio: *cosa c'hai dopo il dado?*;

**2b** Domande *wh-* con risposta *aut-aut*, ad esempio: *ma c'hai un dado o una pallina?*;

**3** Proposizione di tipo affermativo con apporto di nuova informazione;

**3b** Proposizione di tipo affermativo senza apporto di nuova informazione<sup>6</sup>;

**4** Imperative;

**5** Risposte.

Sono state realizzate alcune schematizzazioni grafiche per ogni *file* analizzato, che comprendono:

- numero del *file*;
- tipologia di frase;
- trascrizione ToBI;
- valore in semitoni dei movimenti prosodici nella porzione finale di frase;
- tempo, espresso in *ms*, del movimento identificato (quando possibile);
- valore minimo, massimo, iniziale e finale espresso in Hertz della curva in oggetto;
- media della curva;
- relative deviazioni standard;
- una o più vocali riconosciute come prominenti nella frase;
- posizione delle suddette vocali all'interno della frase (1 per la posizione iniziale di frase, 2 per la posizione mediana e 3 per la tonica finale di frase);
- durata e gradiente di modulazione in ST della vocale stessa;
- direzione del movimento che su di essa si realizza;
- valore di intensità ed espressione di tale valore rispetto al contesto (valore massimo, minimo, basso, alto);
- ripetizione di tali informazioni per il dominio della sillaba a cui si riferisce la vocale prominente;
- annotazioni varie ed eventuali.

Un esempio di quanto annotato è riportato di seguito<sup>7</sup>; la griglia riporta le informazioni generali relative alla curva di  $f_0$ : numero di *file*, tipologia frasale, trascrizione ToBI,

<sup>5</sup> Le tipologie frasali sono state successivamente ridotte, in virtù della qualità del materiale in esame.

<sup>6</sup> Durante le operazioni di analisi le due diverse classi, 3 e 3b sono state spesso accorpate.

modulazione del tono di confine (ST T%), modulazione associata al *pitch accent* prominente (ST T\*), valori minimi, massimi, iniziali e finali, media della curva (M.C.) e deviazione standard (D.S.).

| N°<br><i>file</i> | tipologia | ToBI | ST<br>T% | ST<br>T* | min | max | iniz | fin | M.C. | D.S. |
|-------------------|-----------|------|----------|----------|-----|-----|------|-----|------|------|
|-------------------|-----------|------|----------|----------|-----|-----|------|-----|------|------|

In riferimento alla vocale tonica di sillaba prominente sono stati misurati i valori di:

- qualità del timbro (analisi impressionistica);
- posizione all'interno della parola;
- durata in *ms*;
- modulazione ad essa allineata (in semitoni);
- qualità del movimento (discendente, ascendente, piatto, complesso);
- intensità globale;
- intensità relativa.

| V<br>[+prom] | posizione | durata | ST | Qlt mov. | intensità | Int.relativa |
|--------------|-----------|--------|----|----------|-----------|--------------|
|--------------|-----------|--------|----|----------|-----------|--------------|

In funzione di un dominio poco più ampio, sono state indagate le qualità della sillaba entro la quale la vocale è contenuta, ovvero la sua posizione, la sua durata, la modulazione in semitoni e la qualità del movimento:

| Posizione | durata | ST | Qlt movimento |
|-----------|--------|----|---------------|
|-----------|--------|----|---------------|

Lo schema così proposto permette di suddividere i vari contesti prosodici per varietà diatopica, locutore, contesto e tipologia di frase.

Data la situazione conversazionale, si riscontra un'alta concentrazione di contesti di tipo 3, ovvero contesti assertivi, con una percentuale quasi superiore ai due terzi dell'intero *corpus*. Gli enunciati presentano spesso una struttura sintattica poco complessa e ricca di sintagmi preposizionali, considerato il particolare valore informativo attribuito alla posizione degli oggetti e alla ricerca del percorso da seguire. Il sintagma può assumere spesso posizione iniziale o finale di frase, a causa di fenomeni di focalizzazione, come ad esempio in: *C'hai un dado tu poi sopra? Sopra hai un dado tu poi?*.

Molto frequenti sono gli esempi di enunciati non conclusi o sospesi, che provocano un allungamento, soprattutto vocalico, dei segmenti in posizione finale. Tale fenomeno sembra essere provocato non solo da ripensamenti, incertezze e cambi di struttura del discorso, tipici del parlato spontaneo, ma anche dalle caratteristiche fisiche del gioco stesso: dal momento che il percorso che deve essere effettuato non giunge a termine fino al suo "arrivo", la conversazione mantiene spesso un tono generale non conclusivo.

In alcuni *files* relativi a parlanti romani si registra la ripetizione di alcuni elementi in posizione finale di frase a scopo rafforzativo, che possono essere riconosciuti come una caratteristica diatopica: *Cioè (i)n pratica finito 'sto giro dee stelle ce sta l'arrivo 'n pratica; Va be', diciamo che comunque per arrivare alla torta li devo aggirare, li devo. A*

<sup>7</sup> La suddivisione in tre griglie distinte si ha ai soli fini grafici di stampa.

ciò si aggiungono alcune espressioni marcate, del tipo: *Comunque a bicicletta 'n se la filamo de pezza*.

Viceversa alcuni enunciati prodotti da parlanti milanesi presentano delle espressioni di ripetizione, diatopicamente marcate, del tipo: *Come è che è messa.....?*; *Dov'è che è.....*, e la postposizione del soggetto: *Stai guardando cosa tu?*.

L'analisi del segnale è stata condotta attraverso l'utilizzo del software PRAAT, scaricabile dalla rete: [www.praat.org](http://www.praat.org) e continuamente aggiornato dagli stessi autori. Questo ha permesso di applicare successivamente, in maniera del tutto automatica, la trascrizione INTSINT della curva intonativa attraverso simboli, come proposto da Hirst e Esperter (1993). Nella figura seguente se ne riporta un esempio:

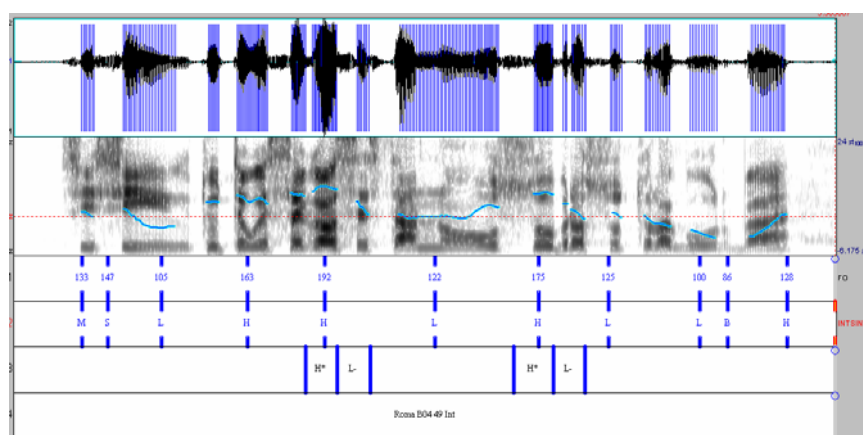


Figura 1. Oscillogramma, spettrogramma, griglia di trascrizione dei valori in Hz raggiunti dalla curva della frequenza fondamentale, simboli INTSINT (da trattamento automatico del segnale) e trascrizione ToBI (manuale) di alcuni punti strategici della frase. Sull'ultima riga si riporta il nome del file.

Come si osserva, alla raffigurazione del segnale (forma d'onda, *pulses*, spettrogramma, curva della frequenza fondamentale (in azzurro)) si aggiunge una griglia sottostante, che riporta i valori in Hz automaticamente estratti dal programma e l'interpretazione dei suddetti punti attraverso i simboli della trascrizione fonetica.

L'unità prosodica di riferimento è stata definita *clausola intonativa*, vale a dire un costituente avente indipendenza propria, ma non sempre sinonimo di sintagma intonativo. Il termine si differenzia da quello comunemente adottato di *sintagma intonativo* poiché vuole rendere esplicita la possibilità di ricoprire un dominio talvolta ridotto.

Per quanto riguarda la prominza e la sua individuazione, credo che per adesso non si possa far riferimento a programmi di riconoscimento automatico da parte di una macchina, visto che i tentativi che hanno dato alcuni risultati rilevanti (Tamburini, 2003), non sono facilmente reperibili e necessitano un complesso lavoro di *training* su una voce specifica, limitando così l'applicazione ad un parlato appositamente preparato.

Da un punto di vista strettamente acustico, una delle principali caratteristiche della marca di prominza è certamente la presenza di un movimento significativo sulla curva di  $f_0$  o il raggiungimento di un particolare livello tonale (alto o basso che sia), in grado di donare all'elemento salienza percettiva. Tali variazioni locali sono in grado di attirare

l'attenzione del parlante e, in una particolare posizione dell'enunciato rendono esplicite marche funzionali.

Ruolo determinante in questo senso viene rivestito dal parametro di *scaling* che risulta capace di recuperare differenze diatopiche entro la stessa tipologia frasale (contesti di domande *y-n questions* di Roma e Milano), ma anche variazioni funzionali entro la stessa varietà, come nel caso dell'espressione della continuità o la distinzione *chech-queries/y-n questions* per la varietà di Milano e di Roma.

In quest'ottica viene infatti interpretata anche la diversificazione tra contesti interrogativi e contesti assertivi: il contorno melodico di enunciati dichiarativi conclusivi presenta un andamento progressivamente discendente, che ha una sua relativa escursione melodica in senso inversamente proporzionale alla durata della frase; qualsiasi altra forma di modulazione del parlato, che investa la porzione finale, e soprattutto le ultime tre sillabe dell'enunciato, è distintiva per la marca di sospensione o interrogazione.

Sistemi linguistici diversi o varietà diatopiche diverse sembrano fare un uso variegato dei movimenti finali in contesti interrogativi e non continuativi, esprimendo comunque la stessa funzione attraverso la modulazione o il parametro di *scaling*.

### 3. UNA PRECISAZIONE SULLA PROMINENZA

La prominente, comunque venga espressa, sembra possa essere scomposta in tre aspetti principali, afferenti sia al campo fonetico che funzionale. Innanzitutto sembra appropriato trattare la prominente come fenomeno prosodico e non soltanto intonativo, poiché alcuni meccanismi del livello soprasegmentale sembrano essere strettamente inscindibili dal piano segmentale sottostante. Secondariamente, la prominente deve essere analizzata in base alle sue diverse funzioni e non rispetto a valori gerarchicamente assunti. Il grado di "salienza percettiva" (cfr. Marotta *et al.*, 2003 e 2004) viene attribuito grazie a mutamenti prosodici (ed in questo senso l'intonazione svolge un ruolo di primo ordine) in grado di attirare l'attenzione dei parlanti. Tali mutamenti sono capaci di apportare informazione nel messaggio comunicativo, sia sul piano fonologico stretto, che sul piano semantico e ritmico.

In riferimento al primo aspetto si intende per prominente funzionale o grammaticale quell'insieme di tratti acustici che permettono l'espressione chiara della marca tipologica dell'enunciato e la qualità dei suoi confini. Queste informazioni sono di solito concentrate sulla porzione finale del sintagma intonativo e sono in grado di distinguere frasi assertive ed interrogative, sospensive, conclusive etc.

Per prominente metrico-ritmica si intende invece la struttura melodica definita da un insieme di restrizioni e regole, strettamente correlate alle peculiarità ritmiche della varietà che, anche in funzione della distribuzione della durata segmentale, sono in grado di gestire la scansione del parlato, affinché questo non risulti piatto, innaturale e monotono. Le suddette regole sembrano agire in un ambito di "importanza" minore rispetto alla prominente funzionale e informativa, poiché non sono in grado di apportare alcun cambiamento significativo né sul piano pragmatico, né grammaticale in senso stretto. Questo particolare aspetto della prominente prosodica può essere definito *tempo* del parlato, in quanto presenta delle strutturazioni ed organizzazioni tendenzialmente simili a quanto avviene in musica; il *tempo* così espresso sfugge all'unità di misura precostituita ed è direttamente correlato a variazioni melodiche di un certo *range* significativo.

Per quanto riguarda la prominente informativa, lo scenario si presenta più complesso, poiché svariato ne è il dominio, come pure le modalità di realizzazione. Per prominente informativa si intendono tutti quei contesti in cui un determinato elemento frasale riesce

ad emergere dal resto del materiale segmentale, per svolgere una funzione essenzialmente pragmatica, nel processo di comunicazione. È in questa accezione che molto spesso il fenomeno coincide con quanto comunemente definito *focus*.

In realtà, per poter individuare i contesti di possibile compresenza dei fenomeni connessi con la prominenza è indispensabile puntualizzare quanto segue: la “salianza percettiva” viene realizzata in maniera diversa, a seconda della differente posizione dell’elemento bersaglio nella frase; essa può essere ricorsiva e pertanto presente più volte, entro uno stesso sintagma intonativo; l’elemento saliente può essere descritto attraverso una serie di tratti, non strutturati in senso gerarchico. Nel caso di focalizzazione larga di enunciato in forma assertiva (ovvero tutti quei contesti dove sia svariata la formulazione delle domande per le quali l’enunciato costituisce una risposta pertinente), l’assenza di un movimento saliente sull’ultima sillaba tonica di frase non può essere considerato sinonimo di mancata espressione prosodica; al contrario, il livello raggiunto da  $f_0$  in quella particolare porzione di frase assume valenza funzionale in quanto precedente un confine. La durata del segmento vocalico dell’ultima sillaba tonica si distingue per quantità dalle altre circostanti, evidenziandone la particolarità; il livello raggiunto da  $f_0$  in quella specifica porzione di frase assume valenza funzionale, in relazione al *range* di frequenze realizzate dal parlante. Diremo allora che la salienza informativa, in grado di marcare uno o più elementi all’interno della frase, non fa uso del parametro di *scaling* ogni qualvolta essa si concentri sull’ultima sillaba tonica di enunciato; lo stesso parametro risulta invece indispensabile ogni volta si voglia attribuire salienza e maggiore rilievo ad una specifica parola in differente posizione frasale. L’esplicitazione di tale marca attraverso forti modulazioni sulla curva di  $f_0$  o il raggiungimento di livelli particolarmente alti entro il PR, costituisce in letteratura, la forma più comune di focalizzazione.

In conclusione, potremmo affermare che prominenza informativa e salienza percettiva veicolata dal parametro di *scaling* non sono sempre sinonimi e che l’una comprende l’altra, unitamente a fattori strettamente grammaticali, come il lessico, la semantica, la sintassi.

Per il riconoscimento dei vari contesti e per la resa esplicita delle dinamiche di occorrenza, si potrebbe pensare ad una trascrizione in tratti che renda chiara la distribuzione della prominenza: il tratto [+info] sarà in grado di individuare i bersagli lessicali, sintattici, semantici e prosodici costituenti la nuova informazione dell’enunciato. Il tratto può essere ricorsivo, ma la sua presenza su un unico elemento della frase, permette di riconoscere un contesto a focalizzazione stretta o contrastiva<sup>8</sup>. Il tratto [+tempo] individua i punti della frase dove la scansione metrica-ritmica si ancora, secondo regole distinte essenzialmente su base diatopica. Unitamente al tratto potrebbe essere indispensabile la specificazione della presenza di alcuni confini intermedi. Infine, il tratto [+funzionale] permette di individuare le eventuali marche tipologiche e la presenza di confini.

#### 4. SULLA TEORIA AUTOSEGMENTALE

Ritornando adesso al trattamento dei dati in oggetto, per quanto concerne l’indagine fonologica, l’impianto teorico seguito è la teoria autosegmentale, con l’adozione della

---

<sup>8</sup> Potremmo aggiungere che, interpretando le indagini di Avesani (2003 e 2004), il riconoscimento delle diverse focalizzazioni potrebbe essere reso attraverso la posizione dell’elemento ricevente il tratto positivo: *focus contrastivo* in caso di posizione iniziale di enunciato, *focus ristretto* nel caso di posizione diversa.

trascrizione ToBI, anche se in una forma non del tutto tradizionale; si adotta ad esempio il suggerimento di Marotta e Sorianello (2001) per quanto riguarda la trascrizione dei toni di confine intermedi solo se non ridondante. Inoltre, seguendo Marotta (2003) e Walters (2003), per ciò che concerne la resa esplicita del grado di escursione melodica o *scaling*, che si realizza in taluni contesti, sono stati adottati i diacritici ↓ oppure ↑, ogni volta che si riscontra una escursione tonale significativa, discendente o ascendente<sup>9</sup>. Oltre a ciò, si sottolinea l'esigenza di indicare i valori dei livelli tonali [H, L] dei singoli accenti intonativi non tanto in rapporto al contesto locale, ma al *pitch range* dell'intera frase: infatti, dato il ruolo rilevante attribuito al parametro di *scaling*, una volta riconosciute alcune dinamiche prosodiche in grado di rendere esplicite alcune funzioni grammaticali, sembra opportuno riportare tali variazioni anche attraverso la trascrizione della curva della frequenza fondamentale.

L'adozione di entrambi i modelli, ovvero l'estrazione dell'estrazione automatica dei simboli tonali attraverso INTSINT e la trascrizione manuale ToBI non entrano in contraddizione teorica, in quanto le due trascrizioni assolvono, come già ricordato in precedenza, due funzioni diverse: da una parte si intendono identificare i parametri necessari all'individuazione delle marche diatopiche reperibili nell'indagine acustica per la costituzione di un modello di riconoscimento automatico, d'altro canto si rende opportuna una interpretazione fonologica dei fenomeni, per una diretta comparazione con altre varietà.

## 5. CONTESTI INTERROGATIVI: Y-N QUESTIONS

Per la tipologia interrogativa di *domande sì-no* sono state individuate alcune caratteristiche generali presenti in entrambe le varietà in esame, ovvero un movimento discendente-ascendente nella porzione finale di frase. L'andamento ha un'alta percentuale di distribuzione (84%). Nello schema che segue vengono riassunte le principali caratteristiche ed i parametri coinvolti:

---

<sup>9</sup>Il simbolo non intende, come nella versione proposta da Walters (2003), identificare escursioni aventi un preciso grado di valore in semitoni (Walters propone una freccia per *scaling* compresi tra i 3 e i 6 ST, e ↓↓ due frecce nel caso di escursioni superiori a 7 ST) ma la presenza di un'escursione melodica riconosciuta significativa per la varietà in esame.



## ➤ ***Y-n questions*** ...

### Roma e Milano: andamento finale **discendente-ascendente**

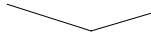
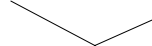
|           | Roma                                                                              | Milano                                                                             |
|-----------|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------|
| Movimento |  |  |
| ST        | 3,6      1,4                                                                      | 5,5      3,2                                                                       |
| ToBI      | H*+L      H%                                                                      | H+L*      H%                                                                       |
| PR        | 6,63                                                                              | 8,84                                                                               |
| Fo minimo | 107                                                                               | 93                                                                                 |
| Fo finale | 121                                                                               | 114                                                                                |

Figura 2 Schema riassuntivo dei valori medi delle caratteristiche prosodiche in grado di marcare la variazione diatopica.

La variazione diatopica si concentra sui seguenti punti:

- sul grado di escursione melodica media relativo ai due diversi movimenti (nello schema: ST);
- sulla qualità dell'ancoraggio tra il movimento di  $f_0$  e i segmenti (in trascrizione ToBI: H\*+L H% per Roma e H+L\* H% per Milano);
- sul *range* di frequenza medio (PR);
- sul valore medio raggiunto dai livelli più bassi ( $f_0$  min.);
- sul valore medio raggiunto in posizione finale ( $f_0$  fin.).

A ciò si aggiungono fenomeni interni di deaccentazione di alcune vocali, la collocazione media del punto più alto e più a destra della curva di  $f_0$  (che risulta all'interno dei confini di una vocale tonica costituente un PA per Roma ed esterna ad essa, nel caso della varietà milanese) e la presenza, per il parlato di Roma, di salienza percettiva, ovvero variazioni melodiche consistenti, su parole semanticamente scariche, come articoli e preposizioni, pronomi etc.

La trascrizione fonologica più frequente che si incontra nei contesti realizzati da parlanti romani è la seguente: H\*<sub>(n)</sub> H\*+L H%, dove <sub>(n)</sub> sta per un PA ricorsivo, come illustrato nella figura seguente.

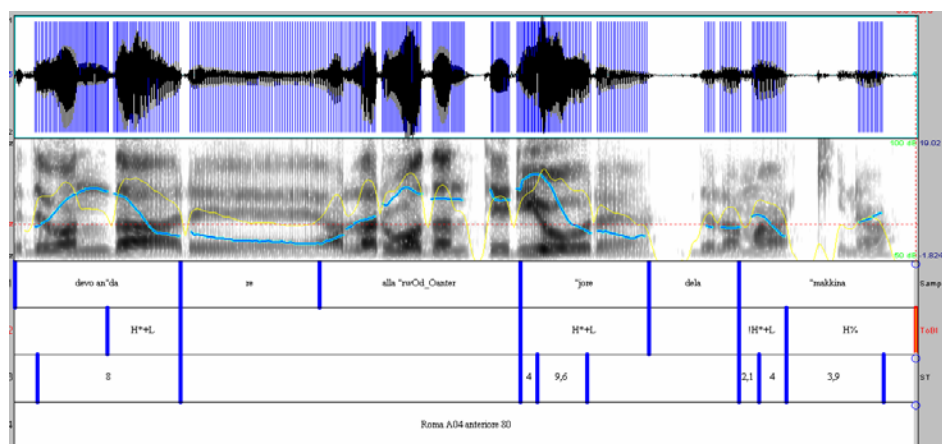


Figura 3 Oscillogramma, spettrogramma, griglia di trascrizione SAMPA, ToBI (manuale) di alcuni punti strategici della frase. Valori in semitoni dei movimenti salienti e nome del *file*, di una frase interrogativa prodotta da un parlante romano. 

I contesti interrogativi di richiesta di conferma possono presentare la coda sia in forma esplicita che sottintesa, ciò nonostante la struttura dei PA si mantiene costante (ovvero caratterizzata da  $H^*_{(n)} H^*+L$ ), presentando una variazione soltanto sul tono di confine che può essere L%, oppure H%. La diversa realizzazione del tono di confine non comporta alcuna variazione per la resa della tipologia interrogativa. Attraverso alcune prove di manipolazione del segnale, è emerso come solo la realizzazione di un tono basso, entro un livello particolarmente basso del PR del parlante, realizzato sull'ultima sillaba tonica della frase, sia in grado di trasformare il contesto da interrogativo ad assertivo. Nella figura seguente si riporta l'andamento di un contesto interrogativo (segnale originale in grigio) e l'andamento di un contesto dichiarativo (segnale modificato, in verde).

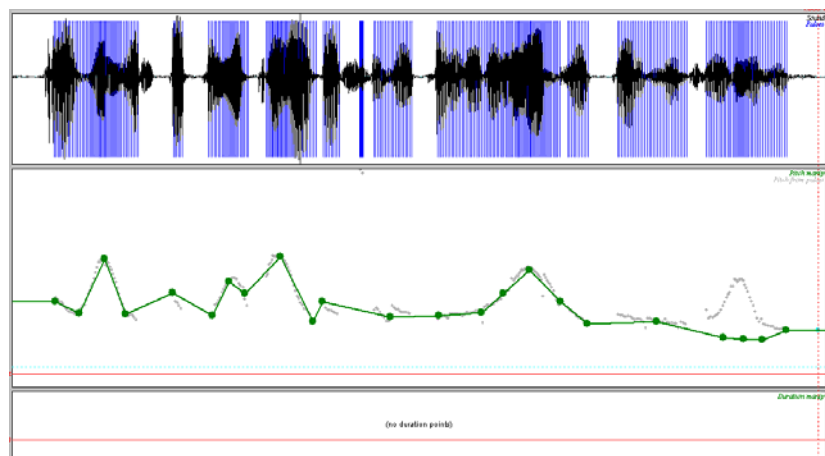


Figura 4 Oscillogramma, curva di  $f_0$  originale (in grigio) e curva di  $f_0$  manipolata (in verde) di una frase prodotta da un parlante romano, tipologia di domanda a richiesta di conferma con coda sottintesa.

Possiamo concludere che la marca interrogativa nella varietà romana di italiano, è caratterizzata da un movimento discendente-ascendente e se definiamo  $\alpha$  il primo movimento e  $\beta$  il secondo:

- $\alpha > \beta$  presso un livello mediamente alto del PR del parlante (vs contesti sospensivi);
- $\alpha$  non si immette mai all'interno un movimento discendente iniziato in precedenza, ovvero sul percorso progressivo di *baseline* (vs contesti dichiarativi);
- se  $\alpha$  è fortemente modulato e associato alla sillaba nucleare come  $H^*+L$ , esso è sufficiente per esprimere la marca interrogativa (vs contesti dichiarativi).

Per quanto concerne il parlato milanese, la trascrizione fonologica dell'andamento è:  $H^*_{(n)} H+L^* H\%$ , anche se non sono rari casi di difficile interpretazione, in cui il movimento discendente inizia e finisce all'interno della sillaba tonica. La presenza di suddetti esempi induce spesso a ipotizzare la trascrizione proposta da Marotta (2000), ovvero  $(H+L)^*$ .

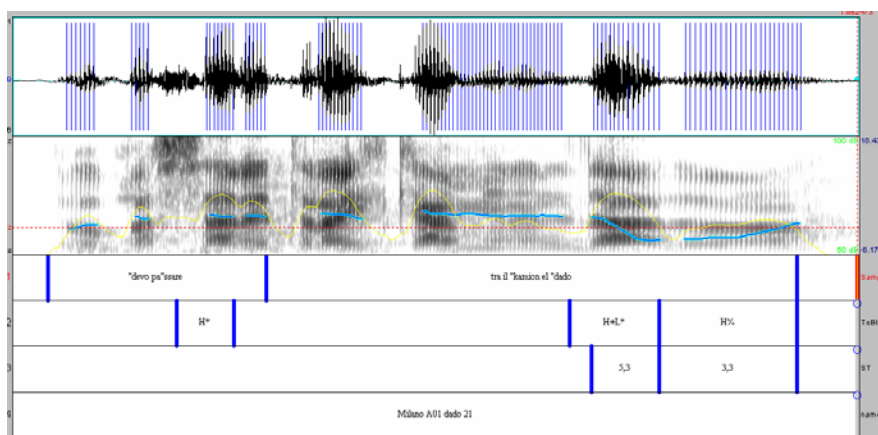


Figura 5 Oscillogramma, spettrogramma, curva di  $f_0$  e dell'intensità, griglia di trascrizione SAMPA, ToBI di alcuni punti strategici della frase. Valori in semitoni dei movimenti salienti e nome del *file*, di una frase interrogativa prodotta da un parlante milanese. 

Sul piano fonetico, si assiste a:

- consistente allungamento della durata della vocale finale e sua realizzazione bitonale;
- presenza del valore di  $f_0$  [+alto] e [+a destra] al di fuori dei confini di vocale in PA;
- escursione melodica tipica, nei movimenti  $\alpha$  e  $\beta$ , decisamente maggiore, rispetto agli esempi di parlato romano e, in media, pari a 5,5 ST per  $\alpha$  e 3,2 ST per  $\beta$ ;
- PR più esteso e spostato verso il basso (i valori medi alti rimangono invariati);
- valori medi finali più bassi;
- caratteristiche spettrali di alcuni segmenti “bandiera” come [ae] in *barchetta*, *forchetta* etc...

## 6. VOCALI BITONALI

La realizzazione di vocali bitonali, ovvero vocali che si allineano a due movimenti distinti della curva di  $f_0$ , distintamente percepiti, sembra non essere una caratteristica distintiva unicamente del parlato di Milano. Ciò nonostante sembra che all'interno di questa varietà venga attribuito al fenomeno sia un valore fonetico che funzionale. Il dato, particolarmente interessante, merita alcune precisazioni: nel parlato di Milano si ritrovano contesti come *Si uniscono tutte le chiazze a te??* (si veda la figura 6), oppure come *Ah, c'è l'ombra?* (si veda la figura 7), dove alla vocale finale si associano due movimenti sulla curva della frequenza fondamentale, separatamente percepiti. Per il primo caso, riportato nella figura 6, si può ipotizzare che la bitonalità sia condizione necessaria al riempimento del tono di confine H%; essendo infatti vocale tonica finale ed essendo questa comunemente associata da un tono L\*, semplice o complesso che sia, si rende indispensabile una prosecuzione del materiale segmentale e melodico, per la buona realizzazione del contorno intonativo necessario alla marca interrogativa.

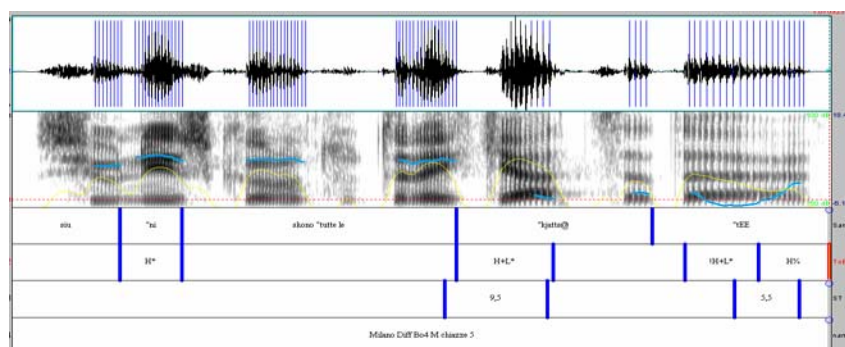


Figura 6 Oscillogramma, spettrogramma, curva di  $f_0$  e dell'intensità, griglia di trascrizione SAMPA, ToBI (manuale) di alcuni punti strategici della frase. Valori in semitoni dei movimenti salienti e nome del *file*, di una frase interrogativa prodotta da un parlante milanese: *Si uniscono tutte le chiazze a te?* 

La suddetta necessità non sembra presentarsi invece nel contesto seguente:

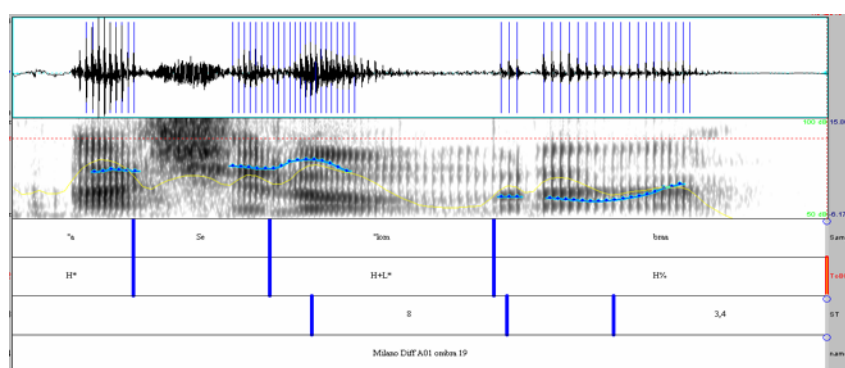


Figura 7 Oscillogramma, spettrogramma, curva di  $f_0$  e dell'intensità, griglia di trascrizione SAMPA, ToBI (manuale) di alcuni punti strategici della frase. Valori in semitoni dei movimenti salienti e nome del *file*, di una frase interrogativa prodotta da un parlante milanese: *Ah, c'è l'ombra?* 

Premesso che casi come riportato in figura 6 sono presenti anche nella varietà romana e che casi come nella figura 7 sembrano essere diatopicamente marcati come [+milanese], resta aperta la questione relativa al rapporto di causa-effetto che intercorre tra struttura segmentale e suprasegmentale, ovvero: la bipartizione tonale, implica un processo di riarticolazione segmentale? Per il momento sembra impossibile dare una risposta certa, soprattutto a causa del numero relativamente basso di occorrenze presenti nel *corpus* in oggetto, che non permette di fare generalizzazioni aventi valenza statistica. Ciò nonostante il dato sembra particolarmente interessante. I primi risultati che emergono da tentativi realizzati attraverso la manipolazione del segnale, sembrano essere incoraggianti: sembra infatti di poter affermare che a Milano, la presenza vs assenza di accento sulla vocale finale non costituisca un limite all'occorrenza del fenomeno. Il contrario pare invece essere tipico della varietà romana che fa talora uso del bitonalismo, e forse anche della riarticolazione, come riempimento del tono di confine, nel caso di vocale tonica finale. A Milano si hanno

vocali atone finali dove: a) il bitonalismo induce la percezione di due segmenti distinti, nonostante l'unità segmentale; b) il bitonalismo induce la riarticolazione del segmento; quest'ultimo caso sembra essere particolarmente marcato in senso diatopico.

## 7. CONTESTI INTERROGATIVI: DOMANDE CODA

I contesti interrogativi di richiesta di conferma con esplicitazione della coda, sono caratterizzati dal contorno  $H^*_{(n)} H^*+L H\%$ , per la varietà romana, dove  $H\%$  risulta fortemente modulato. Prove di manipolazione del segnale hanno evidenziato come una diversa modalità di ancoraggio dei movimenti della curva di  $f_0$ , ovvero un PA  $H+L^*$  modificherebbe il senso della frase, da interrogativo a dichiarativo. Per la varietà milanese invece gli stessi contesti sono stati trascritti come  $H^*_{(n)} L^* H\%$ .

Dall'analisi delle caratteristiche fonetiche della prosodia dei contesti analizzati, suddivisi per varietà diatopica e tipologia frasale, si può concludere che sono stati per adesso individuati alcuni parametri (escursione melodica, modalità di ancoraggio, estensione del PR generale,  $f_0$  minimo,  $f_0$  finale, media della curva) in grado di costituire informazione sufficiente alla costituzione di un modello di riconoscimento automatico, grazie al quale le medie statistiche vengono poste in relazione ai dati provenienti da un singolo enunciato. Gli stessi parametri possono essere utilizzati anche per l'identificazione di alcune marche linguistiche all'interno della stessa varietà, come risulta dallo schema seguente:

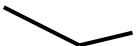
| <i>Yes-no Questions</i> |                                                                                     |                                                                                     | <i>Check queries</i> |                                                                                      |                                                                                       |
|-------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|----------------------|--------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
|                         | Roma                                                                                | Milano                                                                              |                      | Roma                                                                                 | Milano                                                                                |
| Mov.                    |  |  | Mov.                 |  |  |
| ST                      | 3,6 1,4                                                                             | 5,5 3,2                                                                             | ST                   | 6,1 6,7                                                                              | 3,7 1,8                                                                               |
| ToBI                    | $H^*+L H\%$                                                                         | $H+L^* H\%$                                                                         | ToBI                 | $H^*+L H\%$                                                                          | $L^*+L H\%$                                                                           |
| PR                      | 6,63                                                                                | 8,84                                                                                | PR                   | 9,11                                                                                 | 6,64                                                                                  |
| Fo min                  | 107                                                                                 | 93                                                                                  | Fo min               | 126                                                                                  | 97                                                                                    |
| Fo fin                  | 121                                                                                 | 114                                                                                 | Fo fin               | 139                                                                                  | 115                                                                                   |
|                         |                                                                                     |                                                                                     | M.C.                 | 128                                                                                  | 110                                                                                   |

Figura 8 Rappresentazione schematica delle principali caratteristiche prosodiche dell'andamento dei contorni intonativi finali di frase per le due diverse varietà diatopiche e tipologie frasali.

Nonostante vengano riportati in questa sede solamente i risultati provenienti dall'analisi di due varietà diatopiche, ovvero il parlato di Roma ed il parlato di Milano per i contesti interrogativi, l'individuazione dei suddetti parametri permette già di individuare due

distinte strategie che possono essere adottate per la costituzione di un modello di riconoscimento automatico.

In prima istanza, il calcolo delle medie statistiche dei valori riportati in precedenza, relativi alla varietà, sesso del parlante e tipologia frasale, permetteranno un confronto diretto con enunciati naturali, dei quali verranno calcolate le distanze. I valori di ogni singola frase possono cioè essere posti a diretto confronto con le medie dei valori emersi dall'analisi del *corpus* qui presentato. A questo computo, che si basa sulla misurazione di parametri distribuiti sull'interno enunciato, potrà seguire un'indagine specifica su alcuni punti della frase definiti strategici.

Se infatti, come abbiamo visto le due varietà si distinguono per comportamenti intonativi differenti, all'interno dei domini temporali dei PA riconosciuti, si potrà procedere alla etichettatura manuale di alcune porzioni frasali, sulle quali si concentreranno poi i calcoli automatici specifici. Ad un'indagine ad ampio raggio che ricerchi le caratteristiche generali dell'enunciato, sarà affiancata una ricerca puntuale sulle caratteristiche prosodiche di alcune porzioni di frase, come i confini, le sillabe nucleari e prominenti. Il progetto, al momento in corso (cfr. Sardelli, in stampa), mira all'applicazione di un modello di riconoscimento automatico su un numero sempre maggiore di varietà diatopiche e tipologie frasali.

In conclusione, sembra importante ricordare come i tentativi di sintesi del segnale vocale hanno dimostrato che nel passaggio dalla forma interrogativa alla forma dichiarativa (e viceversa) di un enunciato sono necessari i seguenti elementi:

- 1) la presenza di un contorno tonale significativo;
- 2) la corretta associazione (ancoraggio) tra andamento intonativo e struttura segmentale, secondo la specifica variante diatopica;
- 3) la riduzione o estensione temporale dei singoli segmenti;
- 4) il corretto posizionamento del contorno, entro le linee di confine del PR del parlante, secondo la specifica tipologia frasale.

Da un'indagine pilota sembra di poter ipotizzare infine che la presenza di una particolare qualità vocale ovvero [+ creaky] costituisca indice sociofonetico del parlato fortemente marcato in diatopia. La particolarità potrebbe essere indotta dal *phonation setting* di alcuni segmenti e dalla cospicua concentrazione di livelli tonali bassi.

## 8. BIBLIOGRAFIA

Avesani, C., 2003. La prosodia del *focus* contrastivo. Un accento particolare?. In G. Marotta e N. Nocchi (a c. d.) *La Coarticolazione. Atti delle XXIII Giornate di Studio del Gruppo di Fonetica Sperimentale*, Pisa, 157-168.

Avesani, C.; Vayra, M., 2004. Focus ristretto e focus contrastivo in italiano. In F. Albano-Leoni, F. Cutugno, M. Pettorino, R. Savy (a c. d.) *Il Parlato Italiano. Atti del Convegno Nazionale di Napoli*, 13-15 Febbraio 2003, Napoli: D'Auria (CD-ROM).

Berruto, G. 1987. *Sociolinguistica dell'italiano contemporaneo*. Roma: La Nuova Italia Scientifica.

Berruto, G., 1988. Le varietà del repertorio. In A. Sobrero (a c. d.) *Introduzione all'italiano contemporaneo. La variazione e gli usi*. Roma-Bari: Laterza. 3-36.

Carletta J., Izard A., Izard S., Kowtko J., Doherty-Sneddon G., Anderson A. 1996, *HCRC Dialogue structure coding manual*. In *Human Communication Research Center, Technical Report 82*, University of Edinburgh, June 96, pp.1-27.

- Hirst, D.; Espesser, R., 1993. Automatic modelling of fundamental frequency. In *Travaux de l'Istitut de Phonétique d'Aix*, 15, 71-85.
- Marotta, G., 2000. Allineamento e trascrizione dei toni accentuali complessi: una proposta. In *Atti delle X Giornate di Studio del Gruppo di Fonetica Sperimentale*, Napoli, Dicembre 1999, 139-149.
- Marotta, G., 2003. L'illusione prosodica. In *Atti del Convegno di Studi in memoria di Tristano Bolelli*, Pisa, 28-29 Novembre, 237-258.
- Marotta, G.; Sorianello, P., 2001. La teoria autosegmentale nell'analisi dell'intonazione interrogativa in due varietà di italiano toscano (Lucca e Siena). In *Atti del XXXIII Convegno della Società di Linguistica Italiana*, Roma: Bulzoni, 177-204.
- Marotta, G.; Sardelli, E., 2003. Sulla prosodia della domanda con soggetto postverbale in due varietà di italiano toscano. In P. Cosi, E. Magno Caldognetto, A. Zamboni (a c. d.) *Voce, canto, parlato. Studi in onore di F. Ferrero*, 205-212.
- Marotta, G.; Calamai, S.; Sardelli, E., 2004. Non di sola lunghezza. La modulazione di F0 come indice sociofonetico. In A. De Dominicis *et al.* (a c. d.) *Atti delle XIV Giornate del Gruppo di Fonetica Sperimentale*, 1, 210-215.
- Sardelli, E., in stampa. Quali parametri prosodici per il riconoscimento automatico del parlante?. In *Atti della Conferenza Nazionale sul Trattamento Automatico della Lingua*, Roma, 9-10 marzo 2006.
- Tamburini, F., 2003. Automatic prosodic prominence detection in speech using acoustic features: an unsupervised system. In *Proceedings of Eurospeech 2003*, Geneva, 129-132.
- Walters, R. J., 2003. On the intonation of a South Wales "Valleys accent" of English. *Journal of the International Phonetic Association*, 33, 2, 211-238.