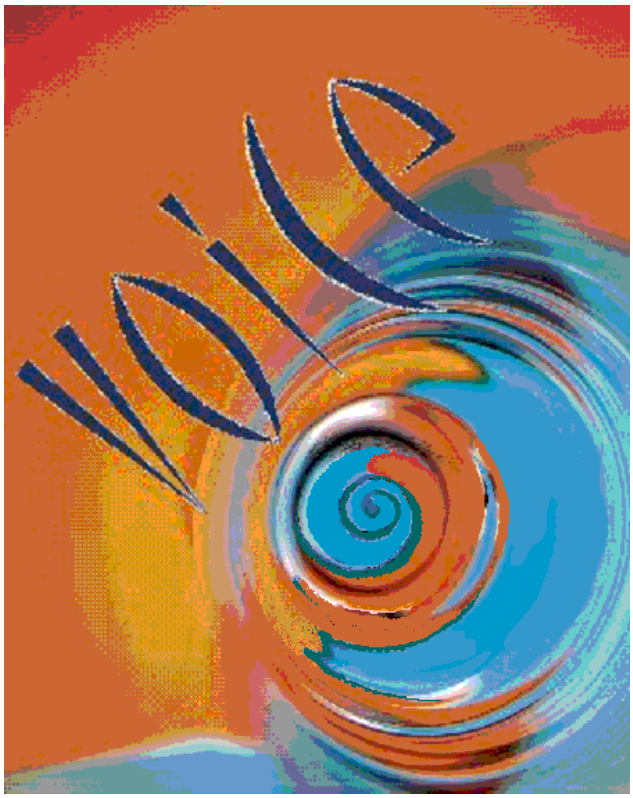
Associazione Italiana di Scienze della Voce



2° Convegno Nazionale

ANALISI PROSODICA
teorie, modelli e sistemi di annotazione

RIASSUNTI DELLE COMUNICAZIONI

Dati empirici e modelli fonologici: il caso dei dittonghi discendenti in sillaba chiusa.

Giovanni Abete

Friedrich-Schiller-Universität Jena

I dittonghi discendenti in sillaba chiusa costituiscono da sempre un problema nel quadro linguistico romanzo. Se la romanistica tradizionale si è potuta (e dovuta) accontentare di trattarli come eccezioni, oggi essi pongono questioni ben più gravi per i recenti modelli di fonologia neo-generativista. In particolare, la fonologia autosegmentale rischia di vedere contraddette le sue predizioni sulla struttura della sillaba, relativamente al numero di segmenti che possono occupare la posizione di coda.

Nel tentativo di far quadrare i dati con la teoria, si è diffuso negli ultimi anni un atteggiamento "negazionista", sostanzialmente volto a dimostrare che i dittonghi in questione o non sono segmenti realmente lunghi, o non si ritrovano in sillabe realmente chiuse. In questo contesto la fonetica ha assunto un ruolo del tutto ancillare e subordinato all'ottenimento di dati che si adattassero al modello fonologico dominante.

Partendo da una prospettiva completamente diversa, questo contributo intende, in primo luogo, mostrare le evidenze fonetiche dell'esistenza di dittonghi discendenti in sillaba chiusa nel dialetto di Pozzuoli e, in secondo luogo, proporre una linea di ricerca differente, che possa portare a una comprensione integrata tanto delle regolarità, quanto delle "anomalie".

La percezione dell'accento lessicale: un test sull'italiano a confronto con lo spagnolo.

Iolanda Alfano

Università degli studi di Salerno

Sebbene le conseguenze della variazione e covariazione di uno o più parametri acustici sul piano percettivo siano tutt'altro che lineari, grazie a vari studi sperimentali, è stato dimostrato, ad esempio, che la frequenza fondamentale svolge un ruolo dominante nella percezione dell'accento lessicale sia in inglese che in francese (Fante *et al*. 1991, Hasegawa & Hata 1992).

Uno studio effettuato sulla lingua spagnola dall'Equipe di Fonetica della UAB (Università Autonoma di Barcellona, Llisterri *et al*. 2004) attraverso la sistematica alterazione dei parametri ritenuti responsabili del processo percettivo (durata, frequenza e intensità), ha messo in luce come, nonostante a livello acustico la durata risulti il parametro predominante nella realizzazione dell'accento lessicale, sul piano percettivo f_0 svolga un ruolo necessario, sebbene non sufficiente.

Il lavoro che presentiamo nasce dall'idea di realizzare un'analoga indagine sull'italiano, condividendo, fatte le dovute distinzioni, le metodologie di costruzione del corpus e di conduzione del test adottate nel lavoro sullo spagnolo.

Partendo da un *corpus* di 45 trisillabi con struttura CVCVCV, pronunciati in isolamento e suscettibili di diversa accentazione (es: *légami-legàmi, lavòro-lavorò, sémino-semìno-seminò,* ecc), abbiamo calcolato durata e frequenza dei segmenti vocalici di ciascuna parola. Associando ad una parola i valori di tali parametri propri di un altro profilo accentuale (es: proparossitona con valori di parossitona, parossitona con valori di ossitona), abbiamo realizzato stimoli sintetici (costituiti da parole e non parole) proposti in un test di percezione (prova di identificazione e di discriminazione). Con l'intento di studiare gli effetti sulla percezione della variazione del fattore temporale, della frequenza e dei valori congiunti di ambedue i parametri, abbiamo manipolato i suddetti correlati prima isolatamente e poi in concomitanza. Parlanti nativi di italiano avevano il compito di

a) indicare la sede accentuale;
b) classificare coppie di stimoli come "uguali" o "diversi" in merito alla collocazione dell'accento.

In accordo con quanto indica la letteratura (Bertinetto 1981), i risultati mostrano, per l'italiano, la netta predominanza del parametro durata nell'individuazione della sede accentuale in parole e non parole; la manipolazione della sola frequenza sembra, invece, del tutto insufficiente a provocare un cambio di schema percepito. La percentuale dei casi in cui i soggetti avvertono la modifica è, poi, sensibilmente superiore in conseguenza della covariazione di durata e frequenza.

Sembra importante osservare, inoltre, che la percezione dipende, in parte, anche dallo schema accentuale considerato: risulta più nettamente e facilmente individuabile il cambio di sede nella coppia proparossitona-parossitona (*légami-legàmi*) rispetto alla coppia parossitona-ossitona (*lavòro-lavorò*).

BERTINETTO, P. M. (1981), *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici*, Accademia della Crusca, Firenze.

FANT, G., KRUCKENBERG, A., NORD, L. (1991), "Durational correlates of stress in Swedish, French and English", *Journal of Phonetics*, 19, 3-4: pp. 351-365.

HASEGAWA, Y. & HATA, K. (1992), "Fundamental frequency as an acoustic cue to accent perception", *Language and Speech*, 35, 1-2, pp. 87-98.

LLISTERRI, J. - MACHUCA, M. - de la MOTA, C. - RIERA, M. - RÍOS, A. (2004). "La percepción del acento léxico en español", in *Homenaje a Antonio Quilis*, Universidad de Valladolid – Consejo Superior de Investigaciones Científicas – Universidad Nacional de Educación a Distancia (en prensa).

La prosodia degli enunciati dichiarativi e interrogativi in tre aree dialettali dell'Italia centro-meridionale (Abruzzo, Basilicata e Campania)

Francesco Avolio[1] & Antonio Romano[2]

[1]*Dip. di Storia e metodologie comparate - Università dell'Aquila, Italia;*
[2]*Dip. di Scienze del Linguaggio - Università di Torino, Italia*
avolio@cc.univaq.it, antonio.romano@unito.it

Questo contributo si situa nell'ambito del progetto *AMPER* e riporta principalmente i risultati di un'inchiesta preliminare.

Dato che lo scopo del progetto è quello di procedere al confronto tra le strategie prosodiche di diverse varietà dialettali e di varianti regionali delle lingue romanze, è stata predisposta la raccolta di materiali sonori in alcune aree d'Italia con l'obiettivo di saggiarne la variabilità geoprosodica.

La ricerca in corso è il risultato di un'interessante esperienza di collaborazione che si prefigge l'analisi delle configurazioni intonative degli enunciati dichiarativi e interrogativi in diverse aree dell'Italia centro-meridionale. In particolare in questa fase siamo in grado di pubblicare i primi risultati su due varietà lucane (Aliano e San Mauro Forte) che abbiamo avuto modo di studiare in rapporto con alcuni campioni raccolti anche in Abruzzo (Tèramo) e in Campania (Ravello).

I nostri dati concordano con quelli delle descrizioni dialettologiche tradizionali (com'è facile immaginare, elementi di discriminazione sono ad esempio i fenomeni di dittongazione e frangimento dei nuclei vocalici e le caratteristiche di isocronismo, accentuale vs. sillabico). Globalmente, possiamo ritenere che i materiali raccolti presentino delle caratteristiche interessanti nell'opposizione tra dichiarativa e interrogativa e nella diversificazione tra le soluzioni intonative dialettali rappresentate. In particolare, nel nostro caso i tentativi di sintesi della sola prosodia prototipica, seppure limitatamente ai tipi di produzione osservati, ci consentono di affinare l'attenzione nei riguardi di quelle relazioni melodiche, temporali e dinamiche che si rivelano salienti ai fini della caratterizzazione geoprosodica. I numerosi test condotti per definire le condizioni di riproduzione dei tratti prosodici tipici delle aree studiate permettono infatti di migliorare la descrizione degli stereotipi con l'inclusione di elementi solitamente ignorati o esclusi dalle ricerche in questo settore (ad es. i contributi di strascichi, scoppi etc.).

Anche se allo stato attuale resta anche da effettuare un confronto con i risultati di cui disponiamo grazie a studi sull'intonazione già pubblicati (soprattutto per il napoletano, il barese e l'abruzzese centro-meridionale), i nostri dati, ancora parziali e provvisori, offrono però uno spunto importante per saggiare anche in Italia la presenza di varietà dialettali la cui prosodia non sembra poter essere limitata allo studio di profili melodici e schemi ritmici.

Un'interessante ipotesi di lavoro è quella che ci spinge a studiare le modalità d'inclusione, in una concezione generale di prosodia, dei contributi delle riduzioni e dei frangimenti, così come la presa in conto di elementi fricativi ed esplosivi che a volte sopperiscono funzionalmente alla perdita di elementi vocalici.

AVOLIO, Francesco: "Gli «indicatori geografici»" come fonte per gli studi dialettali: esempi lucani", In M.T. Greco (a cura di), *Gli "indicatori geografici" della Basilicata nord-occidentale nei territori delle Comunità montane del Marmo e del Melandro*, Napoli, Università degli Studi "L'Orientale", Quaderni di *AION* (nuova serie - 6), 27-44, 2003.

AVOLIO, Francesco: "Ma nuje cómme parlamme? Problemi di descrizione e classificazione dello spazio dialettale «campano»", *Romance Philology*, 54, 1-28, 2000.

ROMANO, Antonio: "Utilisation des données *AMPER* pour une description de la variation linguistique : tests de perception et contrôles statistiques", *Géolinguistique*, no. 3 (hors série: Projet *AMPER - Atlas Multimédia Prosodique de l'Espace Roman*), 39-64, 2005.

ROMANO, Antonio, LAI, Jean Pierre, & ROULLET, Stefania: "La méthodologie *AMPER*", *Géolinguistique*, no. 3 (hors série: Projet *AMPER - Atlas Multimédia Prosodique de l'Espace Roman*), 1-5, 2005.

Qualità di voce e indebolimento consonantico: un caso di correlazione in *Scouse*?

Marlen Barth e Massimiliano Barbera

Università di Pisa

L'inglese di Liverpool, noto anche come *Scouse*, presenta caratteristiche peculiari sia a livello segmentale, sia a livello paralinguistico. Legato alla *working class* di origine irlandese, è nato come varietà sociolinguisticamente marcata, ma nel corso del XX secolo ha subito una sensibile diffusione diastratica, tanto da filtrare nell'uso linguistico dell'intera comunità cittadina. Uno degli aspetti più rilevanti è costituito dall'indebolimento consonantico che investe tendenzialmente le occlusive sorde, producendo come *output* segmenti affricati o fricativi; in alcuni contesti fonosintattici si perviene addirittura alla cancellazione. Questi stessi processi di lenizione fanno capo ad un generale quadro di *lax voice* che compromette l'occlusione completa in fase articolatoria. L'analisi acustica segmentale (Marotta & Barth 2005) ha in effetti messo in evidenza una elevata incidenza percentuale di allofoni leniti corrispondenti ai fonemi occlusivi.

In letteratura si registrano, in relazione allo *Scouse*, scarsi e sparsi riferimenti all'impiego di qualità vocali non modali come fattore di marcatezza sociolinguistica: in modo particolare, la diffusa velarizzazione di tutti i segmenti consonantici sembra interagire con il *setting* fonatorio, determinando un tipo vocale generalmente qualificabile come voce adenoidale.

In questo lavoro si svilupperanno le considerazioni sopra indicate, prendendo in esame un *corpus* costituito da un campione di parlato spontaneo prodotto da sei locutori nativi (tre femmine e tre maschi). L'indagine sperimentale mira a valutare la *voice quality*, prendendo spunto dalle etichette interpretative di Laver (1980). Si applicano quindi schemi e parametri di analisi tali da consentire la registrazione dei correlati spettroacustici dei tipi fonatori individuati. In coerenza con i dati relativi alla frequenza statistica della lenizione, sembrano potersi cogliere alcuni elementi di differenziazione relativi al genere anche per quanto concerne la caratterizzazione vocale.

Epstein, M. A. (2002), *Voice quality and prosody in English.* (A dissertation submitted in partial satisfaction of the requirements for the degree of Ph.D. in Linguistics), University of California, Los Angeles.

Gordon, M., Ladefoged, P. (2001), *Phonation types: a cross-linguistic overview.* Journal of Phonetics, 29, pp.383-406.

Honeybone, P. (2001). Lenition Inhibition in Liverpool English. In: *English Language and Linguistics,* 5/2: 213-249.

Knowles, G. (1974). *Scouse, the urban dialect of Liverpool*. Unpublished PhD dissertation, University of Leeds.

Knowles, G. (1978). The nature of phonological variables in Scouse. In: Trudgill, P. (ed.) *Sociolinguistic Patterns in British English*. London: Edward Arnold, 80-90.

Laver, J. (1980), *The phonetic description of voice quality.* Cambridge University Press, New York.

Marotta, G. (2004), *La lenizione nell'inglese parlato a Liverpool (Scouse).* Atti del 4° congresso di studi dell'Associazione Italiana di Linguistica Applicata, Guerra Edizioni, Perugina.

Marotta, G. & Barth, M. (2005), *Acoustic and sociolinguistic aspects of lenition in Liverpool English.* In corso di pubblicazione.

Focus Contrastivo nella periferia sinistra della frase:
un solo accento, ma non solo un accento

*Giuliano Bocci e **Cinzia Avesani

**Università di Siena; **ISTC-CNR Padova*

L'approccio cartografico alle strutture sintattiche ha individuato nella periferia sinistra della frase la proiezione di Focus Contrastivo (FC) specializzata nel segnalare alle interfacce esterne (fonologia e sistema concettuale) l'articolazione della frase in Focus Contrastivo – Presupposizione (Rizzi 1997, 2004). Lo scopo di questo lavoro è di indagare se esistano caratteristiche prosodiche stabilmente associate dal sistema fonologico all'attivazione della proiezione sintattica di FC e di fornirne una descrizione all'interno della teoria autosegmentale-metrica, nell'intento di raccogliere indizi circa i processi di interfaccia tra sintassi, prosodia e semantica. La prosodia del Focus Contrastivo (principalmente in posizione post-verbale) in italiano e nelle sue varietà regionali è stata negli ultimi anni più volte oggetto di indagine sotto diversi rispetti: per gli aspetti intonativi si vedano Avesani e Vayra (2004); Grice, D'Imperio, Savino e Avesani, (2005); D'Imperio (1999; 2002); Gili Fivela (2002) e Gili Fivela e Savino (2003); per gli aspetti di costituenza prosodica, Frascarelli (2000) e Nespor e Guasti (2002). Nel presente lavoro abbiamo voluto prendere in considerazione sia le caratteristiche ritmiche sia le caratteristiche intonative delle frasi con FC nella periferia sinistra, raccordando l'indagine prosodica a precisi assunti circa le strutture sintattiche soggiacenti (Rizzi, 1997; Bocci, in preparazione).

La presente ricerca, che si configura come un *case study*, ha analizzato un corpus costituito da 160 frasi lette da una unica parlante senese all'interno di piccoli dialoghi sceneggiati. Le frasi raccolte comprendono frasi con FC in posizione iniziale di frase, frasi con Topic con dislocazione clitica a sinistra (Cinque, 1990) e frasi in Focus Ampio. Dalle frasi del corpus, segmentate a livello di fonema per tutta la loro lunghezza e sottoposte ad analisi percettiva, sono state estratte le informazioni riguardanti durata, intensità, valori di F0 dei target tonali individuati e allineamento dei target alla struttura segmentale. L'analisi ha mostrato che l'attivazione della proiezione di FC nella periferia sinistra impone sistematicamente a tutta l'articolazione FC-Presupposizione proprietà prosodiche ben distinte da quelle delle frasi in FA o con Topic. Le caratteristiche prosodiche delle frasi in FA o con Topic, assunte come base di confronto per quelle in FC, sono risultate del tutto coerenti con quanto attestato in letteratura (*inter alia,* Avesani, 1990). Il confronto delle durate del materiale ai confini dei costituenti prosodici in frasi con FA e FC ha permesso di stabilire la costituenza ritmica delle frasi in FC, evidenziando l'assenza di un confine di sintagma intonativo (IP) a destra del costituente in FC, diversamente da quanto proposto da Nespor e Guasti (2002) per il FC post-verbale e, limitatamente a certe condizioni sintattiche, da Frascarelli (2000). Quanto emerso dalle analisi ritmiche è del tutto coerente con i risultati dell'indagine sulla costituenza intonativa delle frasi in FC. Le caratteristiche dell'allineamento dei target tonali LH che costituiscono l'accento intonativo del costituente focalizzato sono risultate non del tutto trasparenti per la trascrizione fonologica (Arvaniti et al., 1998, 2000; Marotta, 2000; Frota, 2000 ed in particolare Gili Fivela, 2002). Tuttavia, considerata la diversa distribuzione di H rispetto alle caratteristiche della sillabica tonica, e considerata la nostra analisi dell'accento intonativo nel primo costituente in frasi in FA, proponiamo di trascrivere l'accento del FC come L+H*. Conseguenza immediata dell'adozione di tale trascrizione è la possibilità di trattare in modo omogeneo l'accento del FC nelle varietà fiorentina e senese, sia che esso occupi la periferia sinistra della frase sia che si collochi in posizione finale di enunciato (per il fiorentino: Avesani e Vayra, 2004).

In conclusione, le proprietà prosodiche di frasi con FC non sembrano poter essere derivate composizionalmente a partire da quelle di frasi in FA, poiché i due tipi di frasi presentano differenze non solo locali. Allo stesso tempo, la differenza tra FA e FC non è riconducibile unicamente al luogo di assegnazione della testa nucleare di IP, poiché gli andamenti degli accenti nucleari sono tipologicamente opposti: ascendente per FC vs. discendente per FA.

Arvaniti A., Ladd, R. D. e Mennen, I. 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics*. Vol. 26, pp. 3-25.

Arvaniti A., Ladd, R. D. e Mennen, I. 2000. What is a starred tone: Evidence from Greek. In Broe M. & Pierrehumbert J. (a cura di), *Papers in Laboratory Phonology V*. Cambridge: Cambridge University Press.

Avesani C. 1990. A contribution to the synthesis of italian intonation. *Proceedings ICSLP 90 - 1990 International Conference on Spoken Language Processing*. Kobe, Japan, vol. 1, pp. 834-836.

Avesani, C. e Vayra, M. 2003. Broad, narrow and contrastive focus in Florentine. In M. J. Solé, D. Recasens e J. Romero (a cura di), *Proceedings of the 15th .International Congress of Phonetic Sciences (*Barcellona, 3-9 agosto 2003), Vol. 2*,* Causal Productions Pty Ltd, pp. 1803-06.

Bocci, G. in preparazione. Some remarks on the cartography of the left periphery of the clause in Italian. *Rivista di Grammatica Generativa*.

Cinque, G.1990. *Types of A'-Dependencies*. Cambridge (Mass.): MIT Press.

D'Imperio, M. 1999. Tonal structure and pitch targets In J. Ohala (a cura di), *Proceedings of the 14th International*. San Francisco, vol. 3, pp. 1757–1760.

D'Imperio, M. 2002. Italian intonation: an overview and some questions. *Probus.* Vol. 14, n. 1, pp. 37-69.

Frascarelli, M. 2000. *The syntax-phonology interface in focus and topic constructions in Italian*. Dordrecht: Kluwer Academic Publishers.

Frota, S. 2000. *Prosody and Focus in European Portuguese. Phonological phrasing and intonation*. New York: Garland.

Gili Fivela, B. 2002. Tonal alignment in two Pisa Italian peak accents. In *Proceedings of the Speech Prosody 2002 Conference*. Aix-en-Provence: Laboratoire Parole et Langage, pp. 339-342.

Gili Fivela, B. e Savino, M. 2003. Segments, syllables and tonal alignment: A study on two varieties of Italian. In *Proceedings of the 15 th International Congress of Phonetic Sciences* (ICPhS'03). Barcelona, pp. 2933-2936.

Grice M., D'Imperio M., Savino M., Avesani C. 2004. Strategies for intonation labelling across varieties of Italian. In Jun S.-A. (a cura di), *Prosodic typology*. Oxford: Oxford University Press, pp. 55-83.

Ladd, D.R., Mennen, I. e Schepman, A. 2000. Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America*. Vol. 107, pp. 2685-2696.

Marotta, G. 2000. Allineamento e trascrizione dei toni accentuali complessi: una proposta. In *Proceedings of X Giornate di Studio del Gruppo di Fonetica Sperimentale*. Napoli, pp. 139-149.

Nespor, M. e Guasti, M. T. 2002. Focus to stress alignment. *Lingue e Linguaggio*. Vol. 1, pp. 79-106.

Rizzi, L. 1997. The Fine Structure of the Left Periphery. In L. Haegeman (a cura di), *Elements of Grammar*, 281-337. Dordrecht: Kluwer.

Rizzi, L. 2004. (a cura di). *The Structure of CP and IP – The Cartography of Syntactic Structures vol.2.* New York: Oxford University Press.

Costruzioni verbali con elemento dislocato a destra. Una o due unità intonative? Esempi di italiano L1 ed L2.

Elisabetta Bonvino

Università degli Studi di Roma Tre

Gli studi sulla dislocazione a destra in italiano sono poco numerosi, generalmente non si basano sull'analisi prosodica di dati di parlato e non comprendono la dislocazione a destra del soggetto. Nel caso del soggetto infatti si riscontrano delle difficoltà a riconoscere la dislocazione a causa della mancanza del clitico di ripresa. Questo lavoro prende le mosse da miei precedenti studi che hanno attribuito alla prosodia un ruolo cruciale nella distinzione tra soggetto posposto e soggetto dislocato a destra, identificando le caratteristiche sintattiche, semantiche e prosodiche delle costruzioni comportanti il soggetto dislocato a destra.

Il lavoro qui presentato si articola in due fasi. Nella prima fase l'analisi prosodica viene estesa a tutti i tipi di dislocazione a destra (dislocazione dell'oggetto diretto e indiretto) riscontrati in un corpus di parlato italiano L1 di due ore e 15 minuti.
L'analisi strumentale prende in considerazione i seguenti parametri prosodici:

I. andamento generale del contorno intonativo (ascendente o discendente o piatto);
II. valori di F0 e durata delle toniche del verbo, dell'eventuale elemento non dislocato che segue il verbo e dell'elemento dislocato;
III. valori delle eventuali postoniche dell'elemento dislocato e del verbo;
IV. presenza di eventuali pause tra verbo ed elemento dislocato e loro durata;
V. durata totale dell'enunciato.

Gli obiettivi di tale studio sono due:

I. Verificare se le caratteristiche prosodiche identificate per le dislocazioni del soggetto abbiano una portata più ampia, che comprenda i diversi tipi di dislocazione a destra;
II. Partendo dai dati, discutere l'appartenenza o la non appartenenza dell'elemento dislocato all'unità intonativa di cui fa parte il verbo.

Nella seconda fase del lavoro, i risultati saranno confrontati con i dati di italiano L2 tratti da un corpus di parlato non preparato di apprendenti di livello avanzato. Il corpus è raccolto nell'ambito della Certificazione di Italiano L2 "IT" dell'Università di Roma Tre. Questo confronto ha lo scopo di estendere ulteriormente l'analisi della dislocazione. L'interesse viene dato dal fatto che l'*input* che i parlanti non nativi ricevono e che influenza la loro interlingua è ricco di strutture dislocate. Il raggiungimento di un livello avanzato di competenza presuppone quindi la presenza nell'interlingua degli apprendenti di questi fenomeni sintattici, pragmatici e prosodici.

The ‘logogenicity’ of singing.
Observations about the extemporary poetry of Southern Sardinia.

Paolo Bravi
pa.bravi@tiscali.it

Università di Siena/Cagliari/Perugia
Dottorato di ricerca “Metodologie della ricerca etno-antropologica”

What happens when a verbal line built with a defined metrical form (that is a verse) is performed as a ‘sung verse’? Does it conserve something, from the ‘melodical’ and prosodical point of view, of the intonological structure that is present when it is simply said (or read)? Or the musical form superimposed slaughter the traits of the ordinary linguistic form? In the present work, more versions of the same poetical text, in sung and in read form, were compared. The oral texts analysed - on the basis of a double theoretical and methodological scenery, the intonological one and the musicological one - come from the extemporary sung poetry of Southern Sardinia, particularly from the genre known as *a muttettus* singing, that is the most important, musically various and complex form, and which is performed only by professional improvisers. The aim of the analysis was threefold: first, to give a contribution to the wider matter of the interrelation between forms and codes which we observe in sung verse; second, to give a contribution to the integration of knowledge of verbal intonation through the analysis of sung forms; third, to point out what is, in a structural perspective, the ‘logogenic’ character often noticed in oral sung poetry.

Indagine preliminare sul vocalismo della lingua mòoré (Gur, Niger-Congo)

Silvia Calamai & Pier Marco Bertinetto

Scuola Normale Superiore di Pisa

La lingua mòoré ("Lingua dei Mòosé") è un idioma appartenente al gruppo Gur, parlato nel Burkina Faso da circa il 50% della popolazione (i Mòosé). Il suo vocalismo orale si compone di sette timbri vocalici, lunghi e brevi, per un totale di quattordici fonemi, il cui numero si raddoppia considerando il contrasto orale / nasale: una vocale centrale /a/, due vocali medie /e o/, quattro vocali alte che – sia per la serie breve sia per la serie lunga – presentano una opposizione nel tratto [±ATR]: i ɪ u ʊ.

L'indagine intende fornire un primo quadro descrittivo del vocalismo orale di una lingua mai indagata con mezzi sperimentali, e intende inoltre osservare i correlati acustici del contrasto relativo alla posizione della radice della lingua. In particolare, l'analisi mira a valutare se i correlati acustici del contrasto [±ATR] reperiti in altre lingue del continente africano (vedi tra gli altri Fulop *et al.* 1998; Guion *et al.* 2004; Local & Lodge 2004) siano presenti anche nella lingua mòoré, soprattutto per quanto concerne i valori delle prime due formanti e i valori di durata. Interessa osservare se il contrasto [±ATR] ha qualche effetto sul dominio temporale; se anche nel mòoré le vocali [+ATR] hanno valori più elevati della prima formante rispetto alle controparti [-ATR]; se infine esiste una qualche sistematicità nella variazione relativa alla seconda formante.

Il materiale sonoro analizzato è costituto da parlato letto (parole isolate e brevi frasi), prodotto da tre soggetti maschi di giovane età e registrato *in loco*.

Canu G. 1976 *La langue mò:rē. Dialecte de Ouagadougou (Haute-Volta). Description synchronique*, Paris.

Fulop S.A., E. Kari, P. Ladefoged 1998 "An acoustic study of the tongue root contrast in Degema vowels", *Phonetica*, 55: 80-98.

Guion S.G., M.W. Post, D.L. Payne 2004 "Phonetic correlates of tongue root vowel contrasts in Maa", *Journal of Phonetics*, 32, 4: 517-542.

Local J., K. Lodge 2004 "Some auditory and acoustic observations on the phonetics of [ATR] harmony in a speaker of a dialect of Kalenjin", *Journal of the IPA*, 34, 1: 1-16.

Kinda J. 1999 *Moore langue vivante*, Université de Ouagadougou.

Sistemi vocalici in diatopia

G. Clemente°, R. Savy*, S. Calamai^

° *Università degli studi di Napoli "Federico II"; *Università degli studi di Salerno; ^ Laboratorio di Linguistica, SNS Pisa*

Il lavoro intende dare un contributo alla descrizione dei sistemi vocalici di matrice regionale attraverso analisi contrastive in chiave diatopica, e contestualmente ampliare l'indagine sui fenomeni di riduzione vocalica, alla luce di nuove procedure sperimentali.

In un precedente lavoro (Savy et alii 2005), è stato fruttuosamente collaudato un tipo di elaborazione e rappresentazione dei sistemi vocalici e di misura della riduzione, che fornisce dati quantitativi e statistici delle distribuzioni vocaliche. In particolare si cercherà di osservare la modalità della distribuzione dei timbri vocalici 'definiti' e la misura della riduzione in diversi sistemi regionali, al fine di descrivere e rintracciare somiglianze e differenze tra le varietà diatopiche.

Il materiale analizzato (estratto da parlato semispontaneo dei *corpora* Avip-Api e Clips) comprende circa 3.000 produzioni vocaliche di aree regionali diverse: Palermo, Napoli, Lecce, Pisa, Roma, Milano, Torino. L'indagine si concentra, in questa fase, sui fenomeni di riduzione 'non strutturale' (Savy&Cutugno, 1997), nel confronto, quindi, tra vocali 'definite'e timbri classificati come 'indefiniti'.

L'analisi contrastiva dei sistemi regionali di partenza mette in luce alcune differenze fondamentali nella morfologia del sistema come nelle tendenze distributive di ciascun'area vocalica.

L'analisi dei fenomeni di riduzione evidenzia significative differenze nella direzione del fenomeno, in parte imputabili alle difformità intrinseche ai sistemi 'definiti', presi come riferimento. L'indefinitezza si manifesta, tuttavia, come tendenza 'dispersiva' dei dati, in maniera molto simile nelle diverse varietà diatopiche, non riguardo alla direzione, quanto all'entità.

Lo Prejato, M., Clemente, G., Savy, R., 2004, "Su alcuni aspetti della riduzione vocalica nella varietá napoletana", in A. De Dominicis, L. Mori, M. Stefani (acd): Costituzione, gestione e restauro di corpora vocali. Atti delle XIVe Giornate di Studio del G.F.S., Università della Tuscia (Viterbo), 4-6.XII.2003 Roma, Esagrafica, pp. 183-188.

Savy, R. & Cutugno, F. (1997), Ipoarticolazione, riduzione vocalica, centralizzazione: come interagiscono nella variazione diafasica?, in *Atti delle VII Giornate del Gruppo di Fonetica Sperimentale – AIA 1996*, Napoli, 14-15 novembre (F. Cutugno, editor), Roma: Esagrafica, 177-194.

Savy, R., Lo Prejato, M., Clemente, G., 2005, "Per una caratterizzazione e una misura della riduzione vocalica in italiano", in Atti del I Convegno Nazionale AISV, *Misura dei parametri. Aspetti tecnologici ed implicazioni nei modelli linguistici*, Padova, 2-4 dicembre 2004, CDRom.

Studio sull'apprendimento del consonantismo spagnolo da parte di italofoni.

Emanuela Colonna, Barbara Gili Fivela

Università Degli Studi Di Lecce

L'apprendimento delle caratteristiche fonologiche e fonetiche di una lingua straniera rappresenta una difficoltà notevole per qualsiasi parlante, tanto che spesso anche chi dimostra un'ottima competenza linguistica su vari livelli, ad esempio quello morfologico e sintattico, viene comunque identificato come straniero, proprio perché il suo eloquio è caratterizzato da elementi che appartengono ad un sistema fonologico diverso da quello della lingua di apprendimento.
In questo lavoro sono state prese in considerazione le influenze del sistema consonantico italiano nell'apprendimento dello spagnolo. Dopo una dettagliata descrizione delle caratteristiche fonologiche e fonetiche delle consonanti italiane e spagnole, sono stati individuati i contesti nei quali gli italofoni possono incontrare maggiori difficoltà nel produrre foni e fonemi dello spagnolo Castigliano. L'identificazione di tali contesti ha permesso di evidenziare i fattori potenzialmente rilevanti per lo studio - ad esempio la posizione della consonante nella sillaba - e di individuare i criteri utili per la scelta di un insieme di parole da far produrre oralmente ai soggetti intervistati. Il corpus di indagine è composto da 114 parole bersaglio, scelte in base alla posizione segmentale e prosodica degli elementi consonantici. Le parole, presentate in isolamento, sono state lette per tre volte da dieci soggetti individuati all'interno della Facoltà di Lingue dell'Università di Lecce: due parlanti madrileni, appartenenti alla classe docente, e otto studenti che differivano relativamente al livello di conoscenza della lingua spagnola (ad esempio per numero di anni di studio, durata dei soggiorni all'estero, ecc.). L'indagine condotta ha fornito spunti di riflessione piuttosto interessanti. Ciò che emerge chiaramente è la variabilità sia nelle produzioni dei parlanti spagnoli madrelingua, che degli italofoni. Per quanto riguarda questi ultimi, i risultati dello studio mostrano, secondo le aspettative, che l'influenza della lingua materna nelle produzioni in lingua straniera è spesso individuabile, che è più o meno evidente a seconda del livello di competenza linguistica acquisita, e che si registra maggiormente in relazione ad alcuni dei fonemi e dei contesti considerati. Ad esempio, è emerso chiaramente lo status particolare degli elementi fricativi che non esistono nel sistema fonologico italiano - l'interdentale sorda /θ/ e la velare sorda /x/ - ma anche dell'approssimante laterale palatale sonora /ʎ/ che, al contrario, appartiene al suddetto sistema; inoltre l'analisi della variabilità osservata nella ripetizione degli stimoli ha permesso di evidenziare la presenza di meccanismi di auto-correzione e, talvolta, di ipercorrezione.
‘

Un sistema automatico per la localizzazioni delle zone formantiche nella identificazione del parlante.

Gianpaolo Coro[1], Mauro Falcone[2]

[1]*Dip. di Scienze Fisiche - gruppo NLP, Università Federico II, Napoli,*

[2]*Fondazione Ugo Bordoni, Roma*

coro@na.infn.it, falcone@fub.it

Il problema della identificazione e della verifica del parlante assume due aspetti fondamentali a seconda se si stia considerando il caso di applicazioni per accedere a servizi, oppure se si stia considerando il caso di decidere, in tempi e modalità non vincolate, se due campioni di voce appartengano o no allo stesso parlante. Questo ultimo caso è quello tipico del forense ed è quello di nostro interesse. Tecniche statistiche specifiche sono state sviluppate per l'adempimento di tale compito, e queste sono basate sull' analisi dei valori formatici delle vocali. Una serie di esperimenti, in pratica dei casi reali, sono stati eseguiti da decenni in diversi Istituti, ma sempre attraverso la mediazione di un operato esperto che localizzava le vocali da "utilizzare" nel test statistico. Il presente lavoro vuole proporre una metodologia automatica che sostituisca questa fase manuale del lavoro, e a tal fine si propone un metodo basato su procedure simili a quelle operate manualmente e guidate, nella fase di addestramento del sistema, da una base dati realizzata da operatori esperti. L'esperimento è eseguito su materiale telefonico, o su materiale convertito a qualità telefonica, che costituisce un riferimento realistico dello scenario operativo. Tuttavia, poiché non è pensabile che un sistema automatico sia in grado di individuare le stesse vocali che localizza un esperto, anche perché ricordiamo che l'operatore non ha l'obiettivo di trovare tutte le vocali utilizzabili ma solo un insieme "sufficiente", delle opportune metriche di misura delle prestazioni dovranno essere utilizzate.

Il sistema proposto è composto da una serie di moduli che operano sequenzialmente sul segnale vocale. Più in dettaglio, un primo blocco opera una sillabazione del segnale vocale alla ricerca di zone fortemente energetiche che contengano almeno una vocale, un secondo processo individua una o più zone vocalmente prominenti in ciascuna sillaba, mediante un'analisi dell'andamento del pitch e dell'energia. Una fase successiva si occupa del riconoscimento delle zone individuate, sulla base di un filtraggio secondo un banco di filtri legati ad un modello percettivo del suono, e ad un'estrazione successiva di coefficienti cepstrali (**MFCC**). Per ogni segmento vocalico riconosciuto viene operata una stima automatica dei valori formantici, che, insieme alle etichette, vengono utilizzati in una successiva fase di identificazione del parlante. Quest'ultima tipicamente conduce l'elaborazione mettendo a confronto i risultati ottenuti dall'analisi di un segnale contenente una voce non identificata e di un insieme di registrazioni di voci note, e fornisce in uscita una stima della verosimiglianza tra i segnali noti e quello anonimo.

Forma prosodica e funzione pragmatica delle domande polari in napoletano: alcuni dati preliminari.

Claudia Crocco

Università di Salerno

È noto che all'interno del gruppo delle domande polari è possibile distinguere domande con funzioni diverse. Bolinger (1989) ha parlato della differenza tra domande informative e domande di conferma, una distinzione che è stata ripresa in molti studi, che hanno mostrato collegamenti tra forme prosodiche diverse e funzioni diverse della domanda (Grice et al. 1995, Grice e Savino 1995a, b,1997, 2003, Kügler 2003, Crocco e Vermandere 2004).

Nell'annotazione pragmatica del *map task* (Anderson et al. 1992) si distinguono tre tipi di mossa che possono essere realizzati attraverso la domanda polare, *query YN*, *check* e *align*. La prima è una domanda di tipo informativo, mentre le ultime due sono domande di conferma sul compito o sull'interazione dialogica in corso tra i parlanti (Carletta et al. 1996).

Per quel che riguarda la varietà napoletana di italiano, gli studi disponibili, prevalentemente di ambito autosegmentale metrico (D'Imperio 1997, 2001, 2002, Grice et al. 2005), distinguono domande fonologicamente diverse per l'ampiezza del focus e la relativa posizione dell'accento principale. Lo scopo di questo lavoro è mettere in relazione la descrizione corrente della forma fonologica della domanda polare in napoletano con la distinzione sopra citata tra domande con funzione di richiesta di informazione e di conferma.

Il campione esaminato è costituito da un dialogo (10 minuti circa, 131 turni; *route giver* di sesso maschile, *route follower* di sesso femminile) tratto dal corpus AVIP/API e provvisto di annotazione pragmatica *map task* (Ferrari 2003).

Il dialogo è stato integralmente annotato a mano, utilizzando un insieme di etichette modellato sull'annotazione INTSINT (Hirst e Di Cristo, 1998, Hirst 1999, Hirst et al. 2000, Giordano in stampa). Il campione di domande esaminato comprende circa 50 unità tonali in mosse di tipo *query YN*, *check* e *align*.

Dall'analisi del campione e dal confronto tra i dati raccolti per questo studio e alcuni altri utilizzati per un lavoro precedente (Crocco e Vermandere 2004), sono emerse alcune differenze nella configurazione intonativa globale in domande con funzione differenti. In particolare, la posizione e l'ampiezza del focus appaiono influenzati dalla funzione di richiesta di informazione vs. di conferma.

Un sistema automatico per la caratterizzazione degli speaker in flussi multimediali

Leandro D'Anna[1], Gennaro Percannella[1], Carlo Sansone[2], Mario Vento[1]

[1]*Dip. di Ingegneria dell'Informazione e Ingegneria Elettrica, Università di Salerno*
[2]*Dipartimento di Informatica e Sistemistica, Università di Napoli "Federico II"*

Il problema dell'individuazione, nella sola parte audio, della presenza di un solo speaker o di un insieme di più speaker in flussi multimediali di tipo news broadcasting è parte della più ampia classe di problemi legata alla speaker identification. Nel corso degli anni, molte sono state le grandezze analizzate, sia nel tempo che in frequenza, e le tecniche automatiche elaborate per caratterizzare univocamente un singolo speaker. Tutte si possono suddividere in due grandi metodologie: una basata sulla sola analisi della parte audio, l'altra con un approccio integrato audio-video.

La prima metodologia si rivela molto complessa da applicare ai flussi multimediali di tipo news broadcast. Infatti all'interno delle stesse non esiste un omogeneità da un punto di vista acustico in quanto si alterna, ad esempio, audio in ambiente controllato, quale quello da studio, con audio non controllato, quale quello da inviati esterni. Inoltre tale metodologia non è idonea per sistemi che debbano lavorare in tempo quasi reale.

La seconda metodologia, invece, appare essere più promettente in quanto si basa sugli accurati risultati ottenuti dall'analisi video. Quest'ultima, infatti, individuando la porzione del flusso multimediale in cui è presente lo speaker, permette di isolare con un ottima accuratezza le porzioni di parlato da studio consentendo, poi, sia un' analisi audio su porzioni omogenee da un punto di vista acustico e sia un analisi con porzioni di durata molto breve e quindi più idonee ad un sistema in tempo quasi reale.

In questo lavoro, che nasce da un approccio integrato audio-video, viene descritto un esperimento per valutare quanto il solo parametro frequenza fondamentale sia discriminante rispetto alla presenza o meno di due speaker nel flusso audio corrispondente ad un insieme di shot video in cui viene individuata la presenza di uno speaker generico sulla base di sole informazioni video.

Nell'esperimento vengono utilizzati vari algoritmi standard per l'estrazione della f0 e viene utilizzato un semplice criterio a soglie per attenuare sia gli errori di raddoppiamento e sia gli errori di dimezzamento caratteristici di ogni algoritmo per l'estrazione della f0.

La tecnica proposta si basa sull'individuazione di una opportuna distribuzione statistica per la f0 mediante un algoritmo di Maximum Likelihood Estimation (MLE) e sull'utilizzo di un criterio di similarità statistico basato sul test di ipotesi di Kolmorogov-Smirnov.

Il corpus utilizzato nell'esperimento è costituito da telegiornali italiani cui a volte sono presenti due speaker e a volte uno solo.

Una prova fonologica della categorizzazione di un sistema nominale?

Amedeo De Dominicis

Università della Tuscia – Viterbo

1. Scopi

Vorrei illustrare un caso di differenziazione tra radici nominali e radici verbali in cui le prime sono caratterizzate da una specifica proprietà tonale soggiacente la quale è assente nelle seconde.
Il punto interrogativo nel titolo è dovuto al fatto che i fenomeni che verranno presentati sono suscettibili di essere interpretati come espressione di una categorizzazione cognitiva distinta per nomi e verbi, ma anche in alternativa come residui di una evoluzione diacronica; ma non vi è modo di stabilire quale delle due interpretazioni sia vera.
La lingua analizzata è il gizey. Appartiene al gruppo ciadico centrale della famiglia afroasiatica. È parlata nel nord Camerun e sud Ciad da circa 12.000 persone. È una lingua non scritta, tonale e – almeno fino ad ora – non descritta.

2. Dati

Il sistema tonale del gizey comprende due toni fonologici di livello.
Davanti a pausa, le radici nominali e verbali terminanti in vocale vengono automaticamente suffissate da un'occlusiva glottale oppure la vocale finale diventa cricchiata o ancora desonorizzata.
In posizione finale di frase, i nomi vengono suffissati da una marca di definitezza (un 'articolo' espresso da consonante + vocale: Radice + CV). Perciò, in questa posizione la struttura nominale è Radice + {CVʔ, o CV̰, o CṼ̰}. Es.: /gu/ 'albero' → /gu(n)naʔ/, /gu(n)na̰/, /gu(n)nã̰/. (Le forme dell'articolo sono /na/ per il maschile, /da/ o /ta/ per il femminile).
Nelle forme di citazione (nominalizzate), il verbo subisce lo stesso fenomeno: es. /gun/ 'torcere' → /gunaʔ/, /guna̰/, /gunã̰/.
Per quanto riguarda i toni di livello, questi suffissi sono H o L, cioè non sono associati ad un solo tono, ma presentano una grande variazione.
Tuttavia, nei test percettivi condotti con i nativi risulta che essi sono in grado di distinguere chiaramente coppie minime tonali nome vs. verbo, come /gunaʔ/ 'torcere' vs. 'albero'.
La situazione appare piuttosto strana, dal momento che uno stesso suffisso viene realizzato con livelli tonali molto diversi, sebbene a livello percettivo esso continui ad essere nettamente identificato.
Evidentemente, la sua pertinenza fonologica è da ricercare non nel livello, ma nella sua forma o contorno tonale. Questo è quanto ho verificato.
Infatti, il suffisso nominale risulta caratterizzato da un tono di superficie discendente (F), mentre quello verbale da un tono di superficie alto (H).
Attraverso una serie di esperimenti, possiamo inferire che il tono di superficie F dei suffissi nominali è la manifestazione di un tono basso (L) flottante soggiacente, posto alla destra dello *skeleton* tonale della radice nominale e quindi appartenente alla sua struttura fonologica profonda.
A livello lessicale, tale suffisso è dotato di tono H soggiacente. Quest'ultimo affiora in superficie come H nel caso dei verbi. Mentre, nel caso dei nomi, il tono L soggiacente della radice viene collegato al successivo H mediante una linea di riassociazione destrorsa e il risultato è un tono di superficie F.
La mancanza di dati storici su questa lingua e su tutto il gruppo ciadico, rende impossibile stabilire se il fenomeno in esame sia una prova di una diversa categorizzazione cognitiva del sistema nominale rispetto a quello verbale, oppure se il tono L flottante sia residuale, cioè fosse un tempo associato ad una sillaba presente in uno stadio di lingua precedente e collocata a destra della radice nominale; tale sillaba potrebbe essere scomparsa successivamente, lasciando traccia di sé solo attraverso il suo residuo tonale.

Traduzione Automatica del Parlato di Sessioni del Parlamento Europeo

M. Federico, N. Bertoldi, R. Cattoni, M. Cettolo, B. Chen, D. Gupta

ITC-irst, Centro per la Ricerca Scientifica e Tecnologica, Trento

La traduzione automatica è da sempre considerata una delle sfide più affascinanti dell'informatica. Dopo decenni di risultati altalenanti, è stato ottenuto un sensibile miglioramento delle prestazioni grazie all'utilizzo di metodi statistici. Questo successo ha aperto nuove prospettive di applicazione col conseguente aumento degli investimenti nella ricerca. Sebbene la qualità dei testi prodotti dalla macchina sia ancora molto inferiore a quella di un interprete professionista, la traduzione automatica potrebbe divenire un'importante chiave d'accesso all'enorme mole di informazione prodotta sul pianeta. Questa sfida è oggi l'oggetto di due importanti progetti di ricerca, GALE negli Stati Uniti e TC-STAR in Europa. In particolare, TC-STAR (Technology and Corpora for Speech to Speech Translation) è un progetto integrato, di cui ITC-irst è coordinatore, finanziato dalla Commissione Europea all'interno del Sesto Programma Quadro. Il progetto si occupa della traduzione del linguaggio parlato e coinvolge alcuni tra i più importanti laboratori di ricerca nei settori del riconoscimento e sintesi del parlato e della traduzione automatica.

In TC-STAR, l'ITC-irst ha sviluppato un sistema allo stato dell'arte per la traduzione automatica delle sessioni plenarie del Parlamento Europeo. Il progetto si è finora focalizzato sulla traduzione da spagnolo a inglese, ma è già prevista l'estensione ad altre lingue.

Il sistema ITC-irst utilizza corpora paralleli per apprendere traduzioni sia di parole singole che di sequenze (blocchi) di più parole. Il sistema si interfaccia con un sistema di riconoscimento del parlato in grado di produrre ipotesi multiple, sia sotto forma di grafo di parole che di lista di frasi. Il processo di traduzione si basa su un modello log-lineare che combina diversi sottomodelli specializzati. Un algoritmo di ricerca basato sulla programmazione dinamica traduce la frase di ingresso combinando diverse operazioni: ricerca di sottosequenze da tradurre in blocco; generazione di ipotesi di traduzione per ogni blocco; ricerca della posizione migliore per ciascuna traduzione; valutazione della scorrevolezza della frase generata. L'algoritmo, durante la traduzione, esplora moltissime possibilità e sceglie quella che globalmente ha il punteggio migliore. Il sistema ITC-irst è risultato il migliore nell'ambito di una valutazione internazionale svoltasi recentemente negli Stati Uniti, che verteva sulla traduzione di frasi colloquiali.

Durante l'intervento, verrà descritto il sistema sviluppato nell'ambito del progetto TC-STAR. In particolare verranno messe in evidenza le differenze di approccio nei casi di input parlato e scritto.

Uso della funzione di periodicità multicanale per la stima robusta della frequenza fondamentale di parlato a distanza

Federico Flego, Maurizio Omologo

ITC-irst, Centro per la Ricerca Scientifica e Tecnologica, Trento

La stima della frequenza fondamentale (F0) nel campo dell'elaborazione del segnale vocale, viene utilizzata in diverse applicazioni, come per esempio il riconoscimento vocale, l'identificazione del parlatore o l'analisi della prosodia. Stimare accuratamente F0 è un problema noto per il quale esistono diversi algoritmi che offrono ottime prestazioni nel caso di segnali vocali close-talk, in assenza di rumore e di riverbero e quando sia possibile effettuare post-processing.

Nel caso di segnali vocali registrati con microfoni lontani, in presenza di rumore o riverbero, esistono soluzioni che utilizzano schiere di microfoni, imponendo però vincoli sulla posizione del parlatore e sull'orientazione della testa, nonché sull'insieme di microfoni e sulla geometria dello stesso. In questo caso le tecniche utilizzate più comuni sono il il beamforming o il filtraggio spaziale-temporale che permettono di isolare il messaggio vocale e riducono gli effetti indesiderati causati dal rumore fondo, dal riverbero o da altre sorgenti sonore. Uno dei fattori che determina le prestazioni di questi sistemi è la distanza tra i singoli microfoni, che è di solito nell'ordine dei centimetri per evitare o contenere l'effetto di aliasing spaziale. Non sono molte invece le soluzioni proposte che non impongano questo tipo di vincoli. Questo lavoro si inserisce tra queste e affronta il problema della stima di F0 in un contesto di multimicrofonia distribuita senza imporre vincoli spaziali e in presenza di riverbero. Caratteristica, quest'ultima, di qualsiasi scenario reale in ambiente chiuso e che rende molto difficoltosa la stima di F0. Inoltre, grazie alla sua bassa complessità computazionale, è adatto ad essere utilizzato in contesti dove sia richiesta l'elaborazione in tempo reale. Il progetto si inserisce così, nel contesto più ampio dei nuovi ambienti di lavoro, nei quali il cosiddetto “ambiente intelligente”, viene realizzato attraverso un uso estensivo di sensori (telecamere, microfoni, etc.) collegati a computer invisibili, integrati nell'ambiente.

In questo nuovo framework, diventa quindi importante la rimozione di ogni vincolo nella distribuzione dei microfoni nell'ambiente, permettendo quindi la realizzazione dello “ubiquitous computing”. Ciò implica di conseguenza un importante contributo in termini di flessibilità di implementazione delle applicazioni nel campo dell'elaborazione del segnale vocale.

Lo scenario specifico qui considerato e relativo al progetto CHIL, prevede un parlatore immerso in una rete di microfoni distribuita arbitrariamente (distributed microphone Network) che fornisce sequenze multi-canale per la caratterizzazione del soggetto parlante. Le distanze minima e massima tra parlatore e microfoni sono rispettivamente di 1,80 e 2,40 metri, condizionando quindi la fedeltà delle sequenze vocali ottenute per ogni canale. Per la stima di F0, si adotta un modello comune per la sorgente sonora e si assume che l'informazione procedente da ogni singolo microfono ne rappresenti una diversa osservazione. Dopo aver derivato da queste una descrizione congiunta nel dominio delle frequenze, viene introdotta la multichannel periodicity function (MPF) passando al dominio del tempo. Da questa si deriva la stima di F0 per il frame considerato. L'algoritmo proposto valuta inoltre l'affidabilità di ogni singolo canale assegnando a ciascuno di essi un peso per la valutazione finale. L'algoritmo proposto è stato confrontato con altre due tecniche di stima di F0 estese al contesto multicanale. Entrambe si basano sulla funzione di autocorrelazione: la Weighted Autocorrelation (WAUTOC) e YIN, algoritmo appartenente allo stato dell'arte nel settore. L'algoritmo proposto, testato su una base di dati reale, si dimostra il più efficiente in termine di correttezza nella stima di F0 anche nel caso in cui alcuni specifici canali siano stati condizionati artificialmente da forte rumore, dimostrando la sua capacità nello scegliere i canali più affidabili per effettuare la stima. Dal confronto dei risultati, risulta inoltre che gli algoritmi che effettuano la stima di F0 nel dominio del tempo sono i più penalizzati in un ambiente riverberante. L'algoritmo proposto è da considerarsi di tipo ibrido, basato cioè sull'analisi effettuata sia nel dominio delle frequenze che nel del tempo, risultando di conseguenza più robusto nelle condizioni considerate.

Variazione prosodica del connettore pragmatico "magari" : uno studio acustico e percettivo.

Gaillard-Corvaglia Antonella

Laboratorio di Fonetica e Fonologia (UMR 7018) CNRS / Sorbonne Nouvelle, Parigi 3

L'intonazione è una sottocomponente della prosodia, principalmente legata alle variazioni della Frequenza Fondamentale (F0), della Durata, dell'Intensità, del timbro e della qualità della voce, percepibili globalmente dall'orecchio umano. Essa è l'insieme dei fenomeni di accentazione, d'intonazione e di ritmo che permettono di trasmettere l'informazione legata al senso (Jacqueline Vaissière, 2002). Dalle ricerche di fonosemantica risulta che la prosodia agisce direttamente sulla costituzione del senso. Cio' significa allora che in una parola che possiede più significati, il senso sarebbe legato alla variazione dei parametri acustici ?

Poichè l'intonazione fornisce al discorso il suo valore autentico (Lepschy, Appunti sull'intonazione italiana, 1978), ci siamo interrogati sulle possibili relazioni tra le variazioni della prosodia e del significato in una parola *"multisenso"* come il connettore pragmatico "magari".

La prosodia di questa parola è legata ai suoi molteplici significati ? il fattore « senso » ha un effetto significativo sulla variazione dei suoi paramentri acustici : F0, Durata, Intensità ? la sua struttura prosodica è regolata dalla semantica o dalla posizione nella frase?

Lo scopo della nostra ricerca è quello di verificare in che misura una parola avente dei significati diversi secondo il contesto, abbia nello stesso tempo una struttura prosodica che varia con il senso. Inoltre, le differenze prosodiche tra i diversi sensi possono dipendere dalle varie posizioni che la parola occupa all'interno della frase. All'inizo di frase, per esempio, ci si aspetterebbe una F0 e un'Intensità più elevate mentre alla fine una durata più lunga (J. Vaissière, Prosody : Models and Measurements, 1983).

Per rispondere a queste domande, abbiamo utilizzato un corpus formato da cinque frasi tratte da Il nuovo Zingarelli (2005), contenenti le diverse funzioni sintattiche del connettore (interiezione, avverbio interposto, ironico, congiunzione, congiunzione enfatica) .

Le frasi sono state lette da quattro locutrici pugliesi. La metodologia d'analisi applicata è stata la seguente:

1) Per l'analisi acustica, abbiamo: a) segmentato la parola "magari" contenuta in ogni frase, b) misurato la Durata, la F0 e l'Intensità di ogni sillaba. Il test ANOVA effettuato, creando delle molteplici combinazioni, ci ha permesso di comparare i valori relativi al fattore S (senso) a due a due.

2) Per l'analisi percettiva, 2 test fatti per verificare: a) la percezione del connettore fuori contesto e b) l'influenza di ognuno dei tre parametri sulla percezione della parola : 1°) 11 persone hanno ascoltato le frasi complete, poi i diversi "magari" fuori contesto e hanno cercato di risituare il connettore nel contesto giusto. 2°) hanno ascoltato i "magari" modificati per F0, Durata, Intensità e cercato di abbinarli al proprio contesto.

I risultati preliminari mostrano che: 1) la F0 e l'Intensità tendono a variare secondo il senso contrariamente alla Durata, relativamente stabile essendo il correlato acustico più importante dell'accento lessicale; ma dal testo di Fisher risulta che la Durata è relativamente più lunga per l'interiezione e più breve per la congiunzione enfatica; F0 della sillaba tonica è sempre più elevata ma tra tutti, la F0 più alta appartiene alla congiunzione (all'inizio di frase); 2) per la posizione, le tendenze sono: F0 più alta in posizione iniziale e Durata più lunga in posizione finale (ironia); 3) la struttura prosodica cambia secondo il senso.

Per risultati più completi, si potrebbe completare l'analisi con ulteriori accezioni del connettore e con più locutori di diversa provenienza geografica ma i risultati preliminari ci permettono di sostenere che le variazioni prosodiche si producono parallelamente alle variazioni semantiche di "magari" e che il "senso" e la "posizione" hanno un effetto molto significativo sulla sua struttura prosodica.

LA PROSODIA DIRETTIVA IN L2. STUDIO PILOTA

Dalia Gamal

Università di Ain Shams, Il Cairo
daliagamal60@hotmail.com

Il presente studio analizza produzioni semispontanee di apprendenti egiziani di italiano. La scarsità degli studi prosodici nelle lingue seconde hanno dato spazio a ipotesi non verificate di un "inevitabile" *transfer* o trasferimento accentuale e melodico dalla lingua prima. Questo lavoro, dunque, si propone come un primo approccio alla prosodia italiana in parlanti arabofoni e mira alla rilevazione del vero ruolo della lingua prima nella produzione di lingue straniere.

Il *corpus* consta di 58 turni dialogici in arabo egiziano e 116 in italiano L2. Per le analisi prosodiche è stato adoperato il metodo INTSINT e il ToBI-*like* per analizzare l'andamento melodico globale delle TU direttive e la distribuzione degli accenti principali.
Dal confronto tra le produzioni in L1 e L2 emergono delle differenze nell'accentazione forte e nel profilo globale e terminale, per cui ci risulta opportuno evitare le generalizzazioni sull'interferenza fonologica.
A livello sociolinguistico si osserva l'influsso della modalità di apprendimento (spontanea *vs* guidata) nella differenziazione tra alcuni comportamenti prosodici (e anche sintattici) degli informatori.
Infine, si possono trarre conclusioni sulla prosodia direttiva per nulla semplice nelle due lingue.

Analisi prosodica comparata: il map task applicato all'italiano, sloveno e polacco

Antonella Giannini-Massimo Pettorino

Università degli Studi di Napoli L'Orientale

In una precedente ricerca abbiamo confrontato, sul piano del ritmo e della prosodia, lo sloveno con l'italiano utilizzando un parlato dialogico spontaneo, ottenuto mediante il metodo del *map task*. Il materiale analizzato consisteva, per lo sloveno, in tre dialoghi prodotti da tre studenti universitari di sesso maschile e tre di sesso femminile provenienti dalla Stiria-inferiore e, per l'italiano, in tre dialoghi prodotti da parlanti napoletani, baresi e pisani. Dal confronto italiano/sloveno è emerso che lo sloveno rispetto all'italiano, presenta una prevalenza di vocalizzazioni *vs.* prolungamenti vocalici, una maggiore durata sillabica, un più alto indice di produttività, una stessa velocità di eloquio,una minore velocità di articolazione. I dati raccolti in quella ricerca sperimentale, hanno fornito interessanti indicazioni a livello ritmico-prosodico, indicazioni che hanno reso possibile ipotizzare un modello di struttura prosodica del parlato dialogico in italiano e in sloveno. In particolar modo sono emerse due strategie diverse che riflettono una diversa programmazione dell'eloquio: l'italiano preferisce una programmazione a breve termine, mentre lo sloveno una programmazione a lungo termine. A tal proposito fu avanzata l'ipotesi che la diversa programmazione fosse legata ad una minore o maggiore complessità sintattica dell'enunciato, vale a dire che mentre le frasi brevi sono gestite da una programmazione a lungo termine, frasi lunghe e complesse necessitano di una programmazione a breve termine.

In questo lavoro intendiamo proseguire questo tipo di indagine e confrontare, adottando la stessa metodologia, i dati precedentemente ottenuti con quelli ricavati da parlanti polacchi. Poiché il sistema di lingua del polacco prevede, come per lo sloveno, numerose consonanti, il confronto, sul piano ritmico, potrà essere particolarmente interessante.

‘Scaling’ e allineamento dei bersagli tonali nell’identificazione di due accenti discendenti

Barbara Gili Fivela

Università degli studi di Lecce

Il contributo è relativo al ruolo delle modificazioni di allineamento e ‘scaling’ dei bersagli tonali nella percezione di accenti discendenti. Precedenti lavori hanno mostrato come lo ‘scaling’ dei bersagli tonali rappresenti un’informazione rilevante per la categorizzazione di accenti considerati simili dal punto di vista fonologico. In base ad un’analisi autosegmentale-metrica [Bruce 1977; Pierrehumbert, 1980], gli accenti erano accomunati dalla presenza di un tono alto allineato ed associato alla sillaba. I due esperimenti descritti in questo lavoro sono volti a verificare la rilevanza delle informazioni di ‘scaling’ anche nell’identificazione di accenti meno simili da punto di vista fonologico, nella fattispecie accenti che presentino toni diversi associati ed allineati alla sillaba - rispettivamente un tono alto ed un tono basso. Considerando un approccio autosegmentali-metrico, gli accenti sono analizzati come bitonali ed identificati come H*+L e H+L*.

I materiali strumentali all’indagine sono rappresentati da registrazioni di parlato della varietà pisana di italiano. Nelle produzioni di parlanti di questa varietà, i due accenti sono associati a funzioni diverse: il primo esprime ‘focus’ contrastivo, mentre il secondo rappresenta l’accento nucleare finale di frase assertive. Dal punto di vista fonetico, il primo accento è caratterizzato dalla presenza, all’interno della stessa sillaba, di una fase ascendente ed una discendente, e quindi da tre bersagli che, nell’ordine, sono basso-alto-basso; il secondo accento è caratterizzato da una sola fase discendente all’interno della sillaba, quindi da una sequenza di due bersagli alto-basso.

I due esperimenti percettivi sono stati eseguiti proponendo a 9 soggetti pisani degli stimoli nei quali erano state manipolate alcune caratteristiche acustiche, e chiedendo loro di stabilire se gli stimoli potevano essere utilizzati per correggere in modo perentorio un enunciato precedente oppure non erano adatti a questo scopo. Tutte le manipolazioni sono state effettuate per mezzo del programma PRAAT (sviluppato da P.Boersma and D.Weenink - “Institute of Phonetic Sciences of the University of Amsterdam”) e con risintesi PSOLA.

Nel primo esperimento, gli stimoli sono stati ottenuti a partire dall’enunciato ‘mangia il melone’, prodotto con interpretazione contrastiva, mentre nel secondo esperimento essi sono stati generati a partire da una produzione neutra della stessa frase. Una parlante pisana ha realizzato tre enunciati per ogni interpretazione. Nonostante il fatto che la manipolazione abbia, ovviamente, riguardato solo una ripetizione, i valori di altezza tonale e allineamento dei bersagli sono stati ricavati facendo la media dei valori riscontrati nelle tre produzioni della parlante.

Le manipolazioni effettuate nel primo esperimento riguardano:

- l’altezza del bersaglio basso all’inizio della fase ascendente - il livello di frequenza fondamentale è stato gradualmente innalzato;
- l’altezza del picco - il livello di frequenza fondamentale è stato gradualmente abbassato.

Le manipolazioni effettuate nel secondo esperimento sono relative a:

- l’altezza del picco - il livello di frequenza fondamentale è stato gradualmente innalzato;
- l’allineamento del picco – la posizione del picco è stata gradualmente ritardata.

I risultati del primo esperimento mostrano che non si ottiene alcuna percezione categoriale al variare dell’altezza tonale del target basso o dell’abbassamento del bersaglio alto. Il secondo esperimento evidenzia invece che innalzando il valore del bersaglio alto e ritardando il suo allineamento all’interno della sillaba si osserva un’influenza sulla percentuale di identificazione degli accenti, anche se non è possibile osservare un vero e proprio salto categoriale. I risultati sono discussi alla luce dei lavori presenti in letteratura.

Note sulla fonetica del ritmo dell'italiano

Rosa Giordano

Università di Napoli Federico II – Università di Salerno

Sulla realizzazione fonetica del ritmo della lingua italiana sono disponibili studi metodologicamente diversi che hanno dato risultati parzialmente concordanti (i riferimenti bibliografici saranno riportati nella presentazione). Accurate indagini sulla produzione e sulla percezione condotte su parlato di laboratorio hanno individuato nella durata il parametro fisico-acustico pertinente per le opposizioni di natura ritmica; i risultati emersi in vari studi sul parlato connesso e spontaneo sembrano però rendere meno definita la situazione, rivelando la forte variabilità delle durate medie e la consistente sovrapposizione dei margini di deviazione standard di sillabe toniche e atone. Un simile fattore potrebbe indurre a ritenere meno determinanti e pertinenti le variazioni relative a tale parametro. Un secondo aspetto evidenziato in alcuni studi sulla prosodia è la modulazione dell'aumento di durata delle sillabe strutturalmente toniche in dipendenza dalla concomitante occorrenza di prominenza melodica.

Questo lavoro presenta un'analisi ritmica strumentale effettuata su un campione di parlato connesso, selezionato fra i brani di apertura di alcune importanti testate giornalistiche televisive nazionali; nel complesso. L'esame delle variazioni di durata è stato effettuato seguendo un metodo differente rispetto a quello dei precedenti studi sul parlato connesso: partendo dai rilevamenti delle durate di testa, nucleo e coda sillabici, si valutano, infatti, gli scarti tra il valore di ciascuna sillaba e quello della precedente, permettendo in tal modo la registrazione dettagliata dell'effettiva dinamica temporale; si valuta inoltre la variabilità legata ai fattori intonativi ai quali si è fatto cenno.

I risultati confermano l'associazione preminente degli incrementi di durata alle sillabe strutturalmente toniche e, inoltre, la progressione negli aumenti temporali di tali sillabe in relazione al loro diverso *status* sul piano intonativo, pur permettendo di constatare, in generale, la fortissima variabilità delle durate nelle classi ritmiche prese in considerazione, così come già rilevato in studi precedenti.

Un fenomeno interessante è rappresentato, infine, dai casi di mancanza di prominenza di sillabe strutturalmente toniche, che potrebbero essere ritenuti indice di una tendenza alla destrutturazione fonetica anche sul piano prosodico.

Lo sviluppo fonetico in relazione agli stadi di produzione della parola: studio pilota di una bambina italiana

*Sara Giulivi, °Claudio Zmarich, **Mario Vayra, °Edda Farnetani

**Dipartimento di Linguistica, Università di Firenze*
°Istituto di Scienze e Tecnologie della Cognizione del C.N.R., Sede di Padova
***Dipartimento di Studi linguistici e Orientali, Università di Bologna*

L'inizio dell'influenza linguo-specifica e le modalità con cui essa si manifesta rappresentano un aspetto aperto alla verifica sperimentale, specificamente di tipo inter-linguistico. Perché ci sia influenza linguo-specifica bisogna dimostrare che: 1) le differenze fonetiche tra i gruppi nazionali sono maggiori delle differenze all'interno dei gruppi; 2) queste differenze riflettono i *patterns* caratteristici di ciascuna lingua (de Boysson-Bardies et al., 1992).

Al fine di distinguere tra proprietà universali e proprietà linguo-specifiche, una strategia consolidata è quella di quantificare le relazioni tra le strutture fonetiche e fonotattiche presenti nel *babbling* dei bambini cresciuti in ambienti linguistici diversi e quelle presenti nelle rispettive lingue materne. Tale strategia di ricerca ha portato Vihman & de Boysson-Bardies (1994), ad esempio, a individuare un'influenza positiva della lingua nativa già a 9-10 mesi, allorché i foni nativi aumentano, e un'influenza negativa a partire dai 12 mesi circa, allorché i foni non nativi diminuiscono.

Per l'italiano non esistono molti dati, e anche quelli esistenti partono purtroppo solo dai 10 mesi: troppo tardi probabilmente per individuare l'inizio dell'influenza della lingua nativa.

Questo studio è parte di un progetto di ricerca più ampio, su bambini dai sei mesi (inizio del *babbling*) ai ventisette. Si presentano qui i risultati delle prime analisi che sono parte della tesi di dottorato del primo autore. Il soggetto indagato è una bambina audioregistrata e trascritta foneticamente a 6, 8, 10, 12, 14 e 16 mesi di età.

I risultati relativi al calcolo delle frequenze dei diversi tipi sillabici e segmentali in rapporto alla loro posizione entro la sillaba e la parola non saranno qui espressi, come in altre occasioni, in funzione dell'età, bensì in fuzione dell'evoluzione lessicale del bambino, e organizzati in tre diversi stadi riconosciuti in letteratura: lo stadio delle 0 parole (da 0 a 3), lo stadio delle 4 parole (da 4 a 14), lo stadio delle 15 parole (da 15 a 24). I risultati potranno così essere direttamente confrontati con i dati interlinguistici presentati in de Boysson-Bardies et al. (1992). La marcia di avvicinamento della bambina verso il sistema fonologico dell'italiano sarà valutata tramite un confronto della sua produzione vocale con i dati relativi alle frequenze di occorrenza delle strutture foniche del lessico italiano, tratti da una lista di parole desunte dal *Primo vocabolario del bambino* (Caselli e Casadio 1995), e da alcuni studi di frequenza disponibili per la lingua italiana.

Boysson-Bardies de, B., Vihman, M. M., Roug-Hellichius, L., Durand, C., Landberg, I., Arao, F. (1992), Material evidence of infant selection from the target language, in *Phonological Development. Models, Research, Implications* (C. A. Ferguson, L. Menn & C. Stoel-Gammon, editor*s*), Timonium: York Press, 369-391.

Caselli, M.C. & Casadio, P. (1995), *Il Primo Vocabolario del bambino*, Milano: Franco Angeli.

Vihman, M., Boysson-Bardies de, B. (1994), The nature and origins of ambient language influence on infant vocal production and early words, *Phonetica*, 51, 159-69.

Riconoscimento di emozioni nel parlato per mezzo di parametri prosodici

Roberto Gretter, Dino Seppi

ITC-irst, Pantè di Povo, Trento

{gretter, seppi}@itc.it

Il presente lavoro si propone di descrivere la realizzazione di un sistema automatico per il riconoscimento di emozioni nel parlato. La peculiarità del progetto consiste nell'aver utilizzato database di parlato spontaneo, in particolare di registrazioni di interazioni vocali uomo-macchina. L'intero progetto può essere suddiviso in tre differenti fasi: la raccolta e l'etichettatura dei dati, l'estrazione di parametri prosodici dalle registrazioni audio e la classificazione emozionale di ciascuna frase.

Sono stati considerati due database. Il primo consiste in registrazioni di utenti italiani presso *call-center* automatici. Questi dati sono stati selezionati ed etichettati da annotatori professionisti in base alle emozioni espresse da ciascun parlatore. Il secondo database consiste invece in registrazioni di utenti durante la messa a punto di un sistema di dialogo automatico. Questa seconda raccolta, in lingua tedesca, contiene un numero più significativo di frasi non emotivamente neutre ed é stata utilizzata per un raffronto con i dati italiani. La scelta di fare uso di parlato spontaneo comporta inevitabilmente alcuni problemi: la presenza preponderante di frasi emotivamente neutre che rende i database molto sbilanciati, la disomogenea distribuzione delle emozioni rilevate e i bassi livelli di consenso registrati tra gli annotatori.

Ulteriori problematiche derivano dall'estrazione automatica di parametri significativi: a tutt'oggi, infatti, le tecniche utilizzate in letteratura non consentono di fare affidamento su un insieme limitato di *features* robusti. L'approccio adottato prevede quindi il calcolo di un gran numero di parametri, anche molto correlati tra loro, per il successivo addestramento di un classificatore automatico. Tali parametri, comunemente utilizzati in letteratura, derivano da funzioni del segnale audio, come energia, frequenza fondamentale, durata ed eventuale presenza di pause. Per una codifica più robusta delle variazioni di questi valori abbiamo preferito procedere alla loro estrazione a livello di parola. Questo ha comportato l'uso precedente di un segmentatore automatico.

Date le potenziali ambiguità delle annotazioni e le lacune nelle informazioni codificate dai parametri acustici il compito svolto dalla classificazione diviene difficile se non critico. Per questo motivo abbiamo testato due diversi tipi di classificatori: le reti neurali e gli alberi binari di classificazione (CART); per entrambi forniamo i risultati nelle configurazioni che si sono rivelate più robuste. Nonostante entrambi i metodi proposti pongano degli svantaggi se applicati a dati sparsi e molto sbilanciati, siamo riusciti a ottenere, per entrambi i database utilizzati, risultati equiparabili e prestazioni più che soddisfacenti.

In conclusione restano ancora irrisolti numerosi problemi di carattere tecnico e concettuale, tra cui la significatività e il rumore codificato dei parametri prosodico-acustici utilizzati e l'affidabilità dell'etichettatura manuale dei dati. Questo fatto indica che sforzi futuri per il miglioramento dei risultati andranno direzionati soprattutto verso un'attenta e più approfondita esamina dei parametri, che andranno modificati, selezionati e affiancati da nuovi di altra natura anche non prettamente prosodica.

Le occlusive bilabiali in salentino (Puglia): uno studio acustico e percettivo

Kamiyama Takeki, Gaillard-Corvaglia Antonella

Laboratorio di Fonetica e Fonologia (UMR 7018) CNRS / Sorbonne Nouvelle, Paris 3

I dialetti salentini sono caratterizzati dalla presenza dell'opposizione fonologica tra le consonanti geminate e quelle semplici non solo all'interno della parola ma anche in posizione iniziale (Mancarella 1974, Romano 2003). Questa manca pero' nel caso delle occlusive sonore. È interessante notare infatti come la realizzazione fonetica della consonante bilabiale sonora, sia sempre rinforzata (Mancarella 1974).

Con la nostra ricerca preliminare (Kamiyama e Gaillard-Corvaglia, 2005, "in corso di pubblicazione"), effettuata su un'informatrice originaria di Taurisano (LE), abbiamo osservato che: 1) la durata dell'occlusione della /b/ detta "forte" in dialetto è effetivamente più lunga della /b/ semplice in italiano; 2) durante un test di percezione al quale hanno partecipato 15 italiani originari delle tre zone d'Italia (Nord, Centro, Sud), le sillabe /VbV/ del dialetto, tratte dal corpus registrato, sono state percepite più spesso come una /b/ doppia dell'italiano che come una /b/ semplice.

Se la realizzazione fonetica della /b/ salentina detta "forte" è geminata, potremmo porci le domande seguenti: si trovano in dialetto salentino dei casi di realizzazione foneticamente semplice dell'occlusiva bilabiale sonora? che cosa succede allora alla /p/, occlusiva fonologicamente sorda?

Allo scopo di rispondere a queste domande, ci proponiamo di analizzare acusticamente e percettivamente un corpus più esteso che comprende entrambe le occlusive bilabiali in diverse posizioni (pre-tonica, post-tonica, inizio, interno di parola). Il corpus, composto da 53 parole in italiano (di cui 32 contengono una /b/ e 21 una /p/), è stato pronunciato dall'intervistatrice e tradotto spontaneamente da 4 informatori originari del paese appena citato; ogni parola si trovava all'interno di una frase cornice come: /tiku ... ma no tiku .../ ("dico ... ma non dico ...").

La metodologia di analisi adottata è la seguente: in primo luogo, si è misurata la durata dell'occlusione delle bilabiali (cercando di stabilire nel migliore dei modi la parte consonantica quando l'occlusione non è completa) e poi quella della vocale che precede la bilabiale. Si è potuto osservare quindi grazie all'oscillogramma, allo spectrogramma e all'andamento della Fondamentale, se la fonazione continuasse durante l'occlusione e se questa fosse completa oppure no.

In secondo luogo, al fine di studiare la percezione dei due fonemi, si è effettuato un test di percezione su 15 italiani originari delle tre zone d'Italia (Nord, Centro, Sud) e 15 locutori nativi del francese utilizzando delle sequenze VbV e VpV in dialetto, tratte sempre dal corpus registrato. I soggetti hanno avuto il compito di dire se si trattasse di una /b/ semplice o doppia, oppure di una /p/ (nel caso dei francesi, si è chiesto solo se si trattasse di una /b/ o di una /p/, poichè non esiste in francese l'opposizione semplice e geminata).

I risultati preliminari delle analisi acustiche mostrano che: 1) la durata della /b/ è più lunga di quella della /p/; 2) l'occlusione della /b/ è completa nella maggior parte dei casi, ed è completamente sonora; 3) l'occlusione della /p/ è spesso incompleta, ed è frequentemente sonorizzata (a metà o del tutto).

Queste tendenze potrebbero essere verificate ulteriormente, misurando, per esempio, la fonazione grazie ad un elettro-glottogramma e verificando lo stato dell'occlusione attraverso la misura della pressione intra-orale.

Asimmetria nella percezione vocalica di L1: uno studio di sintesi articolatoria

Charalampos Karypidis[1], Angelica Costagliola[1-2], Gilles Guglielmi[3]

[1]*LPP - UMR 7018, CNRS / Université de la Sorbonne Nouvelle- Paris 3*
[2]*Università degli Studi Lecce - Dip.Filologia, Linguistica e Letteratura*
[3]*ARP – UFRL Université Paris 7 Denis Diderot*
ch_karypidis@yahoo.com, angelicacostagliola@yahoo.it, gillesgug@yahoo.fr

Il nostro contributo prenderà in esame il fenomeno della asimmetria percettiva, così riassunto da Polka e Bohn (2003): "Asymmetries in vowel perception occur such that discrimination of a vowel change presented in one direction is easier compared to the same change presented in the reverse direction... the more peripheral vowel within a contrast serves as a reference or perceptual anchor.".
In studi recenti, per la sintesi dei continua vocalici si è utilizzato il frazionamento della distanza Euclidea F1/F2 tra due prototipi in punti equidistanti (quei prototipi che meglio possono rappresentare due differenti categorie vocaliche). Tuttavia, i suoni ottenuti non erano abbastanza naturali, visto che ad alcuni di loro erano state assegnate delle combinazioni di valori formantici che non potevano essere prodotte da un canale orale umano. Nondimeno, l'assegnazione di valori fissi per F3-F4-F5 (rispettivamente 3010, 3300, e 3850 Hz,) ha generato falsi picchi spettrali (vicino a quelli di /i/), inducendo in questo modo gli informatori ad identificare più /i/ di quante ne avrebbero dovuto realmente identificare. Un recente studio sui prototipi vocalici (Karypidis et al., *in preparazione*) suggerisce che /i/ ha una zona di percezione molto ristretta, nonostante le sue caratteristiche di stabilità e perifericità acustiche, e nonostante l'assenza di /e/ medio-alta nel sistema.
Prendendo in considerazione queste incongruenze metodologiche, abbiamo scelto di elaborare i nostri stimoli usando il modello articolatorio di Maeda (1988), conosciuto anche col nome di VTCALCs. Abbiamo usato i parametri articolatori forniti da VTCALCs su 10 stimoli: a partire dal prototipo sintetico /i/ (stimolo numero 1, il più periferico) abbiamo progressivamente spostato i parametri (altezza della mascella e posizione della lingua) verso il prototipo sintetico /e/ (stimolo numero 10, il meno periferico). Successivamente abbiamo sottoposto il continuum delle 10 vocali a 34 informatori nativi francofoni (età media 30 anni), somministrando:
a) un test di identificazione all'interno del quale gli informatori dovevano identificare come /i/ oppure /e/, 7 ripetizioni di ciascuno stimolo presentato in ordine casuale (10x7=70 stimoli);
b) un test di discriminazione all'interno del quale agli informatori venivano presentate tutte le combinazioni possibili di vocali (34) che differivano di uno o due stimoli lungo il continuum e in entrambi gli ordini (i.e. stimoli 1-2, 1-3, 2-1, 3-1,etc.). Successivamente veniva chiesto loro se le due vocali che sentivano erano uguali oppure differenti. L'intervallo di silenzio tra gli stimoli è stato fissato a 250 ms ed ogni coppia è stata presentata cinque volte (34x5=170 coppie).
Ogni test è stato preceduto da un breve training.
I risultati del test di identificazione rivelano una chiara transizione quantica (categoriale) da un tipo fonologico all'altro (la frontiera abbraccia due stimoli, i numeri 4 e 5) insieme a dei tassi di identificazione estremamente alti (oltre 94%) per le vocali non di frontiera.
I risultati del test di discriminazione dimostrano che: a) l'asimmetria (cioè l'effetto dell'ordine) ricorre solo quando la distanza nel continuum è nell'ordine di uno stimolo ($p=0.0116$ con Wilcoxon Signed Ranks) e cioè soltanto quando la differenza acustica è minima, e b) la discriminazione è significativamente più agevole in entrambi gli ordini, cioè con lo stimolo vocalico più periferico dei due in prima o in seconda posizione ($p=0.0006$ e $p=0.0024$ rispettivamente con Mann-Whitney) e quando la distanza nel continuum è nell'ordine di due stimoli. Infine, la percentuale di discriminazione tende ad essere inferiore quando uno dei prototipi è presentato come elemento della coppia, in accordo con Kuhl (1991): "the best instances of a phonetic category play a special role in perception.".

Karypidis, Ch., Costagliola, A. and Colazo-Simon, A., (*in preparazione*). "Vowel prototype assimilation: a cross-linguistic perceptual study of five-vowel systems.".
Kuhl, P. K., (1991), "Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not." Perception & Psychophysics 50(2): 93-107.
Maeda, Sh., (1988), "Improved articulatory models.", J. Acoust. Soc. Am. 84(1):146.
Polka, L. and Bohn, O.-S., (2003). "Asymmetries in vowel perception". Speech communication 41: 221-231.

La durata consonantica nel dialetto di Lizzano in Belvedere (BO)

Michele Loporcaro, Rachele Delucchi, Nadia Nocchi, Tania Paciaroni, Stephan Schmid

Università di Zurigo

La degeminazione delle consonanti geminate viene considerata una delle isoglosse definitorie della Romània occidentale. Tuttavia, vi sono in quest'area alcune varietà dialettali in cui il processo di degeminazione non si è del tutto compiuto. Un esempio è costituito dal dialetto lombardo alpino di Soglio, in Val Bregaglia (Canton Grigioni, Svizzera); questo comune rappresenta una classica 'area laterale', data la sua collocazione estrema al margine settentrionale dell'Italoromània. Un'indagine sul campo a Soglio e la successiva analisi acustica di parlato appositamente elicitato hanno permesso di verificare per questa parlata tanto la persistenza di geminate etimologiche (ereditate cioè dal latino) quanto la creazione di nuove geminate anetimologiche in determinate condizioni prosodiche e fonotattiche (Loporcaro et al. 2005).

Esiste però un'altra area marginale dell'Italoromània settentrionale per la quale la bibliografia dialettologica documenta la conservazione almeno parziale della lunghezza consonantica: si tratta della zona immediatamente a Sud della linea La Spezia-Rimini, che abbraccia da Ovest i dialetti del crinale appenninico tosco-emiliano sino a raggiungere l'Adriatico intorno a Senigallia. Entro questa zona abbiamo scelto per il nostro sondaggio il dialetto parlato a Lizzano in Belvedere, sull'Appennino bolognese, già ben descritto da Malagoli (1930). Il lizzanese presenta un quadro molto simile a quello del dialetto di Soglio, per cui alle geminate latine nei parossitoni (ad esempio *vacca*) si aggiungono nuove geminate nei proparossitoni (ad esempio *laggrima*). A proposito della durata di queste consonanti il Malagoli annotava che "la consonante lunga lizzanese suona come una consonante e mezza toscana" (p. 131).

Al fine di verificare strumentalmente tale osservazione, abbiamo condotto un'indagine sul campo nel settembre 2004. A quattro parlanti è stato chiesto di tradurre in dialetto 105 frasi, nelle quali erano inserite in posizione non finale delle parole contenenti vari tipi di consonanti geminate e scempie.

In questo contributo presentiamo i primi risultati dell'analisi acustica, per la quale si tiene conto dei seguenti parametri: durata dell'intero segmento consonantico, durata della vocale precedente, rapporto fra durata vocalica e consonantica. In base ai dati sinora analizzati sembra che …

Riferimenti bibliografici

Loporcaro, M., T. Paciaroni, S. Schmid (2005). Consonanti geminate in un dialetto lombardo alpino. In: P. Cosi (a cura di). *Misura dei parametri – aspetti tecnologici e implicazioni nei modelli linguistici* (Atti del 1° Convegno AISV, Padova, 2-4 dicembre 2004).

Malagoli, G. (1930). Fonologia del dialetto di Lizzano in Belvedere (Appennino bolognese). *L'Italia dialettale* 6: 125-196.

“ANALYSIS BY SYNTHESIS: UNO STUDIO DELLE COMPONENTI DELLA COMUNICAZIONE MULTIMODALE TRAMITE LA SINTESI AUDIOVISIVA DI UNA FACCIA PARLANTE”

E. Magno Caldognetto, Piero Cosi, Federica Cavicchio

ISTC-CNR, Sezione di Padova, VIA aNGHINONI N.10, 35121 PADOVA

La sintesi di una Faccia Parlante può costituire una sofisticata e complessa metodologia di indagine della coproduzione di informazione extra-, para- e linguistica nella modalità acustico-uditiva e ottico-visiva e di valutazione del loro ruolo comunicativo.

Saranno presentati esempi di sintesi bimodale “incrementale”: partendo da una Faccia che riproduce i movimenti labiali articolatori si aggiungeranno i movimenti mandibolari, le modificazioni paralinguistiche emotive, i movimenti di testa, occhi, sopracciglia, fronte correlati alla struttura dell’informazione nell’enunciato (la cosiddetta *visual prosody*) e quelli che veicolano l’informazione sulle emozioni, ponendoli in relazione con la sintesi acustica, in funzione anche di futuri test di valutazione.

Saranno discussi innanzi tutto i diversi problemi fonetici implicati da una sintesi che, in funzione di applicazioni didattiche dell’italiano come L1 o L2 o a scopo riabilitativo, oltre ad essere gradevole, deve essere anche naturale, robusta e linguo-specifica.

Per ottenere queste caratteristiche, per quanto riguarda l’articolazione labiale, dovrebbero essere implementate le conoscenze sulla coproduzione di movimenti labiali e mandibolari per le consonanti bilabiali e labiodentali e di lingua, mandibola e labbra per le vocali dell’italiano e sulla coarticolazione CV.

Il confronto tra i dati sperimentali attualmente disponibili per l’italiano e i prodotti di esperimenti di risintesi evidenzierà il divario tra le nostre attuali conoscenze analitiche e la quantità dei dati indispensabili per la stesura di regole di sintesi pienamente soddisfacenti.

Per quanto riguarda l’utilizzazione della Faccia Parlante nell’e-learning, in particolare in chat e forum, e in tutte le applicazioni di interfacce uomo-macchina che vogliano riproporre situazioni di interazioni faccia-a-faccia, la sintesi bimodale di informazioni linguistiche segmentali e sopra segmentali deve essere integrata da dati sugli effetti della coproduzione di configurazioni labiali emotive e di bersagli articolatori linguistici e da regole sulla cooccorrenza di questi nuovi atteggiamenti labiali con specifici atteggiamenti facciali.

Interface Interpretation and the Interpretation of the PF Interface

Lunella Mereu – Mara Frascarelli

Università Roma Tre

This paper is concerned with the complex interaction between syntax, intonation and discourse grammar and it aims at discussing some of the aspects which emerge at the interface between these levels of analysis. Recent developments in linguistics have made it clear that a deeper understanding of language phenomena requires the investigation of the complex relationship between the syntax, pragmatics and phonology of sentences and utterances. Moreover, the growing attention on discourse grammar has assigned a crucial role to performance, so that language analysis is more and more based on 'real' data, drawn from spontaneous speech and spoken language. More specifically, a strong hypothesis is leading recent investigation, which can be summarized as follows:

a) there is a *systematic connection* between syntax and intonation in the identification of *discourse categories*;
b) this connection *is not direct*, but mediated by prosodic phrasing, that transfers syntactic material to interface interpretation.

A number of interesting frameworks have been proposed in the literature to analyze intonation. In particular, in our investigations upon the prosodic marking of discourse functions in Italian, we approached Martin's (1981) Winpitch software, ToBI framework (Beckman and Pierrehumbert 1986) and the INTSINT system (Hirst and Di Cristo, 1998). Since working on intonation means to deal with physical data that can be measured with appropriate machinery, we expected this could avoid (or, at least, reduce) cross-theoretical variation and lead to sound and indisputable generalizations. On the other hand, some major problems arise in this respect.

Let us briefly consider the point at issue. In the ToBI framework, the analysis of the intonational contour is based on the identification of a restricted number of tonal events, that represent elements of significant variation in the curve. These elements are *pitch accents* – H and L, also in binary combinations – and *boundary tones*. According to INTSINT, on the other hand, *pitch patterns* are coded using a broader set of tonal symbols (Top, Mid, Bottom, Higher, Same, Lower, Upstepped, Downstepped). These tones are combined in different ways and distinguished for their 'quality' (i.e., absolute *vs.* relative tones).

Different approaches thus use different representations, with pros and cons from a phonological point of view. However, when the focus of investigation is the search for universals in a cross-level perspective, a number of problems remain unsolved concerning the way in which data are interpreted at the PF interface, namely:

1. Are ToBI pitch accents and INTSINT tonal patterns comparable?
2. What is the *semiotic* value of the elements used for phonological representations?
3. Are there *other* significant phonetic properties for interface interpretation?
4. Are there intonation universals? In other words, is there a 'basic' tonal structure? And, if not, what determines cross-linguistic variation?
5. Should *Economy* considerations be taken into account? In other words, can tonal differences be interpreted as the consequence of the *interplay* between phonology and *other levels* of the grammar? For example, the presence/absence of a certain tonal event can be dependent on the realization of a *marked* syntactic structure? And, in this case, *what kind* of tone should be considered as the *unmarked* value?

These issues will be discussed in the presentation, making specific reference to the interface analysis of Focus and Topic constituents in Italian.

Beckman, M. E. and J. Pierrehumbert (1986) Intonational Structure in Japanese and English, *Phonology Yearbook*, 3, 255-309.

Hirst, D. and Di Cristo, A. (1998) *Intonation Systems. A Survey of Twenty Languages*, Cambridge: Cambridge University Press.

Martin, Ph. (1981). Pour une théorie de l'intonation : l'intonation est-elle une structure congruente à la syntaxe?. In Rossi, M., Di Cristo, A., Hirst, D., Martin, Ph., Nishinuma (eds.) *L'intonation de l'acoustique à la sémantique*, 234-271. Paris : Klincksieck.

MODELLIZZAZIONE DELLA PROSODIA E DEL TIMBRO PER LA SINTESI DEL PARLATO EMOTIVO

Mauro Nicolao, Carlo Drioli, Piero Cosi

Istituto di Scienze e Tecnologie della Cognizione - Sede di Padova "Fonetica e Dialettologia"
Consiglio Nazionale delle Ricerche

Viene descritta una procedura per la creazione di una funzione di trasformazione di un segnale vocale neutro in uno caratterizzato emotivamente. Questa funzione è stata sviluppata sulla base di un modello statistico, a mistura di funzioni gaussiane, dello spettro del segnale vocale.
Sono utilizzati, come segnali di riferimento per l'allenamento del modello, due database di segnali vocali creati "ad hoc": uno registrato da un parlatore, simulando l'emozione della rabbia, e uno neutro, con la stessa intonazione e durata dei fonemi, ottenuto con un sintetizzatore vocale per concatenazione di difoni, che utilizza la "voce" dello stesso parlatore.
Il modello a mistura di gaussiane, addestrato sui coefficienti mel-cepstrali estratti dal segnale neutro, è utilizzato per dividere questo spazio acustico in classi fonetiche equivalenti e per calcolare, per ogni classe identificata, i parametri delle funzioni di conversione.
Il metodo di trasformazione del segnale nel dominio delle frequenze ha fornito delle ottime prestazioni, come è stato dimostrato da un test percettivo in cui un segnale neutro convertito è stato riconosciuto come "arrabbiato".

Le consonanti labiodentali dello svizzero tedesco

Nadia Nocchi, Stephan Schmid

Università di Zurigo

Uno dei tratti peculiari del consonantismo dello svizzero tedesco consiste nella assenza del tratto di sonorità per le consonanti occlusive fricative e affricate. Le ostruenti possono essere tuttavia definite attraverso la contrapposizione fonologica *fortis* vs. *lenis*, oppure attraverso la semplice differenza tra lunghe e brevi (cfr. Willi 1995; Kraehenmann 2001). In ogni caso si tratta di un' opposizione binaria, con una sola eccezione: le consonanti labiodentali.

Dieth (1950: 362) propone per le labiodentali una tripartizione in cui accanto a [f] (per esempio in [ɒɻhɬ] 'aperto') e [ɷ8] ([ɒɻɷ8ɬ] 'forno') compare un terzo tipo di suono [x] trascritto con il diacritico [̀] (per esempio in [xʊ̀ɐu] 'bianco'); secondo l'Autore (1950: 203), di questo fonema esiste anche una variante definita come 'degenerierte Halbvokal', cioè come [X] 'bilabiale senza protrusione' (per esempio in]\X–ɑ̀l: 'svevo'). In base alle osservazioni del fonetista svizzero, sembra che questa consonante sia caratterizzata dalla presenza di sonorità, da notevole intensità e da breve durata. Queste caratteristiche ben si sposano con la definizione presente nella letteratura fonetica contemporanea per le approssimanti (cfr. Ladefoged 1975; Martínez Celdrán 2004). In questo caso, il suono in questione corrisponderebbe a [X].

Per verificare tale ipotesi abbiamo analizzato i correlati fonetici delle consonanti labiodentali dello svizzero tedesco. A tale proposito, sono stati registrati cinque parlanti maschi provenienti da quattro diversi cantoni della Svizzera tedesca (Argovia, Turgovia, Grigioni e Zurigo), ai quali è stato chiesto di tradurre dieci frasi dal tedesco standard (*Hochdeutsch*) nei loro rispettivi dialetti; per ogni frase tradotta, abbiamo richiesto tre ripetizioni. L'analisi acustica si è concentrata sulla misurazione della durata e dell'intensità dei segmenti; per l'approssimante labiodentale si è proceduto anche alla misurazione dei valori di F1, F2, F3 per verificare eventuali fenomeni di coarticolazione con le vocali contigue al segmento.

I risultati raccolti mostrano che per discriminare i tre tipi di consonanti il parametro essenziale è la durata che decresce nell'ordine]h͡ >]x͡: >]X͡; inoltre, l'approssimante si differenzia chiaramente dalle fricative in quanto caratterizzata da struttura formantica e quindi da barra di sonorità, da valori di intensità elevati, in presenza di durata ridotta.

DIETH, Eugen (1950). *Vademecum der Phonetik*. Bern.

KRAEHENMANN, Astrid (2001). Swiss German stops: geminates all over the word. *Phonology* 18: 109-145.

MARTÍNEZ-CELDRAN, Eugenio (2004). Problems in the classification of approximants. *Journal of the International Phonetic Association* 34: 201-210.

LADEFOGED, Peter (1975). *A Course in Phonetics*, New York.

WILLI, Urs (1995). "Lenis" und "Fortis" im Zürichdeutschen aus phonetischer Sicht. In: LÖFFLER, H. (a cura di.): *Alemannische Dialektforschung*. Basel: 253-265.

Uno strumento per l'annotazione e la modellizzazione prosodica di enunciati marcati per un sistema di sintesi vocale

Andrea Panizza – Francesca Tini Brunozzi – Enrico Zovato

Loquendo S.p.A. – Torino

andre.panizza@virgilio.it, tini.brunozzi@tin.it, enrico.zovato@loquendo.com

Questo lavoro si inserisce nell'ambito di un obbiettivo più vasto rivolto all'individuazione dei correlati acustico-prosodici tra stili di testo e stili di lettura per un sistema di sintesi vocale (TTS) [3] [8]. In particolare, si fa qui riferimento a uno strumento per l'annotazione e la modellizzazione prosodica di enunciati sintatticamente marcati allo scopo di replicare a livello acustico-prosodico determinate funzioni pragmatiche [2] [7].

In particolare, la specifica necessità di migliorare la lettura di frasi interrogative con un sistema TTS, ha richiesto l'analisi prosodica di costrutti sintattici interrogativi (in italiano le domande di tipo polare, wh-). Tale analisi è stata orientata da un progetto di frasi interrogative intese come domande 'vere', con funzione interrogativa pragmatica, e non, piuttosto, come domande retoriche inadeguate allo scopo di questo lavoro. L'analisi è stata poi condotta sul relativo *corpus* di segnali vocali (registrati in studio con speaker professionisti) al fine di ottenere una modellizzazione dei parametri e degli andamenti prosodici che fosse riproducibile con un sistema automatico di sintesi vocale.

Il progetto del *corpus* prevede che gli enunciati interrogativi vengano appositamente letti sia con intonazione dichiarativa sia con intonazione interrogativa, con l'obbiettivo di porre a confronto i due macro modelli prosodici.

Allo scopo di annotare le differenze morfologiche tra i gruppi di frasi, è stato sviluppato un tool di annotazione prosodica che permettesse di descrivere qualitativamente e morfologicamente gli andamenti di F0, energia e durata.

Questo strumento fornisce in modo automatico la segmentazione in sillabe (intese come unità acustiche e non grammaticali) [4] e per ciascuna di esse i valori di energia, durata e F0 [6]. Inoltre, per l'intero enunciato, viene riprodotta la curva della frequenza fondamentale in base alla quale l'annotatore fornisce l'informazione morfologica. Al fine di garantire la maggior uniformità possibile tra i diversi annotatori, è stato definito un alfabeto di etichette morfologiche riconducibili in modo univoco a una serie di gamme di valori. Si tratta di un procedimento di annotazione che, come altri sistemi sicuramente più esaurienti e precisi [1], mira in primo luogo a fornire semplici indicazioni sulla pendenza della curva di F0. Gli sviluppi di questo lavoro saranno l'utilizzo dei dati forniti in parte automaticamente e principalmente derivanti dall'annotazione manuale, per la definizione di modelli prosodici che possano pilotare efficacemente la selezione di unità acustiche in un sistema di sintesi vocale. In tal senso, se le dimensioni del *corpus* saranno sufficienti, il ricorso a strumenti di classificazione automatica potrebbe rivelarsi efficace [5].

[1] Avesani C., Cosi P., Fauri E., Gretter R., Mana N., Rocchi S., Rossi F. e Tesser F. 2004, "Definizione ed annotazione prosodica di un database di parlato-letto usando il formalismo ToBI", in Albano Leoni F., Cutugno F., Pettorino M., Savy R., *Atti del Convegno Nazionale "Il parlato italiano"*, Napoli, M. D'Auria Editore, p. CD- ROM M01.

[2] Avesani C., Vayra M., 2000, "Costruzioni marcate e non-marcate in italiano: il ruolo dell'intonazione", X Giornate di Studio del GFS (Gruppo di Fonetica Sperimentale), Napoli, pp. 1-15.

[3] Balestri M., Pacchiotti A., Quazza S., Salza P., and Sandri S., 1999, "Choose the Best to Modify the Least: a New Generation Concatenative Synthesis System", in *Proceedings of EUROSPEECH 1999*, Budapest, pp. 2291-2294.

[4] Bertinetto P.M., 1981, *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici*, Firenze, Accademia della Crusca.

[5] Cosi P., Avesani C., Tesser F., Gretter R., Pianesi F., 2003, "On the Use of Cart-Tree for Prosodic Predictions in the Italian Festival TTS", in Cosi P., Magno Caldognetto E., Zamboni A., 2003, in *Voce, Canto, Parlato – Studi in onore di Franco Ferrero*, UNIPRESS, Padova, pp. 73-81.

[6] Cutugno F., D'Anna L., Petrillo M., Zovato E., 2002, "APA: Towards an automatic tool for prosodic analysis", Speech Prosody 2002, Aix-en-Provence.

[7] Firenzuoli V., 2001, "Verso un nuovo approccio allo studio dell'intonazione a partire da corpora di parlato: esempi di profili intonativi di valore illocutivo dell'italiano", in Maraschio N., Poggi Salani T., (a cura di), 2001, *Atti del XXXIV Congresso internazionale degli studi della SLI*, 19/21 ottobre 2000 - Firenze, Bulzoni, Roma.

[8] Gili Fivela B., Quazza S., 1997, "Text-to-Prosody Parsing in an Italian Synthesizer. Recent Improvements", *Proceedings of EUROSPEECH '97*, Rhodes, Greece, September Vol. 2, pp. 987-990.

UN SISTEMA DI SPEAKER IDENTIFICATION PER LA SEGMENTAZIONE AUTOMATICA DI VIDEOGIORNALI

Gennaro Percannella[1], Carlo Sansone[2], Domenico Sorrentino[1], Mario Vento[1]

[1]Dipartimento di Ingegneria dell'Informazione e Ingegneria Elettrica, Università di Salerno
[2]Dipartimento di Informatica e Sistemistica, Università di Napoli "Federico II"

In questo lavoro è presentato un sistema di speaker identification che opera in tempo reale. Il presente lavoro è inquadrato nel framework più generale legato alla segmentazione ed alla indicizzazione automatica dei videogiornali. In tale contesto, l'impiego dell'informazione audio può permettere una migliore individuazione delle notizie, rispetto al caso in cui si faccia uso della sola informazione video: ad esempio, si pensi alla situazione in cui le immagini di un servizio scorrono sullo schermo, mentre lo speaker che le sta commentando non è visibile.

Il sistema proposto effettua una classificazione di tipo *text-independent*: in pratica, per identificare lo speaker non è richiesto che questi pronunci necessariamente uno specifico insieme di parole. In particolare, a valle di una prima fase di pre-elaborazione del segnale, vengono estratte le seguenti feature: Linear Predictive Cepstral Coefficients (LPCC), Post Filter Linear (PF) e Mel Filtered Cepstral Coefficients (MFCC). Tali feature permettono di cogliere le proprietà più discriminanti dello speaker sia nel dominio del tempo che in quello della frequenza.

La classificazione è effettuata attraverso una rete neurale LVQ. In fase di addestramento è stato, inoltre, impiegato un algoritmo FSCL che consente un uso ottimale dei neuroni in quanto ne evita il sottoutilizzo.

Per la sperimentazione del sistema è stato realizzato un dataset costituito da segmenti audio estratti dalle registrazione di dodici telegiornali nazionali, in cui sono presenti dieci differenti anchor di entrambi i sessi. Le acquisizioni dell'audio di ogni singolo presentatore, sono effettuate in edizioni diverse dello stesso notiziario, supponendo che le caratteristiche principali dell'audio, non cambiano nelle varie registrazioni.

Le prove sperimentali hanno dimostrato che l'uso combinato di tutte le feature citate in precedenza consente di ottenere prestazioni superiori rispetto all'uso di un singolo set di feature, raggiungendo un tasso di riconoscimento superiore al 99% quando la lunghezza dei brani audio da classificare dura almeno 1,5 secondi.

Effects of Raddoppiamento Fonosintattico on Tonal Alignment in Neapolitan Italian.

Caterina Petrone

Laboratoire Parole et Langage - CNRS, Aix-en-Provence, France
caterina.petrone@lpl.univ-aix.fr

Raddoppiamento Fonosintattico (RF) is a phenomenon found in Central and Southern varieties of Italian, by which an oxytone word1 lengthens the initial consonant of word2. One possible interpretation is that RS is a resyllabification process, depending on the same constraints on syllable structure as within-word gemination. If this is correct, we also expect both phenomena to be signaled by the same cues. As a consequence of syllable structure and vowel length differences, in Neapolitan Italian the peak of the L*+H nuclear accent is later in words containing a geminate (*nonno*) than in words containing a singleton consonant (*nono*). This paper reports a pilot study on effects of RS on tonal alignment in Neapolitan Italian. The results show that, as in the case of items containing geminates, the H target is later when RS is applied. This suggests that RS can be regarded as a process of syllable restructuring.

Analisi preliminare delle strutture tonali del ditammari (Benin)

Antonio Romano & Opportune Mouti

Dip. di Scienze del Linguaggio - Università di Torino, Italia
antonio.romano@unito.it

Il ditammari è la lingua parlata dai Batammariba o Batammarba, appartenenti a un gruppo etnico (Tammari) di stanza tra il Bénin e il Togo.

Le cifre riguardanti il numero di parlanti di questa lingua nel Bénin oscillano significativamente a seconda che si tenga conto delle stime della base di dati "Ethnologue" (20000 parlanti) o del censimento ufficiale del governo del Bénin del 1999 (più di 300000 Batammariba; è invece nel Togo che il numero di Batammariba è stimato a circa 27000).

Il Ditammari, che è una lingua tonale, appartiene al gruppo voltaico (Gur) della famiglia Niger-Kordofaniana (*Ethnologue* la classifica nel *phylum* Niger-Congo, del gruppo Gur e del sotto-gruppo Oti-Volta: codice ISO/DIS 639-3: tbz).

Le varietà di cui qui si tiene conto, sono soprattutto quelle delle comunità di Natitingou e di Boukoumbé (altopiano dell'Atacora). La parlata di quest'ultimo centro (popolato dal sotto-gruppo Bacaaba) è particolarmente interessante perché questo comune è sicuramente stato il primo scalo importante nella migrazione del popolo Tammari dalla periferia dell'impero Mossi verso l'attuale Bénin (XVI s.). La varietà di questa comunità, designata Bacaaba per via dei contatti con i vicini gruppi di lingua dicaa, è tuttavia poco influenzata da questi e si propone come un ditammari più conservativo.

È tuttavia l'altra varietà, detta ditenb— e parlata dal sotto-gruppo Batemb—ba (a Natitingou), che dispone di documenti scritti e di un'ortografia (testi scolastici e Bibbia, v. N'DAH, 2002). Oltre alle variazioni tonali, le principali caratteristiche segmentali della varietà scritta risiedono nell'evoluzione di /a/ in /G/ e nell'assenza di dittongamento di /–%(es.: al taãnw–ãa 'cane' dei Bacaaba corrisponde il taãn–ãG dei Batemb—ba).

Le strutture tonali di questa varietà mostrano una relativa semplicità di definizione e di realizzazione (anche se il problema della loro notazione viene complicato dalla presenza di vocali lunghe o doppie che dànno luogo a combinazioni di toni o a toni modulati; cfr. CREISSEL, 1994) che abbiamo studiato sulla base di più di 300 forme e delle loro realizzazioni da parte di due parlanti (un uomo e una donna, registrati a cura dell'autrice OM).

Tuttavia, in questo lavoro, ci proponiamo di discutere il problema della classificazione dei toni lessicali di questa lingua distinguendo un piano tonetico da uno tonologico.

Infatti, se su un piano tonetico diversi autori distinguono tre o quattro toni (cfr. per es. NICOLE, 1983; NATA, 1991; KOUAGOU, 1991), è dalla riflessione sulle allotonie distribuzionali e individuali che emerge il ruolo funzionale di soli tre toni.

CREISSELS Denis, *Aperçu sur les structures phonologiques des langues négro-africaines*, Grenoble, Ellug, 1994.

KOUAGOU Philippe N., "Speech, Phonology and their impacts on Otammari Students learning English as a Foreign Language", *Mémoire de maîtrise,* Cotonou, Université Nationale du Bénin, 1990-1991.

NATA Théophile, *Abrégé de grammaire ou nouveau guide de lecture du Ditammari*, Cotonou, Université Nationale du Bénin, 1991.

N'DAH Antoine (éd.), **Kaà-nkáwcãi a'náantù** *Syllabaire ditammari* - ditCãnmabGbìipátri ti d–&&dì ka di yu katú, Tomes 1 et 2, *CNLD* (Commission Nationale Linguistique Ditammari), République du Bénin (BennGG Tenkpàti), 2001-2002.

NICOLE Jacques, *Etudes linguistique préliminaires dans quelques langues du Togo*, Lomé, Société Internationale de Linguistique, 1983.

Uno studio degli esiti metafonici nei dialetti dell'area Lausberg: un'introspezione sulla natura della sillaba

Luciano Romito, Stillo Francesca, Lio Rosita e Galatà Vincenzo

Laboratorio di Fonetica, Università della Calabria

Tra i processi fonetici e fonologici che si sviluppano all'interno di una lingua, i fenomeni di armonizzazione tra due o più foni rivestono un ruolo molto importante. Tra questi vi è la *metafonia*, un fenomeno che interessa non solo la maggior parte delle lingue, ma anche molti dialetti italiani.

Lo scopo di questo lavoro è: 1) accertare l'esistenza della metafonia in una particolare zona della Calabria che per il suo grado di arcaicità è stata definita 'zona arcaica calabro-lucana' o 'zona Lausberg' dal nome del linguista che l'ha esplorata e analizzata per primo, e che si trova al confine tra la Lucania del sud e la Calabria del nord; 2) verificare, attraverso analisi acustiche, se la metafonia è avvenuta per innalzamento o per dittongazione; in questo secondo caso verificare se il dittongo formato è di tipo ascendente o discendente; 3) verificare se esiti come ['li:t:u] sono conseguenze di una metafonia per innalzamento (e la cosa risulterebbe alquanto anomala per la lunghezza della vocale tonica in sillaba chiusa) o se invece non siano conseguenza di un monottongamento di un precedente processo di metafonia per dittongamento (in questo caso è necessario indagare su vecchie registrazioni effettuando una sorta di Analisi Acustica diacronica).

Inoltre, oltre all'importanza morfologica che riveste il processo di metafonia in questi dialetti con l'avvenuta caduta della vocale finale, risulta interessante constatare come spinte di tipo fonetico possano forzare opposizioni fonologiche consolidate (vocali toniche brevi in sillaba chiusa e lunghe in sillaba aperta) e come la teoria possa spiegare situazioni di passaggio da un sistema all'altro.

Due realtà linguistiche urbane a confronto: quali parametri prosodici per un modello plausibile?

Elena Sardelli

Università di Pisa – Fondazione Ugo Bordoni

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. Sono stati analizzati dialoghi di parlato semi-spontaneo (*corpus* CLIPS)[1] di parlanti maschi di Roma e Milano e sono stati riconosciuti dei "turni" e "clausole intonative" compiute (aventi cioè indipendenza propria, ma non sempre sinonimi di sintagmi intonativi). 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 diatopiche. 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.

Le frasi sono state distinte in base alla maggiore ricorrenza e essenzialmente in: a) dichiarative, b) domande *sì-no*, c) domande *wh-*. Per entrambe le varietà in oggetto l'andamento finale di frase per i contesti di tipo b), ovvero domande *sì-no*, presenta un movimento discendente-ascendente, che si differenzia in base al gradiente di modulazione ad esso associato (diversi valori medi misurati in ST dei due diversi movimenti). La trascrizione ToBI di tali andamenti presenta differenze di associazione, essendo H*+L H% per i parlanti romani e H+L* H% per quelli milanesi. Per i contesti romani il movimento discendente si allinea, nella grande maggioranza dei casi, con l'inizio della vocale relativa.

Per i contesti di tipo c), domande *wh-*, per i quali il movimento finale rimane discendente-ascendente, elemento discriminante sembra essere l'escursione melodica che su di essi si realizza: i parlanti milanesi realizzano forti escursioni tonali sul primo movimento (in media 7,8 ST) rispetto ai parlanti romani ed effettuano una risalita tonale finale meno pronunciata rispetto alle produzioni romane.

Per i contesti di tipo a), ovvero frasi di tipo dichiarativo (per le quali si è dovuto realizzare una particolare suddivisione dei turni intonativi registrati, al fine di una maggiore capacità comparativa) sono state riconosciute le sillabe prominenti, trascritte come H* sia per i parlanti romani che per i parlanti milanesi. Si è notato però che mentre in contesti romani tale H* è un tono associato al più alto raggiunto da Fo (nella porzione selezionata), stessa cosa non può dirsi per la varietà milanese. Infatti per questi contesti la vocale che riceve prominenza si associa ad un andamento tonale essenzialmente stabile, presso la parte alta del *range* del parlante, al quale segue il picco massimo di Fo. L'analisi, al momento in corso, dovrebbe essere in grado di produrre buoni risultati soprattutto per quanto riguarda il riconoscimento di contesti interrogativi. Per ciò concerne i contesti dichiarativi invece, sembra ancora rimanere *condicio sine qua non* la fattibilità di modelli in grado di recuperare specifici processi di ancoraggio al livello segmentale.

---

[1] Per gentile concessione del C.I.R.A.S.S., Università degli Studi di Napoli Federico II.

Spectrum shapes and place of articulation of the voiceless stops /t/ and /ṭ/ in Moroccan Arabic

Karim Shoul

Laboratoire de phonétique et phonologie, La Sorbonne Nouvelle, Paris, France

The present paper is an acoustic study of the voiceless stop /t/ and its corresponding emphatic /ṭ/ through spectrum shapes in Moroccan Arabic (MA). It is known that both of /t/ and /ṭ/ are alveolar stops and the difference between them is the presence of a secondary pharyngeal articulation which accompanies the emphatic (see Ghazeli 1977, among others). We tried tosee whether the alveolar articulation of our consonants can be reflected in a spectral shape.

According to Stevens (1980) and Ohde et Stevens (1983), consonant release signals the stop consonant place of articulation. Therefore, the alveolar stops are produced through spectra with a release having a rising amplitude in high frequencies. We found that the release of both /t/ and /ṭ/ in MA have a rising amplitude at the frequency 5000 Hz which means that both of the stops are alveolars. However, the amplitude in the high frequencies is higher for /t/ than for /ṭ/. This is due to the fact the non emphatic /t/ is produced with a longer anterior cavity.

Stevens (1989) believes that the highest amplitude spectral peak at the release of a stop is a cue to alveolar articulation if it corresponds to F4, F5 or higher formant of the following vowel. We found that the peak having a higher amplitude for /t/ and /ṭ/ manifested at the frequency 5000 Hz in spectra. This frequency corresponds to F4 or F5 of the following vowel /a/. Therefore, both of the non emphatic /t/ and the emphatic /ṭ/ are produced with the same alveolar articulation even though the latter has, in addition, a secondary pharyngeal constriction.

Gerarchie temporali e livelli di annotazione fonetica

Serena Soldo & Francesco Cutugno

Dipartimento di Scienze Fisiche – gruppo NLP, Università degli Studi di Napoli – "Federico II"
{cutugno@na.infn.it, brixie@blu.it}

Il presente lavoro fa parte di uno studio più ampio nell'ambito dell'annotazione linguistica e in particolare dei grafi di annotazione formalizzati da Bird e Liberman (2000).

Il concetto di gerarchia è insito nella struttura formale dei grafi di annotazione e si riferisce specificamente alle differenti estensioni del dominio temporale di ogni livello di *labelling*. Almeno in linea teorica, un livello di annotazione che influenza un intervallo temporale più breve dovrebbe avere una gerarchia, riflessa anche nel "peso" linguistico dell'elemento associato, inferiore a quella evidenziata da un livello le cui unità si estendono su intervalli più grandi. Nella osservazione empirica del dato linguistico, in realtà, ci si scontra con l'impossibilità di indicare una gerarchia che valga in maniera generale, poiché il linguaggio è caratterizzato da una variabilità molto forte che impedisce l'individuazione di una struttura gerarchica fissa. Nel nostro lavoro mostreremo inizialmente come l'organizzazione dei grafi possa talvolta mettere in evidenza sovrapposizioni temporali fra differenti livelli di annotazione tali da indurre stravolgimenti della gerarchia temporale rispetto a quella che ci si aspetterebbe comunemente. Ad esempio ci si trova non di rado di fronte a casi in cui le sillabe esercitano un dominio temporale più ampio delle parole in cui si trovano.

Un ulteriore elemento che rende più articolato e complesso il tema delle gerarchie è la presenza, nei grafi, di livelli di annotazione istantanei solitamente connessi a eventi di natura prosodica.

La riduzione delle dimensioni temporali di dominio di un evento fonico verso il limite istantaneo non pone queste unità ad un livello gerarchico minimo, bensì introduce il problema del rapporto fra domini reali e domini virtuali, cioè della determinazione della reale influenza linguistica di eventi culminativi o terminali che nella loro "istantaneità" conservano comunque memoria di una evoluzione significativa.

Proporremo un metodo per la determinazione delle gerarchie tra i livelli di annotazione linguistica di livello fonetico, sia segmentale che soprasegmentale mediante un'analisi statistica di corpora dialogici. In altre parole, proveremo ad estrapolare la gerarchia vigente in un particolare corpus, analizzando statisticamente i dati che esso contiene. Definiremo differenti tipi di gerarchie: definiremo e misureremo la gerarchia media di un intero corpus, la confronteremo con quelle ricavate da sottoparti e produrremo un metodo per sondare la struttura gerarchica in uno specifico istante temporale.

Mostreremo come questo metodo introduce un importante vantaggio nella interrogazione dei corpora linguistici. La nostra proposta include infatti anche una estensione del linguaggio di interrogazione AGQL implementato nelle librerie AGTK. Questa estensione permette di esprimere anche query basate sulle gerarchie.

SINTESI VOCALE CONCATENATIVA PER L'ITALIANO TRAMITE MODELLO SINUSOIDALE

Giacomo Sommavilla, Carlo Drioli, Piero Cosi

Istituto di Scienze e Tecnologie della Cognizione - Sede di Padova "Fonetica e Dialettologia"
Consiglio Nazionale delle Ricerche, Padova, Italy

L'argomento di questo lavoro consiste nello sviluppo di un sintetizzatore per l'italiano da testo scritto (Text-To-Speech - TTS) operante nel dominio delle frequenze.

L'architettura del sistema è pensata per un'integrazione con il sistema FESTIVAL per cui sono state utilizzate le sue stesse procedure per la trattazione del testo, per la conversione grafema-fonema, per l'accentazione, la sillabificazione e per il "*prosody matching*", cioè per la trasformazione dei parametri di prosodia (intonazione e durata memorizzate nel corrispondente file nel formato "*.PHO*") calcolati nell'analisi del linguaggio naturale in variabili utili a pilotare le funzioni di "*pitch shifting*" e "*time stretching*". L'architettura del sistema si differenzia per il motore vero e proprio di sintesi audio e concatenazione di difoni che avviene nel dominio delle frequenze. La sintesi audio viene, infatti, effettuata, non nel dominio del tempo come con MBROLA, ma mediante la anti-trasformata dal dominio frequenziale a quello temporale.

Gli elementi concatenativi di base (*difoni*) del corpus MBROLA, utilizzato nella versione attualmente distribuita di FESTIVAL per l'italiano, sono stati convertiti e trasformati nel dominio della frequenza mediante analisi SMS (*Spectral Modeling Synthesis*) per cui ad ogni difono corrisponde un particolare file nel formato "*.SDIF*". Il formato SDIF (*Sound Description Interface Format*) è adatto ai nostri scopi perchè permette una compressione *senza-perdite* dei valori calcolati dall'analisi SMS incapsulati nella rappresentazione spettrale di frame audio consecutivi. In particolare, i tre vettori che memorizzano ampiezze, frequenze e fasi delle parziali e l'inviluppo spettrale per parte residuale associato ogni frame sono sicuramente i parametri più importanti di questa rappresentazione. Il sistema di analisi e sintesi è stato sviluppato all'interno del framework CLAM (ideale per l'approccio SMS), realizzato dal Music Technology Group (MTG) della Universitat Pompeu Fabra di Barcellona.

Ad una maggiore complessità computazionale, dovuta al fatto che nella sintesi audio sono coinvolte le procedure di ricostruzione dello spettro e di inversione mediante IFFT, si possono opporre i vantaggi che derivano dall'operare nel dominio della frequenza (rispetto alle operazioni effettuate nel dominio temporale) e in particolare, quelli derivanti dall'utilizzazione della tecnica SMS, possono essere così schematizzati:

- una rappresentazione del segnale più versatile e potente per le elaborazioni prosodiche (tempo, pitch) e timbriche (inviluppo spettrale);
- la possibilità di sfruttare i risultati dell'analisi SMS nella concatenazione dei difoni;
- la distinzione tra parte sinusoidale (deterministica) e residuale (stocastica) che permette di trattare diversamente fonemi vocalizzati da quelli non-vocalizzati.
- l'utilizzo dello stesso database di difoni MBROLA utilizzato nell'attuale versione di FESTIVAL per l'italiano.

I valori di H1-A2 e H1-A3 come correlati della intensità “rivisitata”. Aspetti e problemi.

Arianna Uguzzoni

Alma Mater Studiorum - Università di Bologna

Negli ultimi tempi gli studiosi hanno guardato con occhi nuovi alla intensità sia come grandezza fisica sia come dimensione percepita . A questa “rivisitazione” dell’intensità dei suoni linguistici ho già dedicato una breve nota (in Studi in onore di Franco Ferrero, 2003: 299-302).

Nel presente contributo da un lato mi soffermerò sugli aspetti generali della svolta metodologica derivata dallo spostamento di interesse dalla intensità dell’intero spettro (‘overall intensity’) alla distribuzione della intensità in differenti parti dello spettro (‘spectral balance’). Dall’altro farò riferimento ad alcune ricerche condotte su lingue diverse e su fenomeni di varia natura dalle quali risulta una caratteristica comune, cioè l’intensità relativamente più grande nelle bande di frequenza superiori a 500 Hz (‘high frequency emphasis’) che si riflette in valori minori di H1 - A2 e H1 - A3 (con H si indica l’ampiezza delle armoniche, con A l’ampiezza delle formanti). Qualche esempio. In più lingue germaniche lo spettro presenta una enfasi nelle regioni alte che è maggiore nelle vocali accentate rispetto a quella che si osserva nelle vocali non accentate. Lo stesso fenomeno si riscontra nelle vocali rilassate del tedesco che hanno una maggiore enfasi spettrale rispetto alle vocali tese.

E’ in tale cornice che si inserisce una recente ricerca sulla intensità di vocali accentate brevi e vocali accentate lunghe di un dialetto frignanese. Le misurazioni fatte finora consentono di dire, sia pure con cautela, che le vocali brevi si differenziano dalle loro controparti lunghe per valori minori di H1 - A2 e H1 - A3 e che esse sono quindi caratterizzate da enfasi spettrale, oltre che da altre proprietà acustiche e cinematiche trovate nel corso di indagini precedenti. I dati sperimentali presi in considerazione in questa sede sollevano problemi di interpretazione della intensità “rivisitata”, con particolare riguardo alla relazione tra l’incremento dell’intensità nel campo delle medie e alte frequenze e i meccanismi pneumo-fono-articolatorii che vi sono sottesi.

Per esempio nel caso degli effetti esercitati sull’intensità dall’accento lessicale si è ipotizzato che l’enfasi spettrale dipenda soprattutto da un aumento dello sforzo fisiologico richiesto per la produzione di vocali accentate (‘stressed’) e da fattori laringali quali la maggiore velocità nella fase di chiusura della glottide e la più ripida pendenza dell’impulso glottico. Ma per altre distinzioni, che hanno anch’esse come correlati acustici differenze nell’ammontare dei valori di H1 – A2 e H1 – A3, non si sa abbastanza in merito ai fatti fisiologici soggiacenti sia a livello della sorgente sia a livello del filtro. E sarebbe interessante scoprire come coagiscono e interagiscono attività laringale e gesti sovralaringali.

Le frequenze dei foni e delle loro co-occorrenze intra- e inter-sillabiche in due bambini dai 9 ai 27 mesi di eta'.

Claudio Zmarich e Elena Luppari
Istituto di Scienze e Tecnologie della Cognizione - C.N.R., Sede di Padova

Per *babbling* si intende la produzione di una sequenza di sillabe di tipo consonante-vocale (C-V) dotate di organizzazione ritmica e temporale simile a quella del parlato adulto (Vihman, 1996). Secondo un'influente teoria che descrive la struttura del *babbling* (Davis & MacNeilage, 1995), il bambino acquisisce schemi motori e controllo articolatorio da una singola base motoria universale consistente nell'alternanza ritmica tra mandibola aperta e chiusa, che genera un effetto acustico percepito come sillaba. Inizialmente la struttura intrasillabica di tale "sillaba" permette che le consonanti anteriori (coronali) co-occorrano quasi esclusivamente con le vocali anteriori, le consonanti posteriori (dorsali) con le vocali posteriori e le consonanti labiali con le vocali centrali. Tra sillabe consecutive ci sarebbe una maggior variazione nella dimensione articolatoria "alto/basso" piuttosto che in quella "anteriore/posteriore". La ragione di siffatta organizzazione intra- e inter-sillabica risiederebbe nel vincolo biomeccanico che lega il posizionamento passivo di labbra e lingua al movimento della mandibola, che di fatto trasporta gli altri due articolatori. Queste restrizioni di tipo neurofisiologico potrebbero rendere conto del fatto che il babbling non risente dell'influenza delle lingue native (Vihman, 1996), cioè manifesta proprietà universali. Da qui sorge la necessità di documentare l'inizio dell'influenza linguospecifica e le modalità con cui essa si manifesta.

L'importanza del presente contributo risiede nell'ampiezza dell'arco temporale in cui sono stati audioregistrati due bambini, un maschio e una femmina, a partire dal 9° fino al 27° mese di età, a intervalli di 3 mesi, nel corso di situazioni di gioco, alla presenza delle madri. Questo periodo è cruciale per studiare le relazioni tra babbling (semanticamente opaco) e primo vocabolario, caratterizzato da una chiara e stabile associazione forma-significato. Le produzioni infantili sono state trascritte foneticamente con i simboli IPA e le capacità fonetiche sono state analizzate attraverso due modalità. La prima valuta la produzione dei singoli foni a sua volta con due procedure:

1) per il babbling e per un vocabolario < 10 parole, le frequenze dei tipi vocalici e consonantici vengono calcolate sia complessivamente che in funzione della loro posizione nella sillaba e nella parola (statistiche sui *tokens*);

2) per un vocabolario > 10 parole (*type*), che i due bambini raggiungono ai 15 mesi, e limitandosi alle prime 50, viene calcolato l'inventario fonetico applicando i criteri di Stoel-Gammon (1985): un fono o un gruppo consonantico sono attestati in posizione iniziale e non iniziale di sillaba e di parola solo se presenti in almeno due diverse "parole" (statistiche sui *types*).

La seconda modalità valuta le ipotesi di MacNeilage & Davis: per quella intra-sillabica, è calcolata la frequenza delle associazioni dei foni consonantici con i foni vocalici classificati per luogo di articolazione, mentre per quella inter-sillabica, è valutato se tra le sillabe consecutive non reduplicate (cioè non ripetute uguali) sia effettivamente più frequente la variazione di modo rispetto a quella di luogo.

I risultati, inquadrati in una prospettiva longitudinale, sono messi in relazione alle frequenze di occorrenza delle strutture foniche dell'italiano tratte da (1) una lista di parole dal Primo Vocabolario del Bambino (Caselli e Casadio 1995, appendice B), (2) alcuni studi di frequenza sulla lingua italiana, e vengono alla fine confrontati con i dati sul primo sviluppo fonetico/fonologico di bambini che acquisiscono l'inglese.

Confini di sillaba, confini di parola e lunghezza fonologica in area frignanese: analisi cinematica dei gesti labiali

*Claudio Zmarich e °Arianna Uguzzoni

**Istituto di Scienze e Tecnologie della Cognizione del C.N.R., sede di Padova*
°Alma Mater Studiorum – Università di Bologna

Molti dialetti dell'area emiliana fanno un uso distintivo della lunghezza vocalica. Nel caso specifico dei dialetti frignanesi (Appennino modenese) si osservano tre proprietà distribuzionali. Le opposizioni: (a) sono sottoposte al vincolo della presenza dell'accento di parola; (b) sono limitate ad un sottoinsieme delle qualità vocaliche (/**e**, O, □, **a** / vs /**e**:, O:, □:, **a**:/); (c) sono possibili sia in sillaba non finale ['CV(:)CV] sia in sillaba finale [CV(:)C e CV(:)]. Le caratteristiche acustiche del fenomeno frignanese sono state oggetto di varie indagini (Uguzzoni,Busà, RID 1995: 7-39).

Il presente lavoro continua l'analisi cinematica dei gesti labiali che già aveva permesso agli autori di spiegare le distinzioni di lunghezza vocalica come dovute a una diversa coordinazione temporale (o rapporto di fase) tra il gesto di apertura, relativo alla vocale, e il gesto di chiusura, relativo alla consonante (Zmarich, Uguzzoni, Ferrari, Atti delle XIII Giornate GFS, Pisa, 2003:295-306). L'analisi che allora aveva dovuto limitarsi ai soli fatti articolatori pertinenti alla distinzione di lunghezza vocalica questa volta potrà essere allargata allo studio del comportamento articolatorio derivato dall'interazione tra quest'ultima variabile e le variabili costituite dalla presenza di confini diversi sia di sillaba sia di parola. Questo studio è di tipo innovativo perché in passato l'influenza di tali fattori prosodici sulla dimensione temporale del parlato (la più modificata) è stata quasi sempre indagata soltanto a livello acustico.

La registrazione cinematica è stata eseguita su un parlante nativo (il secondo autore), che ha prodotto enunciati in cui coppie di pseudo parole target che si opponevano per la durata delle vocali accentate (cioè /**a**/, /O/, /□/ ϖσ /**a**:/, /O:/, /□:/) erano inserite in frasi cornice. Le parole target erano relative a 3 set, il primo costituito da bisillabi piani ('CVCa), il secondo da monosillabi in sillaba chiusa (CVC), e il terzo da monosillabi in sillaba aperta (CV). In quest'ultimo tipo la vocale era seguita senza pause intermedie dall'occlusiva bilabiale che cominciava la parte destra della frase cornice.

Per ciascuno dei gesti di apertura (da C a V) e di chiusura (da V a C) da un lato sono state misurate durata, ampiezza e velocità massima, dall'altro sono stati enucleati e applicati indici cinematici di tipo indiretto che rivestono un'importanza primaria per la teoria della Fonologia Articolatoria, quali la *stiffness* (stimata dal valore del rapporto tra velocità massima e ampiezza del gesto fonetico) e il rapporto di fase tra gesti consecutivi.