



**Parlare con le persone, parlare alle macchine:
la dimensione interazionale della comunicazione verbale**

*People talking to people, people talking to machines:
the interactional dimension of speech*

6° Convegno Nazionale dell'Associazione Italiana di Scienze della Voce

Napoli Centro Congressi di Via Partenope

3-5 febbraio 2010

Booklet of abstracts

Organizzato da

Università degli Studi di Napoli "Federico II"

Università degli Studi di Salerno

Associazione Italiana di Scienze della Voce

Booklet of abstracts

SESSIONE A TEMA.....5

- F. Orletti *Multimodalità e punti di transizione nei turni di conversazione*
- P. Sorianoello *Il tipo esclamativo. Analisi e percezione delle risorse prosodiche*
- E. Zovato, P. L. Salza, K. Bao Nguyen, M. Danieli *Caratterizzazione di due stili di eloquio emozionale in base all'enfasi e ai confini prosodici percepiti*
- I. Alfano, R. Savy *Le richieste di informazione nel parlato dialogico: un'analisi prosodica in contesti pragmatici orientati in lingua italiana e spagnola*
- C. Ericsson. G. Abete *Aspetti fonetici e interazionali di un marcatore discorsivo di area napoletana*
- G. Mignini *Mixed initiative spoken dialogue system in the sales force domain*

SESSIONE TECNOLOGICA.....21

- P. Cosi, M. Nicolao *EVALITA-ISTC comparison of open source tools on clean and noisy digits recognition tasks*
- G. Costantini, A. Paoloni, M. Todisco *Misura oggettiva dell'intelligibilità del parlato in applicazioni forensi*
- A. De Rosa, F. Cutugno *Nuove funzionalità per il portale CLIPS*
- S. Romano *SpLaSH: il sistema general purpose per la gestione di annotazioni time-aligned e text-aligned*
- A. Origlia, V. Galatà *Evaluating speech rate by referring on an intensity-based syllable nuclei search algorithm*

SESSIONE POSTER.....36

- E. Orso, M. Colegari, F. Rapa, C. Zmarich *L'emergere della fonologia nei bambini dai 18 ai 27 mesi: analisi statistica degli errori di sostituzione*
- L. Spreafico *I riempitivi nel parlato plurilingue*
- L. Righero, M. Danieli *Polirematiche interiettive ed espressione delle emozioni*

- A. Origlia, G. Abete *A syllabification algorithm using energy, pitch and harmonicity related features*
- S. Giulivi, M. Vayra, C. Zmarich *Dal babbling alla lingua materna. Tendenze universali e aspetti linguo-specifici nello sviluppo fonetico di due bambini italiani*
- S. Schmid *La neutralizzazione del contrasto di lunghezza delle fricative nel dialetto di Zurigo*
- G. Tisato *NavigAIS - AIS Digital Atlas and Navigation Software*
- L. Mori *Mixed Maltese English: fenomeni di interferenza fonetica lungo il continuum Maltese - Inglese nel repertorio linguistico di Malta*
- V. Galatà, L. Romito *Gli italiani e il decoding di emozioni vocali in quattro lingue europee*
- F. Orletti, A. Paoloni *Identificazione dei segnali*
- Marco A. Piccolino-Boniforti *Distinguere pattern acustici con identica trascrizione fonemica ma struttura morfologica differente: un modello computazionale*

SESSIONE FONETICO-FONOLOGICA.....63

- A. Stella *L'influenza della L1 nell'intonazione della L2: l'allineamento tonale in parlanti con differenti livelli di competenza*
- C. Pisciotta, S. Bernardini, E. Agazzi, N. Crivelli, F. Manni , R. Perosa, D. Stocco, C. Zmarich *Indici fonetici predittivi di balbuzie cronica in età prescolare: studio di un caso*
- G. Schirru *La pendenza spettrale come indice acustico della tensione laringea*
- B. Gili Fivela, S. D'Apolito, A. Stella, F. Sigona *Domain initial strengthening: dati acustici ed articolatori relativi a due varietà di italiano*
- M. C. Felloni, D. Avesani *La percezione della interrogativa globale nell'italiano regionale di Parma*
- V. Retaro *L'abbassamento di [ɛ] nelle varietà dialettali nord-vesuviane: un'analisi socio fonetica*

WORKSHOP "GRUPPO DI FONOLOGIA E LINGUISTICA".....83

- P. M. Bertinetto, I. Ricci, Z. Na *Le nasali sorde dell'ayoreo: prime prospezioni*
- G. Bocci, C. Avesani *Givenness, deaccentazione e il ruolo di L* nell'italiano di Toscana*

- M. Grimaldi, A. Calabrese, F. Sigona, B. Sisinni, L. Garrapa *La variazione micro-parametrica nei processi di armonizzazione del Salento meridionale. Studio articolatorio tramite Articolografo Elettromagnetico 3D ed Ultrasuoni*
- F. Cangemi, R. Delucchi, M. Loporcaro, S. Schmid *Vocalismo finale atono "toscano" nei dialetti del Vallo di Diano (Salerno)*
- L. Mereu *Variazione, universali linguistici e contributo dell'intonazione*
- A. Romano, F. Gambino *Cacuminali calabresi: modi e luoghi d'articolazione alla luce di misurazioni acustiche e indagini per risonanza magnetica (IRM)*
- B. Sisinni *Metodologie negli studi cross-linguistici. Riflessioni sulla relazione fra dati acustici e percettivi*
- A. Uguzzoni *Invarianza e variabilità, ieri e oggi. Spunti per una discussione a più voci*
- P. Mattana, A. Paoloni, M. Grimaldi *Variazione dei parametri neurolinguistici nell'unspoken speech*
- A. De Dominicis *Variazione nelle categorie intonative*

SESSIONE A TEMA

- F. Orletti *Multimodalità e punti di transizione nei turni di conversazione*
- P. Sorianoello *Il tipo esclamativo. Analisi e percezione delle risorse prosodiche*
- E. Zovato, P. L. Salza, K. Bao Nguyen, M. Danieli *Caratterizzazione di due stili di eloquio emozionale in base all'enfasi e ai confini prosodici percepiti*
- I. Alfano, R. Savy *Le richieste di informazione nel parlato dialogico: un'analisi prosodica in contesti pragmatici orientati in lingua italiana e spagnola*
- C. Ericsson, G. Abete *Aspetti fonetici e interazionali di un marcatore discorsivo di area napoletana*
- G. Mignini *Mixed initiative spoken dialogue system in the sales force domain*

MULTIMODALITÀ E PUNTI DI TRANSIZIONE NEI TURNI DI CONVERSAZIONE

Franca Orletti

Università degli Studi di Roma Tre

orletti@uniroma3.it

La relazione affronta un tema centrale della ricerca nell'ambito dell'analisi conversazionale quella della cosiddetta "projectability", concetto che potremmo tradurre in italiano con la possibilità di anticipare gli sviluppi futuri dell'interazione con particolare riguardo per la struttura dei turni. La rilevanza, nella gestione dei turni, della capacità degli interagenti di anticipare l'andamento che prenderà la conversazione, sia nel senso specifico di come e quando verrà completato il turno, sia nel senso più ampio di prevedere le azioni future, era stata sottolineata da Sacks, Schegloff, Jefferson (1974) nel loro classico lavoro sul turno pubblicato in *Language*. Troviamo lì sottolineato sia il carattere interazionale della gestione dei turni, che la proprietà di essere definita localmente, momento per momento. Tutto questo è reso possibile dal fatto che il parlante può parlare in modo tale da permettere ai futuri parlanti di anticipare gli sviluppi future del turno:

"The turn-unit is of a sort which a) employs a specification of minimal sizes, but b) provides for expansion within a unit, c) is stoppable (though not at any point), and d) has transition places discretely recurring with it, e) which can themselves be expanded or contracted; all of these features except the first are loci of interactional determination. By virtue of this character, it is misconceived to treat turns as units characterized by a division of labor in which the speaker determines the unit and its boundaries, with other parties having as their task the recognition of them. Rather, the turn is a unit whose constitution and boundaries involve such a distribution of tasks as we have noted: *that a speaker can talk in such a way as to permit projection of possible completion to be made from his talk, from its start, allowing others to use its transition places to start talk, to pass up talk, to affect directions of talk, etc.; and that their starting to talk, if properly placed, can determine where he ought to stop talk. That is, the turn as a unit is interactively determined.* (Sacks et al., 1974: 726–7) .

Da questa enunciazione iniziale del ruolo della "projectability" nella gestione fluida dei turni, sia nell'alternanza del diritto di parola, che nella individuazione delle unità costitutive dei turni e dei loro confini il tema è stato ulteriormente approfondito nei lavori di linguistica interazionale nati dall'intersezione dell'interesse di studiosi di formazione linguistica, prevalentemente funzionale, e analisti della conversazione. Inizialmente, come in Sacks et al. (1974), le risorse linguistiche alla base della prevedibilità della struttura dei turni erano fondamentalmente sintattiche, successivamente si è riconosciuto il ruolo dei fattori fonetici, prosodici e di natura pragmatica. Ricordiamo a questo proposito i saggi di Auer, (1996); Couper-Kuhlen, (1996, 2001, 2004). Per quanto riguarda gli aspetti fonetici Local (1983), Local et al. (1986) and Local and Kelly (1986).

Lavori come quello di Ford and Thompson (1996; Ford et al., 1996) hanno mostrato il ruolo dell'interazione dei vari fattori, sintassi, prosodia, pragmatica nell'individuazione dei punti di transizione nei turni. A questi sono stati aggiunti aspetti non verbali.

L'attenzione per il ruolo delle risorse multimodali, come lo sguardo, la gestualità, le posture del corpo è presente in gran parte della letteratura più recente sulla

“projectability”(Schegloff,1984, Streeck and Hartge, 1992; Ford et al., 1996; Hayashi et al., 2002)Goodwin, 1981; Lerner, 2003; Rossano, 2005) (Goodwin, 2000).

2. La ricerca

Nell’ambito del quadro teorico di cui abbiamo dato appena qualche cenno, il lavoro che si intende presentare analizza il ruolo dei fattori fonetici, prosodici e multimodali nella gestione dei turni in un corpus di interazioni videoregistrate in lingua italiana. Le interazioni sono tratte da lezioni in una classe elementare con bambini con difficoltà di apprendimento.

Lavori precedenti sullo stesso corpus (Orletti 2006, 2007) hanno mostrato come le insegnanti utilizzino tutte le risorse linguistiche per favorire la partecipazione dei bambini, creando delle intelaiature linguistiche in cui il bambino deve solo collocare una parola o anche solo una parte. I turni diventano veri e propri enunciati a più voci.

Nel presente lavoro l’attenzione verterà sui segnali fonetici, paralinguistici e non verbali che favoriscono la costruzione del turno e le transizioni da parlante a parlante in situazioni in cui il controllo delle risorse linguistiche come la sintassi e il lessico sia limitato.

Si ha così l’intento duplice di affrontare la questione della prevedibilità in una lingua come l’italiano tipologicamente diversa da quelle centrali nella ricerca in linguistica interazionale, dall’altra trovare la conferma della possibilità di interagire fluidamente e di regolare sia l’alternanza dei turni che la strutturazione dei turni in unità costituenti nei casi di risorse linguistiche limitate (Goodwin, 2000; Streeck 2008).

IL TIPO ESCLAMATIVO. ANALISI E PERCEZIONE DELLE RISORSE PROSODICHE

Patrizia Sorianoello
Università degli Studi di Bari
p.sorianoello@lettere.uniba.it

L'esclamazione è una tipologia modale, talora designata secondaria o minore, intrisa da un'ineliminabile componente espressiva ed emotiva. E' una risorsa comunicativa tipica della lingua orale la cui forza illocutiva consiste nel presentare come inaspettato o sorprendente il suo contenuto proposizionale.

Gli aspetti su cui si è concentrata fino ad oggi l'attenzione degli studiosi sono quasi esclusivamente la sintassi e la semantica (cfr. Benincà 1995; Radford 1997; Zanuttini & Portner 2000, 2003; Munaro 2006); scarso è stato invece l'interesse riservato a questo tipo modale relativamente alle sue componenti fonetiche.

Il presente lavoro intende approfondire quest'ultima tematica prendendo in esame le caratteristiche prosodiche di frasi esclamative realizzate da locutori di italiano meridionale. Il corpus si compone di enunciati, privi di lessico emotivo e di operatori focalizzanti, estratti da sceneggiature appositamente elaborate. Il campione comprende due serie di frasi identiche per composizione segmentale e organizzazione sintattica, ma diverse quanto a forza illocutiva: la frase esclamativa è stata posta infatti a raffronto da un lato con la sua controparte assertiva (ad es. il treno è partito vs. il treno è partito!) e dall'altro con l'interrogativa *wh-* (quanti pesci hai pescato? vs. quanti pesci hai pescato!). L'ipotesi di partenza è che, a parità di struttura proposizionale, la riuscita degli atti linguistici considerati avvenga attraverso l'attivazione di precipue risorse prosodiche.

A tal fine, più parametri acustici sono stati impiegati, tra cui la durata totale dell'enunciato, la struttura intonativa (andamento, escursione tonale, distribuzione degli accenti e dei toni di confine), la durata vocalica, l'intensità.

Nell'espressione esclamativa, l'intensità, l' f_0 media e il pitch range totale sono sempre maggiori rispetto a quanto rinvenuto nelle modalità di confronto. Peculiare è l'andamento melodico, alto e sostenuto, di norma presente sulla parte iniziale del contorno. Le esclamative sono inoltre sistematicamente più lunghe delle interrogative, ma non sempre delle assertive.

L'interpretazione dei dati sperimentali è stata integrata con i risultati ottenuti dall'esecuzione di un test di identificazione percettiva. Per un sottocampione di frasi è stata creata, tramite la manipolazione separata e congiunta di f_0 e durata, una serie di stimoli. Gli stimoli verbali ottenuti sono di sei tipi:

- a) frase esclamativa *non wh-* con intonazione assertiva;
- b) frase esclamativa *wh-* con intonazione interrogativa *wh-*;
- c) frase esclamativa *non wh-* con *pattern* temporale assertivo;
- d) frase esclamativa *wh-* con *pattern* temporale interrogativo *wh-*;
- e) frase esclamativa *non wh-* con intonazione e *pattern* temporale assertivi;
- f) frase esclamativa *wh-* con intonazione e *pattern* temporale interrogativi *wh-*.

La valutazione degli stimoli è stata effettuata da un gruppo di uditori ingenui. Durante la somministrazione del test, ai soggetti è stato chiesto di riconoscere il valore modale di ogni

enunciato. I giudizi ottenuti permettono di definire il ruolo svolto dai due parametri nella decodifica uditiva di un'esclamazione.

Nell'individuazione della categoria pragmatica delle frasi, l'azione degli indici acustici manipolati è sinergica, non a caso i punteggi più alti si ottengono quando si interviene su entrambi, ma il loro impatto percettivo non è paritario. La durata, ad esempio, non sembra di per sé bastare a modificare il giudizio espresso sulla categoria modale dell'enunciato: la frase esclamativa continua ad essere valutata tale dagli uditori anche quando il pattern temporale riprodotto è quello dell'assertiva o dell'interrogativa *wh*-equivalente. La durata delle parole (in primis della *V* tonica) su cui verte l'esclamazione è invece apprezzabile. La riduzione temporale di queste sillabe chiave, sebbene non produca un cambio di forza illocutiva, riduce tuttavia la carica enfatica dell'esclamazione, che di conseguenza appare meno 'convincente'. In ragione della natura scalare della frase esclamativa, la durata assume una chiara funzione pragmatica, visto che l'incremento temporale delle sillabe è direttamente proporzionale al grado di sorpresa insito nell'enunciato.

Al contrario, l'andamento intonativo si configura quale valido discrimen nella separazione delle categorie considerate; è sufficiente modificare anche soltanto alcuni tratti del contorno esclamativo per deviare, confondendolo, il giudizio emesso dagli uditori.

La ricerca condotta ha dunque messo in evidenza in che modo la covariazione degli indici acustici possa contribuire al riconoscimento dell'atto modale esclamativo, consentendo altresì di individuare alcune soglie di mutamento percettivo utili a definire i caratteri peculiari delle frasi esclamative 'prototipiche'.

Riferimenti bibliografici

Benincà P. (1995), 'Tipi di frasi – Il tipo esclamativo', in Renzi L., Salvi G. & A. Cardinaletti (acd), *Grande Grammatica Italiana di Consultazione*, vol. III, Il Mulino, Bologna, pp. 127-152.

Benincà P. (1996), 'La struttura della frase esclamativa alla luce del dialetto padovano', in Benincà P., Cinque G., De Mauro T. & N. Vincent (acd), *Italiano e dialetti nel tempo. Saggi di grammatica per Giulio Lepschy*, Bulzoni, Roma, pp. 23-43.

Munaro N. (2006), 'Verbless exclamatives across Romance: standard expectations and tentative evaluations', *Working Paper in Linguistics* 16, pp. 185-209.

Munaro N. & P. Benincà (2002), 'Sintassi e lessico delle esclamative: caratteristiche romanze e indoeuropee', in Marcato G. (acd) *La dialettologia oltre il 2001*, Atti del Convegno di Sappada, Padova, Unipress, pp. 199-202.

Radford A. (1982), 'The syntax of verbal *Wh*-exclamatives in Italian', in Vincent N. & M. Harris (acd), *Studies in the Romance Verb*, Croom Helm, London

Radford A. (1997), 'Per un'analisi delle frasi esclamative in italiano' in Renzi L. & M. Cortelazzo (acd), *La linguistica italiana fuori d'Italia*, Bulzoni, Roma, pp. 93-123.

Zanuttini R. & P. Portner (2000), 'The characterization of exclamative clauses in Paduan', *Language* 76, pp. 123-132.

Zanuttini R. & P. Portner (2003), 'Exclamative clauses: at the syntax-semantics interface', *Language* 79, pp. 1-48.

CARATTERIZZAZIONE DI DUE STILI DI ELOQUIO EMOZIONALE IN BASE ALL'ENFASI E AI CONFINI PROSODICI PERCEPITI

Enrico Zovato¹, Pier Luigi Salza¹, Kim Bao Nguyen², Morena Danieli¹

¹Loquendo S.p.A., ²CELI s.r.l., Torino

Enrico.zovato@loquendo.com, pierluigi.salza@loquendo.com, morena.danieli@loquendo.com

Introduzione

La riproduzione dell'espressività tipica degli stili di eloquio emotivo può migliorare significativamente la contestualizzazione delle risposte di un'interfaccia persona-macchina, facendone un assistente virtuale capace di conversare con l'utente in modo affettivo. A questo scopo, il presente lavoro illustra la caratterizzazione di due stili di parlato emozionale, rispettivamente triste e allegro, sulla base dei fenomeni percepiti come enfasi e confini prosodici. In particolare si sono analizzati: per l'enfasi, i valori medi dei parametri acustici di F0, durata e intensità di ciascuna unità fonetica della parola in esame; per i confini prosodici, la distinzione tra confine forte (presenza di pausa) e confine debole (assenza di pausa, discontinuità di tipo ritmico e/o intonativo). Infine, si è analizzata l'eventuale correlazione tra l'occorrenza dell'enfasi e dei confini prosodici e la categoria grammaticale della parola (Part of Speech).

Corpus di dati vocali e modalità dell'esperimento

I due stili espressivi sono stati riprodotti mediante registrazioni effettuate in laboratorio da parte di 5 soggetti madrelingua parlanti inglese americano. I testi, tratti da romanzi e racconti, sono stati selezionati da un madrelingua in modo che i contenuti semantici fossero in grado di indurre l'appropriato atteggiamento emotivo relativamente ai due stili considerati. I due corpus testuali avevano all'incirca la stessa lunghezza: quello di contenuto allegro era composto da 350 parole, quello di contenuto triste da 388 parole, per un totale di 738 parole. Per comodità di elaborazione degli enunciati, i testi sono stati suddivisi in frasi, risultate ovviamente di lunghezza diversa. Si sono ottenute 42 frasi: solo 14 di contenuto allegro, risultate mediamente lunghe; ben 28 di contenuto triste, risultate mediamente assai più brevi. Ai parlanti (tre femminili e due maschili) si è chiesto di leggere i brani in modo naturale e conforme ai contenuti. Si è poi chiesto loro di leggere gli stessi brani, in sessioni di registrazione separate, adottando uno stile di lettura neutro indipendentemente dai contenuti semantici dei testi, in modo da avere un corpus di riferimento con stile neutro. Il corpus totale ammonta perciò in origine a $(42+28) \times 5 = 420$ enunciati.

Il corpus di enunciati è stato quindi sottoposto ad una valutazione percettiva da parte di numerosi soggetti madrelingua, tale che ciascun enunciato è stato valutato da 3 soggetti. La valutazione, tesa a verificare l'effettiva resa emozionale degli enunciati, è stata condotta utilizzando una scala MOS a 7 punti (1="very sad", 2="sad", 3="slightly sad", 4="neutral", 5="slightly happy", 6="happy", 7="very happy"). Includendo soltanto gli enunciati la cui valutazione è risultata coerente con il corpus originale di appartenenza (tristi gli enunciati con punteggio medio < 3; neutri gli enunciati con punteggio medio > 3,5 e < 4,5; allegri

quelli con punteggio medio > 5), si è ottenuto il seguente corpus di enunciati validi su cui lavorare: 197 effettivamente espressivi (62 allegri e 135 tristi) e 170 effettivamente neutri. Conseguenza di tale selezione è che ora i dati non sono più omogeneamente distribuiti tra i diversi parlanti.

Annotazione, ulteriore selezione ed elaborazione dei dati

I testi degli enunciati validi sono stati quindi annotati manualmente e separatamente da due operatori i quali, sulla base dell'ascolto, hanno etichettato ciascuna parola percepita come portatrice di enfasi e ciascun confine di parola percepito come confine prosodico. Al fine di stimare il grado complessivo medio di concordanza percettiva, si è misurata la congruenza tra le annotazioni dei due operatori, per ciascun enunciato e separatamente per ciascun parametro (enfasi e confine prosodico), tramite il "K test", usato per misurare il livello di concordanza tra le classificazioni di due valutatori rispetto alla concordanza "per caso". Nel caso dei confini prosodici il valore medio di K è risultato superiore a 0,75 indicando quindi un buon livello di coerenza nelle annotazioni, mentre per quanto riguarda le enfasi, il K medio risultante è inferiore a 0,6, indice del margine di soggettività e della relativa difficoltà insita nel compito affidato agli annotatori di identificare percettivamente le parole portatrici di enfasi.

Si è quindi proceduto a selezionare ulteriormente il materiale su cui svolgere le successive analisi, ottenendo i seguenti sottoinsiemi: a) tutte le singole parole che sono state percepite come portatrici di enfasi da entrambi gli annotatori, suddivise per stile di eloquio; b) tutte le coppie di parole il cui confine è stato percepito come confine prosodico da entrambi gli annotatori, suddivise per stile di eloquio. Uno degli annotatori ha poi effettuato le seguenti ulteriori etichettature "oggettive": su entrambi i sottoinsiemi, ha etichettato ciascuna parola in base alla categoria grammaticale (Part of Speech); sul sottoinsieme b) (confini prosodici), ha etichettato ciascun confine in base alla presenza o meno di pausa, valutata strumentalmente. Le etichette grammaticali utilizzate sono: Nome, Pronome, Verbo, Avverbio, Aggettivo, Preposizione, Congiunzione.

Attraverso l'uso di trascrittori automatici, basati su algoritmi ad apprendimento, si è ottenuta la trascrizione fonetica del corpus vocale selezionato; quindi i campioni di voce sono stati segmentati in unità fonetiche per mezzo di un algoritmo di allineamento forzato basato sui modelli di Markov (HMM); etichettatura e allineamento sono stati successivamente supervisionati da un esperto fonetista; infine, per ciascuna unità fonetica sono stati calcolati automaticamente i seguenti parametri: F0 (min, max, medio), RMS energia (max, media), mentre le durate sono state ricavate direttamente dall'allineamento fonetico.

Primi risultati e analisi statistica dei dati

Con riferimento al corpus di enunciati validi e separatamente per ciascuno dei due annotatori, si sono definiti gli indici di frequenza ottenuti dal rapporto tra il numero di eventi annotati (enfasi/confini) e il numero totale di parole, per ciascuna frase. I parametri così calcolati sono stati aggregati per stile di eloquio e successivamente è stata eseguita

l'analisi di varianza sugli indici di frequenza. L'uso alternativo di tecniche di classificazione basate su algoritmi di *machine learning* tipo SVM (Support Vector Machine) risulterebbe impraticabile, per la quantità limitata dei dati a disposizione. Riguardo al fenomeno dell'enfasi, l'analisi di varianza ha evidenziato per entrambi gli annotatori delle differenze statisticamente significative per quanto riguarda gli stili emozionali rispetto allo stile neutro, come è lecito attendersi. Tuttavia è emersa anche una differenza tra lo stile allegro e quello triste, con una maggior frequenza del fenomeno dell'enfasi rilevata nel primo. Per quanto riguarda i confini prosodici, l'analisi della varianza ha evidenziato differenze meno significative, che saranno comunque oggetto di ulteriore indagine.

Ringraziamenti

Questa ricerca è stata parzialmente finanziata dalla Commissione Europea nell'ambito del progetto europeo FP6 IST34434 "Companions". Gli autori sono riconoscenti a Simon Parr.

Riferimenti bibliografici

- Avesani C., Vayra M., & Zmarich C. (2007). On the articulatory bases of prominence in Italian. In J. Trouvain e W. J. Barry, *Proceedings of the XVth International Congress of Phonetic Science* (pp. 981-984). Saarbrücken: Dudweiler, Pirrot GmbH..
- Bevacqua, E., Mancini, M., Niewiadomski, R., & Pelachaud, C. (2007), An expressive eca showing complex emotions. AISB'07 - Artificial and Ambient Intelligence. Newcastle University, Newcastle upon Tyne, UK
- Brugnara, F., Falavigna, D., & Omologo, M. (1993), Automatic Segmentation and Labelling of Speech based on Hidden Markov Models, *Speech Communication*, 12, 357-370
- Campbell, N. (2005), Getting to the heart of the matter: Speech as the expression of affect; rather than just text or language, *Language Resources and Evaluation*, 39, 109-118
- Carletta, J. (1996), Assessing agreement on classification tasks: the kappa statistics, *Computational Linguistics*, 22, 249-254
- Davidson, R.J, Scherer, K.R., and Hill Goldsmith, H. (Eds) (2009), *Handbook of Affective Sciences*, Oxford Univ. Press
- Ekman, P. (1993), An argument for basic emotions, *Cognition and Emotion*, 6, 169-200
- <http://www.companions-project.org>
- Johnstone, T. & Scherer, K.R. (1999), The effects of emotions on voice quality. In *Proceedings of XIV Int. Congress on Phonetic Science*, San Francisco, 2029–2032
- Landis, J.R. & Koch, G.G. (1977), The measurement of observer agreement for categorical data, *Biometrics*, 33, 159-174
- Magno Caldognetto, E. (2002), I correlati fonetici delle emozioni, in *Passioni, emozioni, affetti*. (C. Bazzanella & P. Kobau), McGraw-Hill, 197-213

Poggi, I. & Magno Caldognetto, E. (2004), Il parlato emotivo. Aspetti cognitivi, linguistici e fonetici. In F. Albano Leoni, F. Cutugno, M. Pettorino & R. Savy. (Eds.), Atti del Convegno "Italiano parlato", Napoli: D'Auria Editore, CD-Rom

Russell, J.A. (1980), A circumflex model of affect. In *Journal of personality and social psychology*, 1161 – 1178

Scherer, K.R. (2003), Vocal communication of emotion: A review of research paradigms, *Speech Communication*, 40, 227-256

Salza, P. L., Zovato, E., Danieli, M. (2009), Enfasi e confini prosodici in due stili di eloquio emozionale, Atti del Convegno AISV 2009, Zurigo, CD-Rom

Schröder, M. (2001), Emotional Speech Synthesis: A Review, In *Proceedings of EUROSPEECH 2001*, Aalborg , 561 – 564

LE RICHIESTE DI INFORMAZIONE NEL PARLATO DIALOGICO: UN'ANALISI PROSODICA IN CONTESTI PRAGMATICI ORIENTATI IN LINGUA ITALIANA E SPAGNOLA

Iolanda Alfano¹, Renata Savy²

¹Universitat Autònoma de Barcelona, ²Università degli Studi di Salerno

iolandaalfano@libero.it, rsavy@unisa.it

Il lavoro che qui presentiamo offre una descrizione dei primi risultati di un'analisi pragmatica e prosodica delle diverse strategie di richiesta di informazione e di conferma nel parlato dialogico in lingua spagnola ed italiana. Ad una descrizione dei tipi di richiesta impiegati e alla successiva analisi della loro distribuzione, seguirà uno studio della rispettiva resa prosodica, nonché un'analisi comparativa tra i due sistemi linguistici relativa al rapporto tra i due piani di analisi.

Nel quadro della teoria degli atti linguistici, una *request* rappresenta genericamente “un atto illocutivo mediante il quale un parlante comunica al suo interlocutore che vuole che faccia o dica qualcosa per lui” (Trosborg, 1995: 187); può elicitarne pertanto un contributo azionario (*non-verbal goods*) o comunicativo (*verbal goods*) da parte dell'interlocutore (Edmondson & House, 1981). Le richieste di informazione, di cui ci occupiamo in questo lavoro, si configurano tra gli atti linguistici del secondo tipo.

Per definire le strategie impiegate per le richieste di informazione sul piano pragmatico, vengono di solito presi in considerazione molti aspetti, tra cui, ad esempio il grado di *directness*, *politeness* e forza illocutoria. Questi fattori consentono di definire tipi diversi di richiesta, ma anche di mettere in relazione le scelte strategiche con gli stili comunicativi e le esigenze interazionali.

Le strategie messe in atto dai parlanti dipendono, ovviamente, da una molteplicità di fattori concomitanti e spesso difficili da sviscerare, tra cui tutti gli elementi situazionali, i fattori sociali, quali ad esempio lo status sociale dei parlanti, il tipo di relazione ed il relativo grado di familiarità che li lega (Cenoz e Valencia, 1994), nonché l'*habitus* culturale (Blum-Kulka e Olshtain, 1984; Blum-Kulka *et al.*, 1989; García, 1989; Izaki, 2000; Savy Solís García, 2008) congiuntamente alle scelte stilistiche individuali di ciascun parlante.

La nostra indagine ha come oggetto un piccolo *corpus* pilota di dialoghi *task-oriented*, elicitati con la nota tecnica del “Test delle differenze” (Pean *et al.*, 1993). Nello schema per l'annotazione pragmatica impiegato, Pra.Ti.D (De Leo e Savy, 2007), le richieste di informazione sono etichettate con sei tipi di *moves* (Carletta *et al.*, 1996): domande generiche che possono essere formulate in vario modo (*InfoRequest*), domande introdotte da un pronome interrogativo (*Query_w*), domande che prevedono una risposta sì/no (*Query_y*), vere e proprie richieste di conferma, (*Check*), richieste di chiarimento (*Hold*), verifiche di attenzione e disponibilità dell'interlocutore (*Align*).

Dal punto di vista interazionale, tali mosse delineano un *continuum* tra due meccanismi di interazione principali: 1) quello del cosiddetto *Questioning*, in cui uno o entrambi i partecipanti collaborano con domande e risposte centrate intorno agli elementi/oggetti della vignetta; 2) quello del cosiddetto *Check*, in cui il parlante assume la descrizione della propria vignetta come punto di vista centrale, mettendo in pratica strategie di controllo basate su richieste di conferma.

Dal punto di vista testuale, il primo si realizza in una strutturazione del dialogo piuttosto equilibrata (in termini di durata dei turni degli interlocutori e frequenza dell'avvicendamento), mentre il secondo dà luogo ad una struttura altamente asimmetrica e frammentata, basata su lunghe descrizioni (spiegazioni, asserzioni dichiarative) e domande polari o richieste di conferma, che limitano il campo di azione verbale dell'interlocutore.

Sul versante comunicativo, le due dinamiche sono rappresentative di stili diversi, il primo implicante un maggior coinvolgimento dell'interlocutore e massima cooperazione nello scambio verbale, il secondo improntato ad una gestione, per così dire, più "egocentrica", in cui ciascuno dei due parlanti tende a far prevalere il proprio punto di vista o quantomeno gestisce l'interazione guardando dalla propria prospettiva.

Si vedono quindi disposte sul primo polo le mosse che richiedono una collaborazione maggiore all'interlocutore cui sono indirizzate (richieste generiche di informazione - *InfoRequest* - seguite da domande aperte - *Query_w*)¹ mentre sull'altro quelle maggiormente indirizzate alla verifica ed alla conferma di un'informazione presupposta o già acquisita (attraverso domande polari - *Query_y* - e *Check*)². Le richieste di chiarimento (*Hold*) tendono a situarsi verso il centro del *continuum*, in quanto esprimono dinamicità e scambio equilibrato di informazioni; infine, le mosse di verifica di attenzione e disponibilità dell'interlocutore (*Align*) oscillano tra un polo e l'altro a seconda che focalizzino su un referente esterno o sull'interlocutore stesso.

L'analisi qualitativa e quantitativa del *corpus* contrastivo rivela la presenza di strategie diverse adottate dai parlanti per elicitarne un contributo comunicativo che, pur essendo legate a vari fattori concomitanti, mostrano tendenze diverse nelle due lingue prese in esame. Globalmente tutti i tipi di *request* appaiono ben rappresentati, anche se con percentuali di occorrenza diverse. Le mosse privilegiate nei dialoghi italiani (*Query_y* e *Check*) sembrano partire dal punto di vista del parlante che in quel momento tiene il turno, delineando una modalità di interazione di tipo *Check* basata su strategie di controllo attraverso richieste di conferma; diversamente, nei dialoghi spagnoli prevalgono vere e proprie richieste di informazione su elementi della vignetta (*InfoRequest* e *Query_w*) che tendono a coinvolgere maggiormente l'interlocutore ed a mettere in atto strategie di forte collaborazione, tendenti, dunque, alla modalità di tipo *Questioning*.

Oltre ad un'analisi pragmatica e testuale, per capire a fondo le scelte dei parlanti e definirne, laddove è possibile, le modalità, è imprescindibile uno studio sul piano prosodico, in questo momento in via di sviluppo. L'analisi di questo livello mira innanzitutto a verificare se e in quale misura le diverse strategie di richiesta di informazione si differenziano anche in termini di resa intonativa; in secondo luogo se la eventuale corrispondenza tra funzione pragmatica e profilo melodico manifesta delle costanti confrontabili nelle due lingue in esame.

L'analisi e annotazione prosodica del *corpus* prevede tre livelli: 1) quello del *phrasing* in unità intonative maggiori; 2) la descrizione fonetica della curva di f_0 (attraverso la notazione INTSINT, cfr. Hirst *et al.*, 2000), 3) la determinazione dei gradi di prominenza ritmica e melodica (primo passo verso la definizione di categorie fonologiche eventualmente associabili alle diverse funzioni).

L'analisi, del tutto preliminare, non permette allo stato attuale di trarre conclusioni definitive, soprattutto per quanto riguarda il confronto interlinguistico. Si rivela, tuttavia, significativamente la presenza costante di diverse strategie di focalizzazione prosodica (che

¹ ad esempio: "La coda del cane?" e "Com'è la coda del cane?"

² ad esempio: "La coda del cane è dritta?" e "La coda del cane è dritta, vero?"

passano sia attraverso il livello di *phrasing* sia attraverso le realizzazioni accentuali) in rapporto alle diverse funzioni pragmatiche. Gli stili comunicativi sembrano manifestarsi, quindi, anche sul piano fonico nella prevalenza sistematica di diverse configurazioni prosodiche.

Riferimenti bibliografici

Blum-Kulka, S. - Olshtain, E. (1984). Requests and apologies: A cross-cultural study of speech act realization patterns (CCSARP). *Applied Linguistics*, 5, pp. 196-213.

Blum-Kulka, S. - House J. - Kasper, G. (1989). *Cross Cultural Pragmatics: Request and Apologies*, Norwood, N.J: Ablex.

Carletta, J. - Isard, A. - Isard, S. - Kowtko, J. - Doherty-Sneddon, G. - Anderson, A., (1996) HCRC Dialogue Structure Coding Manual, Technical Report, 82. Human Communication Research Center, University of Edinburgh.

Cenoz, J. - Valencia, J. (1994). Interlanguage pragmatics: The role of linguistic and social psychological elements in the production of English requests and apologies. Unpublished manuscript. Vitoria-Gasteiz, Spain: Dept. of English and German Philology, U. of the Basque Country.

De Leo, S. e Savy, R. (2007) PraTiD_Normario per annotazione pragmatica, in www.parlaritaliano.it (area STRUMENTI, sezione Pra.Ti.D, file PraTiD_Normario per annotazione pragmatica.pdf)

Edmondson, W., House, J. (1981) *Let's talk and talk about it*. München, Wien und Baltimore: Urban und Schwarzenberg.

García, C. (1989). Disagreeing and requesting by Americans and Venezuelans. *Linguistics and Education*, 1, pp. 299-322.

Hirst, D., Di Cristo, A. Espesser, R. (2000). Levels of representation and levels of analysis for the description of intonation systems. in M. Horne (ed) *Prosody: Theory and Experiment*. Studies presented to Gösta Bruce, Kluwer, Dordrecht, pp. 51-88.

Izaki, Y. (2000). Cultural differences of preference and deviations from expectations in requesting: A study of Japanese and French learners of Japanese in contact situations. *Journal of Japanese Language Teaching*, 104, pp. 79-88.

Pean V., Williams S., Eskenazy M. (1993), The design and recording of icy, a corpus for the study of intraspeaker variability and the characterisation of speaking styles. In *Proceedings Of Eurospeech 1993*, Berlin, pp. 627-630.

Savy, R. - Solís García, I. (2008) Strategie pragmatiche in italiano e spagnolo a confronto: una prima analisi su corpus, in Voghera M. (a cura di) *Testi e linguaggi*, Roma: Carocci, pp. 214-242.

Trosborg, A. (1995). *Interlanguage Pragmatics: Requests, Complaints and Apologies*. Berlin, New York: Mouton Gruyter

ASPETTI FONETICI E INTERAZIONALI DI UN MARCATORE DISCORSIVO DI AREA NAPOLETANA

Christine Ericsson¹, Giovanni Abete²

¹ Stockholm University, ² Università di Napoli “Federico II”
ericsson@ling.su.se, giovanni.abete@libero.it

Si propone qui un’analisi fonetica e conversazionale dell’espressione *hai capito?* (d’ora in avanti HC) in un corpus di parlato dialettale raccolto a Pozzuoli (NA). Tale espressione viene utilizzata frequentemente come appendice di fine enunciato e presenta realizzazioni molto variabili, raggiungendo spesso gradi particolarmente avanzati di riduzione. Parallelamente alla riduzione fonetica, HC sembra essere soggetto a un processo di delessicalizzazione piuttosto avanzato e assume funzioni essenzialmente pragmatiche. Tali aspetti verranno indagati in questo contributo attraverso l’analisi dettagliata delle realizzazioni fonetiche di HC e lo studio dei suoi contesti d’uso. Particolare attenzione verrà riservata alla posizione sequenziale di questa espressione all’interno della struttura conversazionale, mettendo in evidenza la relazione che intercorre tra HC e i brevi *feedback* prodotti dall’interlocutore (del tipo *mh*, *eh* etc.). In questo modo, lo studio della funzione di HC verrà inquadrato all’interno di una visione della conversazione come cooperazione, secondo la quale i parlanti interagiscono nella costruzione dei significati e utilizzano i dettagli fonetici come una risorsa per la gestione dei turni e il corretto svolgimento dell’interazione verbale (cfr. ad es. Local et al. 1986).

Come accennato, questa ricerca è stata svolta su un corpus di registrazioni di parlato dialettale, raccolte nella comunità di Pozzuoli. Il corpus è parte di un progetto di dottorato e si inserisce in un piano di ricerca più ampio di documentazione e analisi dei dialetti della Campania (cfr. Sornicola 2002). Esso consiste di circa 6 ore di registrazioni di otto parlanti maschi, di età compresa tra i 26 e i 67 anni. Gli informatori sono tutti pescatori e presentano un grado di istruzione piuttosto basso. Le registrazioni sono state elicitate tramite intervista libera (cfr. Como 2006). Le sessioni si sono svolte all’aperto, sulle imbarcazioni dei pescatori, in una piccola baia riparata dal vento. Si tratta di un contesto conversazionale naturale: qui i pescatori si intrattengono, dopo una notte di pesca, a rammendare le reti e a chiacchierare. Durante l’intervista il raccoglitore ha utilizzato una varietà dialettale vicina a quella degli intervistati. La conversazione è stata incentrata principalmente su argomenti legati alla pesca. Erano sempre presenti uno o più amici della persona intervistata, condizione che ha favorito la spontaneità delle interazioni, attenuando l’imbarazzo prodotto dalla presenza dell’intervistatore (cfr. Milroy 1987: 62). Il parlato elicitato risulta pertanto molto spontaneo e fortemente marcato in senso dialettale. Ad ogni modo, nonostante tutti gli accorgimenti presi per instaurare un contesto conversazionale rilassato, si è consapevoli delle peculiarità della situazione comunicativa imposta dalla tecnica di intervista (cfr. Wolfson 1976), e di queste peculiarità si è tenuto conto sia in fase analitica che in sede interpretativa.

Da queste registrazioni sono stati selezionati per l’analisi 30 minuti di parlato per ciascun parlante. Questo materiale è stato quindi etichettato, individuando tutte le realizzazioni di HC e trascrivendo in traduzione letterale il contesto conversazionale in cui l’espressione è inserita. Sono state etichettate e analizzate anche tutte le altre realizzazioni di *capire* presenti nel corpus, al fine di un opportuno confronto della fonetica di queste con quella dell’espressione qui in esame. Gli interventi dei vari parlanti sono stati trascritti su

rigli diversi e allineati rispetto alla loro posizione sull'asse temporale, così da evidenziare gli aspetti interazionali (cfr. Sacks et al. 1974). Inoltre, è stato annotato come procede la conversazione subito dopo HC, segnalando ad es. se l'intervistato continui a parlare, se l'intervistatore intervenga con un commento o una nuova domanda, se ci sia una sovrapposizione tra i due interlocutori, etc. Di seguito si riporta un esempio di trascrizione³:

N060309a-01: 71.50-80.19

A: tiene più assai/ tiene duecento/ trecento/(.)

I: [ǣ]

A: ¹hai capito?/ e quindi ha raddoppiato e<e>perché poi/ più non prendi niente e più<ù> metti nasse/(.) ²hai capito?/(.)

I: [ǣ]

A conclude il discorso, I non interviene.

Le realizzazioni di HC sono state trascritte foneticamente in maniera molto dettagliata. Quindi le trascrizioni sono state classificate secondo una complessa tipologia in rapporto ai fenomeni di riduzione presenti, alla tipo di resa degli eventuali segmenti consonantici, al numero di sillabe percepite.

In questa sede si è scelto di presentare i dati relativi a quattro degli otto parlanti del campione. Si tratta delle persone che si rivolgono quasi esclusivamente all'intervistatore, limitando gli scambi con gli altri presenti. Le realizzazioni di HC in questa parte del corpus sono 105, a cui vanno aggiunte le altre realizzazioni del verbo *capire* con fonetica e funzione diverse da quelle qui in esame. Questo materiale verrà discusso rispetto a due questioni principali: da un lato si metterà in evidenza come i processi di riduzione che coinvolgono HC abbiano caratteristiche speciali, che non si riscontrano in altre parti del sistema, e siano da ricondurre sostanzialmente alla particolare frequenza d'uso di questa espressione; dall'altro si avanzerà una proposta interpretativa della funzione pragmatica di HC.

Questo studio si inserisce nell'ambito di un filone di ricerca che analizza la variabilità fonetica all'interno di situazioni conversazionali naturali, indagando le relazioni tra forma fonetica e struttura dell'interazione⁴. Inoltre, i dati che si presentano sono di particolare interesse per lo studio degli effetti che la frequenza d'uso ha sulla forma fonetica e sul significato delle espressioni⁵.

³ In questa sequenza il parlante A spiega che al giorno d'oggi un pescatore possiede molte più *nasse* ('tipo di trappole per i pesci') di quante ne possedesse un tempo. La ragione è a suo avviso, come era già stato argomentato in un momento precedente, l'impoverimento del mare: meno pesci il pescatore prende e più *nasse* è costretto a utilizzare.

⁴ Si veda ad es. Couper-Kuhlen/Ford 2004, Local et al. 1986, Local 2003 e 2007.

⁵ Si veda ad es. Bybee 2003 e 2007, Pierrehumbert 2001 e 2002.

Riferimenti bibliografici

Bybee, J. 2003. Mechanisms of change in grammaticalization: the role of frequency. Janda R./Joseph, B. (a cura di), *Handbook of historical linguistics*. Oxford, Blackwell. 602-623.

Bybee, J. 2007. *Frequency of use and the organization of language*. New York, Oxford University Press.

Como, P. 2006. Elicitation techniques for spoken discourse. Brown, K. (a cura di) *Encyclopedia of language and linguistics*, seconda edizione, volume 4. Amsterdam, Elsevier: 105-109.

Couper-Kuhlen, E./Ford, C. E. 2004 (a cura di). *Sound Patterns in Interaction. Cross-linguistic studies from conversation*. Amsterdam/Philadelphia: John Benjamins.

Local, J. K./Kelly J./Wells W. H. G. 1986. "Towards a phonology of conversation: turn-taking in Tyneside English". *Journal of Linguistics*, 22: 411-437.

Local, J. 2003. "Variable domains and variable relevance: interpreting phonetic exponents". *Journal of Phonetics*, 31: 321-339.

Local, J. 2007. Phonetic detail and the organization of talk-in-interaction. *Proceedings of the International Congress of Phonetic Sciences*, XVI: 1-10.

Milroy, L. 1987. *Observing and analysing natural language*. Cambridge (UK), Blackwell.

Pierrehumbert, J. B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. Bybee, J./Hopper, P. (a cura di), *Frequency effects and emergent grammar*. Amsterdam, John Benjamins. 137-157.

Pierrehumbert, J. B. 2002. Word-specific phonetics. *Papers in Laboratory Phonology VII*. Berlin, Mouton de Gruyter. 101-139.

Sacks, H./Schegloff, E./Jefferson, G. 1974. "A simplest systematics for the organization of turn-taking for conversation". *Language*, 50: 696-735.

Sornicola, R. 2002. "La variazione dialettale nell'area costiera napoletana. Il progetto di un archivio di testi dialettali parlati". *Bollettino Linguistico Campano*, 1: 131-155.

Wolfson, N. 1976. "Speech events and natural speech: some implications for sociolinguistic methodology". *Language in Society*, 5: 189-209.

MIXED INITIATIVE SPOKEN DIALOGUE SYSTEM IN THE SALES FORCE DOMAIN

Gianluca Mignini

Università degli Studi di Napoli “Federico II”

mignini50421@hotmail.com

Il progetto realizzato consiste in un sistema di dialogo a iniziativa mista per information retrieval nel dominio della vendita di prodotti alimentari. L'applicazione vocale modella l'interazione tra un rappresentante di vendite “umano” e la società per cui lavora, a cui telefona per ricevere informazioni su clienti, prodotti ed ordini e per aprire nuove ordinazioni.

Il sistema è stato implementato utilizzando la tecnologia JSP per la creazione dinamica di pagine VoiceXML e l'interazione con una base di dati MySQL e una base di conoscenza OWL. La piattaforma VoxNauta della Loquendo ha fornito i moduli necessari per la sintesi ed il riconoscimento vocale.

I sistemi di dialogo sono stati identificati e raggruppati in tre tipologie (Jurafsky e Martin, 2006): system initiative, user initiative e mixed initiative. La maggioranza delle attuali applicazioni vocali sono di tipo system-initiative, in cui il controllo del flusso del dialogo è gestito dal sistema e all'utente è richiesto solo di fornire una sequenza di informazioni necessarie per l'esecuzione di un task. Detti sistemi sono molto robusti ed efficienti in applicazioni in cui gli utenti non conoscono le potenzialità del sistema, ma risultano poco adatti ed efficaci per utenti esperti o per scenari in cui è necessario recuperare informazioni in tempi rapidi senza dover riempire con molteplici frasi i singoli campi. Inoltre per il riconoscimento dei comandi e delle informazioni fornite dall'utente vengono spesso utilizzate grammatiche statiche, difficilmente aggiornabili e poco flessibili.

Il nostro approccio ad iniziativa mista lascia che l'utente abbia l'iniziativa del dialogo, per poi riprendere eventualmente il controllo ed aiutare l'utente ad esplicitare correttamente la richiesta. La grammatica dell'applicazione è dinamicamente creata estraendo dalla base di dati le informazioni relative all'oggetto della richiesta dell'utente, e dalla base ontologica le possibili espressioni pronunciate riconducibili a due classi/tipologie di richieste: inserimento o selezione. L'interpretazione da parte di un middle-layer della richiesta utente si avvarrà delle espressioni contenute nella base ontologica per distinguere tra un'operazione di inserimento nella base di dati (come l'inserimento di un nuovo ordine) e un'operazione di interrogazione (richiesta dei dati di un cliente, di un prodotto, o di un ordine). Un modulo delegato all'interazione con il database riceve dal middle-layer l'interpretazione della richiesta e fornisce i risultati dell'interazione.

L'utilizzo della base ontologica permette maggior flessibilità al sistema, in quanto facilmente estendibile utilizzando strumenti che non richiedono alcuna particolare conoscenza di linguaggi di programmazione, oltre che la possibilità di riuso in altre applicazioni. La dinamicità della grammatica principale garantisce inoltre la coerenza dei dati senza la necessità di alcun intervento manuale a seguito di un aggiornamento dei record del database.

Riferimenti bibliografici

Daniel Jurafsky & James H. Martin (2006), An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.

SESSIONE TECNOLOGICA

P. Cosi, M. Nicolao *EVALITA-ISTC comparison of open source tools on clean and noisy digits recognition tasks*

G. Costantini, A. Paoloni, M. Todisco *Misura oggettiva dell'intelligibilità del parlato in applicazioni forensi*

A. De Rosa, F. Cutugno *Nuove funzionalità per il portale CLIPS*

S. Romano *SpLaSH: il sistema general purpose per la gestione di annotazioni time-aligned e text-aligned*

A. Origlia, V. Galatà *Evaluating speech rate by referring on an intensity-based syllable nuclei search algorithm*

EVALITA-ISTC COMPARISON OF OPEN SOURCE TOOLS ON CLEAN AND NOISY DIGITS RECOGNITION TASKS

Piero Cosi, Mauro Nicolao

Istituto di Scienze e Tecnologie della Cognizione - “Fonetica e Dialettologia” C NR, Padova
piero.cosi@pd.istc.cnr.it, nicolao@pd.istc.cnr.it

EVALITA is a recent initiative devoted to the evaluation of Natural Language and Speech Processing tools for Italian. The general objective of EVALITA is to promote the development of language and speech technologies for the Italian language, providing a shared framework where different systems and approaches can be evaluated in a consistent manner. In this work the results of three open source ASR toolkits (*CSLU Speech Tools*, *CSLR SONIC*, *SPHINX*) working on the EVALITA clean and noisy digits recognition task will be described together with the complete evaluation methodology.

EVALITA⁶ provides a shared framework for the evaluation of different systems and approaches on separate natural language and speech processing tasks for the Italian language. In EVALITA 2009 the following tasks have been evaluated.

- Text tasks: PoS-Tagging, Parsing, Lexical Substitution, Entity Recognition, Textual Entailment
- Speech tasks: Connected Digits Recognition (clean and noisy), Dialogue System Evaluation and Speaker Identity Verification (Application & Forensic)

Regarding the Connected Digits Recognition task, systems are required to recognize sequences of spoken Italian digits (numbers ranging from 0 to 9). Two subtasks are defined, and applicants may choose to participate in any of them:

- Clean digits: in this subtask, the test digits sequences are acquired in clean environment;
- Noisy digits: in this subtask, the test digits sequences are acquired in noisy environment. The type of noise may vary from white noise to traffic, room, etc.

The Corpus consists of digit sequences coming from various Italian Corpora. It has been divided into training, development and test. Furthermore, it has been split into clean and noisy speech, according to the presence of many noise events in the signal or low SNR. Audio files are sampled at 16kHz, 16 bit PCM, mono, Windows wav format. For training and development set, transcription at word level has been provided in two separate text files, containing on each row the audio filename followed by the transcription.

The evaluation process is based on Minimum Edit Distance between the transcriptions coming out from the recognizer and the orthographic annotations. Accuracy will be calculated at word and phrase levels and participants which need to enroll the ASR at finer level than phrase have to provide by themselves for the annotation.

Test data have been provided only for one week before the final evaluation workshop in the form of audio files and two lists of filenames (clean and noisy) and test results have been provided with the ASR participant system name, the list of filenames and the corresponding recognition results.

With respect to the results submitted by the participants measurements of Word Accuracy and Sentence Accuracy have been considered. Word Accuracy is defined as:

$$(1) \quad WA = 100 - \frac{I \square S \square D}{N} \times 100$$

⁶ EVALITA web site: <http://evalita.fbk.eu/>

where, I is the number of inserted words, S is the number of substitutions, D is the number of the deletions, N is the number of words in the reference

Sentence Accuracy: is defined as:

$$(2) \quad SA = \frac{H}{M} \times 100$$

where, H is the number of sentences correctly recognized, M is the number of sentences in the reference. For this paper's purpose, we focused on Word Accuracy.

Three of the most used open source ASR tools were considered in this work, i.e. CSLU Toolkit, SONIC, and SPHINX, simply because promising results were obtained in the past on similar digit recognition tasks. The clean, noisy and clean+noisy (full) EVALITA digit task was chosen mainly because it is a simple and natural starting point to test a common evaluation methodology within the EVALITA framework.

From various experiment results few simple considerations can be made:

- beyond the fact that one of the main difficulties was that of finding similarity among the three ASR systems in order to choose comparable configurations because every system has its own completely distinct architecture and consequently the configuration parameters are hardly comparable, an homogeneous and unique test framework for comparing different Italian ASR systems was quite possible and effective if the results produced by the best WA-score configuration were compared for each of the systems;
- in term of better recognition performance the configuration in which Language Model is not used at all (LM = 0) is the best because this should really simulate complete independence between connected digits while using words as basic recognition unit;
- CSLU Toolkit is incredibly good in recognizing clean digit sequences (99.94%), even if we are expecting lower performance with the real test data;
- SONIC is the best system in all noisy and full situations and we believe this is mainly due to the adoption of the PMVDR features;
- SPHINX is quite more sensible to AM specialization than other systems and clean models can not recognize noisy speech;

Finally we should conclude that the EVALITA campaign was quite effective in forcing various Italian research groups to focus on similar recognition tasks working on common data thus comparing and improving various different recognition methodologies and strategies, and we hope more complex task and data will be exploited in the future.

MISURA OGGETTIVA DELL' INTELLIGIBILITÀ DEL PARLATO IN APPLICAZIONI FORENSI

G. Costantini, A. Paoloni*, M. Todisco

[*pao@fub.it](mailto:pao@fub.it)

1 Introduzione

I sistemi per il restauro dei segnali audio sono stati creati principalmente al fine di migliorare la qualità delle registrazioni storiche contenute nei supporti di diverso tipo. Queste tecniche sono rivolte essenzialmente ad audio di buona qualità, i segnali sono interessati da rumori prevalentemente classificabili come click o fruscio, un rumore bianco con un livello molto inferiore a quello del segnale. Altri rumori che possono essere trovati in vecchi dischi da restaurare sono il brusio, le armoniche di 50 Hz provenienti dalla rete elettrica, e, in alcuni casi, la saturazione ampiezza del segnale con la conseguente distorsione del suono.

I sistemi adottati per il restauro dei segnali variano necessariamente con le applicazioni. Nelle applicazioni forensi l'obiettivo non è quello di produrre segnali di buona qualità audio, ma di migliorare l'intelligibilità delle registrazioni o la possibilità di riconoscimento vocale. La qualità dei segnali, in termini di larghezza di banda e il rapporto segnale / rumore, può essere già abbastanza per provocare il fallimento dei sistemi di restauro. Più in generale, l'esperto non si confronta solo con i rumori, ma con i disturbi. Ciò significa che l'esperto forense non è che fare unicamente con il miglioramento SNR.

In questo documento discuteremo il metodo di valutazione della intelligibilità del parlato attraverso la pulizia del segnale.

2 Tipologia dei disturbi e di soluzioni disponibili

Come accennato in precedenza, i disturbi possono essere i rumori o fenomeni di altra natura. Rumori possono essere additivi o moltiplicativi. Altri fenomeni sono in genere quelli acustici (eco, riverbero). Un'altra classificazione può essere fatta separando i disturbi indotti dal registratore (più il canale) e quelli prodotti dall'ambiente. In seguito, abbiamo rilievo la topologia di disturbo e le soluzioni disponibili.

Per migliorare la qualità del segnale disturbato si ricorre solitamente a tecniche di riduzione del rumore denominate genericamente “filtraggio” o “restauro”.

È bene puntualizzare che per filtraggio si intende un qualsiasi tipo di eliminazione selettiva di contenuti energetici del segnale in bande selezionate. Nella misura in cui in ogni banda spettrale può verificarsi la compresenza di segnale e di rumore, il filtraggio è un'operazione che diminuisce sempre l'informazione e diminuisce pertanto l'entropia. In alcuni casi, che vedremo, l'operazione di filtraggio può però migliorare il rapporto segnale-rumore, laddove la quota energetica eliminata sia prevalentemente quella delle bande occupate dal rumore. Se sfortunatamente le bande sono indistinguibili, i filtri sono del tutto

inutili. L'eliminazione di una parte di rumore "fuori banda" darà l'impressione, di un segnale più pulito e gradevole, ma se la perdita di comprensibilità è determinata dall'interferenza vocale e non dal rumore del canale, l'effetto è trascurabile e in alcuni casi porta ad un peggioramento dell'intelligibilità.

Vi sono appositi apparati di elaborazione del segnale che i laboratori di analisi e restauro dei segnali mettono a disposizione. (CEDAR CAMBRIDGE, Wave z noise, AudioCUBE, etc.) Essi sono indirizzati alle varie tipologie di distorsioni e di alterazioni che vengono riscontrate.

Questi trattamenti, che inseriscono elementi artificiali, non presenti nel messaggio originale, possono facilmente diventare alterazioni indebite, in qualche caso manipolazioni illegali (morphing). Si tratta di:

de-hiss: trattamento per la eliminazione del rumore bianco o rosa a largo spettro. Tipicamente si tratta del rumore di fondo introdotto dalle apparecchiature elettroniche (rumore termico ed altro). Una quota del rumore può derivare dall'ambiente che però, generalmente, produce rumore più "colorato" spettralmente.

de-clicking: elimina il rumore di tipo impulsivo, di breve durata e spesso di alta intensità. È un tipo di rumore inaccettabile e fastidioso nelle riproduzioni musicali o in una colonna sonora di un film.

de-buzz: il rumore presente nell'ambiente può essere generato da motori, ventilatori o altri dispositivi che producono un tipo di rumore continuo o a tratti, caratterizzato spettralmente da un gran numero di armoniche.

de-clipping: Talvolta, per errori di regolazione della strumentazione, il segnale acustico satura gli amplificatori e viene deformato dapprima ed infine tagliato di netto agli apici. All'ascolto questo tipo di distorsione produce una grave compromissione della qualità sonora.

Nell'uso delle tecniche sopra esposte per ciascun particolare regime di valorizzazione del linguaggio, deve essere operato un compromesso tra la quantità di rumore rimosso e la distorsione introdotta come un effetto collaterale. Se viene rimosso troppo rumore, la distorsione può essere peggiore per l'ascoltatore che la permanenza del rumore.

3. Problemi nel campo della riduzione del rumore in applicazioni forensi

Le soluzioni sopra elencate permettono, con efficienza maggiore o minore a seconda dell'importanza e della tipologia del disturbo, del tempo impiegato per correggerlo e della qualità degli strumenti di elaborazione del segnale a disposizione, di migliorare la "qualità"

del segnale audio. Con una migliore qualità audio è sicuramente più semplice effettuare la trascrizione del segnale registrato. È tuttavia probabile che, in quasi tutti i casi, non si sia ottenuto alcun reale incremento della intelligibilità ma piuttosto un peggioramento, più o meno lieve. Boll nel capitolo intitolato "Speech enhancement in the 1980", tratto dal libro *Advances in speech signal processing* Marcel-Dekker, 1991 osserva che: "... il gruppo ha concluso che non vi è attualmente un approccio che migliori l'intelligibilità misurata sulla base di un test diagnostico come il DRT. Tale affermazione probabilmente è ancora valida infatti in un recente lavoro di Hu e Louzou nessuno degli 8 sistemi di restauro del segnale da loro utilizzati è riuscito a migliorare l'intelligibilità.

4. Intelligibilità: valutazione

Un altro problema nella trascrizione forense è quello di misurare l'intelligibilità del parlato. Troppo spesso vediamo trascrizioni che riflettono piuttosto le opinioni del trascrittore che le opinioni di chi ha parlato. La norma ISO 9921 in merito alla "Valutazione della comunicazione verbale" definisce l'intelligibilità del parlato, come: "una misura di efficacia nel comprendere il linguaggio".

Al fine di valutare l'affidabilità di una trascrizione e l'efficacia di un sistema di "restauro del segnale vocale", è necessario definire una misura dell'indice di intelligibilità strettamente correlata con la performance di un gruppo di ascoltatori sul segnale stesso.

Tradizionalmente la intelligibilità consiste nella percentuale di segnale viene effettuata soggettivamente da un gruppo di ascoltatori sulla base della percentuale delle identificazioni corrette (vedi tab. I). Il segnale noto può essere costituito da frasi, parole o semplici suoni privi di significato (logatoms). In quest'ultimo caso sarà verificata l'intelligibilità dei singoli fonemi.

La maggior parte della letteratura viene dal mondo IT, dove il problema è quello di studiare l'impatto del canale di trasmissione e dei codificatori sull'intelligibilità del segnale vocale. Per fare ciò, gli algoritmi proposti spesso utilizzano un approccio dual-sided e operano il confronto tra il segnale iniziale (in chiaro) e il segnale trasmesso.

Tale approccio non è utilizzabile in un contesto legale in quanto il perito ha una sola versione del segnale. In tal modo è necessario operare una valutazione single-sided ossia sulla base del solo segnale in uscita. Per poter valutare le prestazioni dei sistemi di restauro, ma più ancora per poter valutare l'attendibilità di una eventuale trascrizione è necessario poter stimare l'intelligibilità del segnale da trascrivere.

Method	Standard ref.	Type	comments
PB - Phonetically Balanced Word	ISO TR 4870	Subjective	
MRT - Modified Rime Test		Subjective	
DRT - Diagnostic Rime Test		Subjective	
AI - Articulation Index	ANSI S 3.5	objective	instrument based
STI - Speech Transmission Index	IEC 60268-16 - 1998	objective	instrument based
%AL - Articulation loss of Consonants	Peutz	objective	

Tab. I – Speech Intelligibility Tests

Le misurazioni oggettive non misurano la comprensibilità, ma i parametri fisici in grado di predire l'intelligibilità sulla base di un certo modello. I metodi più frequentemente usati sono l'indice articolazione (AI) e la Speech Transmission Index (STI).

L'obiettivo è quello di stimare l'intelligibilità residua con sufficiente precisione.

Si sono pertanto studiati degli esperimenti per valutare la correlazione tra i parametri fisici misurati sul segnale e l'intelligibilità.

5. I corpora

Per poter valutare la correlazione tra le misure oggettive effettuate su segnale rumoroso e l'intelligibilità del segnale stesso valutata come tradizionalmente sulla base della percentuale di segnale correttamente identificato si sono predisposti alcuni corpus tratti dal progetto europeo SAM eurom 1. In particolare sono stati utilizzate le parole rimate e le frasi bilanciate. I due corpus sono stati opportunamente resi rumorosi aggiungendo rumore Rosa, rumore di automobile, rumore di voci e riverberazione. Ciascuna tipologia di rumore compariva in tre diversi gradi di rapporto segnale rumore. Alla fine delle operazioni pertanto si rinviava disporre di 24 differenti corpus che sono stati sottoposti ciascuno ad un gruppo di 10 ascoltatori normoudenti. Il risultato delle prove è riportato nelle tabelle 2-13.

6. L'esperimento

Un primo esperimento è stato effettuato utilizzando per la valutazione dell'intelligibilità il rapporto Segnale Rumore pesato con la curva "A" e l'indice di articolazione "AI".

L'esperimento ha messo in luce la correlazione fra i dati nelle particolari condizioni di banda che sono tipiche dell'applicazione forense.

7. Sintesi e conclusioni

Il presente documento riassume le tecniche di restauro del segnale vocale nelle applicazioni forensi.

Questa applicazione particolare è infatti caratterizzata dai seguenti punti:

- il segnale è molto disturbato, con rapporti S / N vicini a 0;

- il segnale disponibile è solamente quello degradato su canale monofonico;

- è necessario misurare con metodi oggettivi, l'intelligibilità effettiva del segnale.

- L'ultimo punto è fondamentale per garantire l'affidabilità della trascrizione.

Secondo la letteratura attuale sembra impossibile che una qualsiasi elaborazione del segnale possa migliorare l'intelligibilità del segnale vocale, pertanto una particolare attenzione dovrebbe essere data, nel nostro punto di vista, nello stabilire una procedura per valutare oggettivamente l'intelligibilità di un discorso rumoroso.

NUOVE FUNZIONALITÀ PER IL PORTALE CLIPS

Aurelio de Rosa, Francesco Cutugno

Dipartimento di Scienze Fisiche – gruppo NLP, Università degli Studi di Napoli – “Federico II”

aurelioderosa@gmail.com, cutugno@na.infn.it

Il lavoro realizzato è il rifacimento del portale web CLIPS [1] corredato da alcune nuove funzionalità per lo studio e l'analisi statistica di corpora linguistici. Il sito è stato sviluppato in due lingue, italiano ed inglese, così da soddisfare le esigenze di eventuali visitatori stranieri. Esso ha lo scopo di essere punto di riferimento per lo scambio di dati ed informazioni tra studiosi della linguistica grazie ai documenti pubblicati circa l'andamento della ricerca e le novità che gli autori vorranno condividere con l'esterno.

Come detto, il portale è fornito di vari strumenti per l'analisi dei corpora. Il primo è un convertitore che opera su più livelli di annotazioni ed etichettatura (fonetico, sillabico, ed altri) codificati in formato TIMIT, comprese le informazioni temporali, e dà origine ad un AGset, che rispetta appieno gli standard AG (Annotation Graphs – [2]) in formato XML. Gli AG permettono di gestire i corpora di parlato per i quali i metodi di rappresentazione tradizionali e di rappresentare le annotazioni trascritte da un segnale contenente risorse linguistiche [3]. Attualmente solo l'amministratore del portale, attraverso l'apposita interfaccia che gli permetterà di selezionare il corpus di interesse, può generare il documento XML che verrà salvato sul web server. Uno sviluppo futuro decisamente interessante è quello di rendere questo strumento pubblico permettendo di far scaricare al richiedente il file dopo che esso è stato generato.

La seconda funzionalità, attraverso l'implementazione grafica del linguaggio di query AGQL per scopi fonetici, si occupa di eseguire interrogazioni sui file XML creati. In esso è prevista la possibilità di compiere la stessa query contemporaneamente su più corpora e visualizzare i risultati insieme. L'elaborazione è effettuata sui file XML precedentemente creati e presenti sul server CLIPS secondo diversi criteri, ad esempio sul piano fonetico o acustico o cercando delle stringhe all'interno dei vari livelli concatenando tra loro fino a tre livelli di selezione.

Un altro degli applicativi presenti genera statistiche generali inerenti i corpora. Una volta che è stato selezionato il corpus od i corpora dialogici sui quali agire (sugli altri non è possibile) si devono poter creare statistiche, per ogni dialogo presente, circa il nome identificativo, il numero di parole pronunciate, il tempo per il quale parla ogni interlocutore. La selezione è effettuabile dal livello più basso per un prefissato corpus crescendo fino alla scelta di tutti i corpora presenti sul server web. Ogni corpus analizzato mostra in aggiunta alle statistiche sui singoli parlanti, anche un quadro riassuntivo delle statistiche rilevate che nei conteggi include, nel caso ve ne siano, i dati dei suoi sottocorpora. Ciò è stato fatto poiché se si esegue il programma sul livello più alto, che quindi racchiude tutti i corpora presenti, oltre alle singole statistiche si otterrà un quadro riassuntivo di tutto il materiale presente [4].

Un'altra funzionalità proposta consente, partendo da un script fatto in PERL che effettua il parsing in parti del discorso (POS – morfosintassi) di un testo dato in ingresso, di analizzare il livello .WRD di un corpus prescelto e produrre un file XML contenente le POS di quel corpus. Lo script usato è il TreeTagger, uno strumento per annotare il testo con

informazioni sul POS, acronimo di Part of Speech, ed il lemma delle forme. Esso è stato sviluppato da H. Schmid [5] nel progetto TC presso l'Istituto di Linguistica Computazionale dell'Università di Stoccarda per la lingua tedesca. Questo programma è stato esteso in varie lingue ed in particolare il suo uso per la lingua italiana è dovuta al lavoro di M. Baroni [6]. Il dialogo presente nel corpus è il nodo radice dell'albero del documento e ha un "id" che lo identifica univocamente. Il valore di questo campo deve essere lo stesso che è usato per individuare i file appartenenti al dialogo in questione. Ogni turno appartenente al dialogo è suo figlio diretto nell'albero del file ed è anch'esso corredato dell'attributo "id" che è un numero sequenziale rappresentante l'ordine di avvicendamento dei turni. Ogni parola presente nel turno è figlia diretta di quello di appartenenza e occorre dei seguenti attributi: "id" che è un numero indicante l'ordine in cui il termine è stato analizzato all'interno del dialogo; "start" identifica il tempo nel quale il dialogante ha iniziato a pronunciare il termine; "end" corrispondente al tempo in cui la pronuncia ha fine; "POS" indica la funzione della parola (in base all'analisi morfosintattica) all'interno della frase.

L'ultima funzionalità creata, partendo dalla selezione di un corpus, può dare origine a statistiche sul numero di occorrenze dei lemmi presenti in esso e, per ognuno di essi, quante sono le occorrenze delle forme appartenenti a tale lemma. L'applicativo cercherà il file XML che estrae i POS per i quali è indicato anche il lemma, così da analizzarlo e compierà la medesima azione per tutti i sottocorpora presenti (nel caso ve ne fossero). Tutti i lemmi e le forme sono memorizzati in appositi alberi AVL (dal nome degli autori G.M. Adelson-Velsky e E.M. Landis - [7]). La configurazione dell'albero generato è degna di nota, infatti, il principale è un AVL contenente nodi che includono il lemma, il numero di occorrenze trovate fino a quel momento ed il riferimento alla radice dell'albero, anch'esso AVL, dei termini che lo hanno come lemma. I nodi di questo sottoalbero racchiudono i seguenti dati: il termine analizzato, il POS ed il numero di volte che ricorre. Dopo aver prelevato i dati dal documento XML, il primo passo è verificare se il lemma estratto esiste già nell'albero. In caso affermativo, viene solo aggiornato il contatore delle frequenze, altrimenti si inserisce un nuovo nodo. Anche se il lemma fosse già presente, non è detto che lo sia il termine individuato, quindi si effettuerà un'altra ricerca e nel caso non sia presente, verrà inserito.

Riferimenti bibliografici

- [1] Savy R. Cutugno F., "CLIPS: diatopic, diamesic and diaphasic variations of spoken Italian. In proceedings off Corpus Linguistics 2009", Liverpool July 20-23 2009
- [2] Steven Bird and Mark Liberman, "A formal framework for linguistic annotation", Speech Communication 33(1,2), pp 23-60, 2001
- [3] E. Cecere, "Un sistema di gestione e interrogazione di corpora linguistici multilivello: ambiente di sviluppo integrato per querying e gestione", Tesi di laurea in Scienze e tecnologie informatiche, Università di Napoli, Federico II, 2008
- [4] A. De Rosa, "Progetto CLIPS: Un portale per lo studio e l'analisi di corpora linguistici", Tesi di laurea in Scienze e tecnologie informatiche, Università di Napoli, Federico II, 2009
- [5] H. Schmid, <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>, 2009

- [6] Schmid H., Baroni M., Zanchetta E., Stein A., "The enriched treetagger system". *Intelligenza Artificiale*, 2007. v. 4, 2, p. 22-23
- [7] G. Adelson-Velskii and E.M. Landis, "An algorithm for the organization of information", *Doklady Akademii Nauk SSSR*, 146:263–266, 1962

SpLaSH (SPOKEN LANGUAGE SEARCH HAWK): IL SISTEMA GENERAL PURPOSE PER LA GESTIONE DI ANNOTAZIONI TIME-ALIGNED E TEXT-ALIGNED

Sara Romano

Dipartimento di Scienze Fisiche – gruppo NLP, Università degli Studi di Napoli – “Federico II”
sararomano80@gmail.com

Negli ultimi anni, numerosi corpora linguistici sono stati arricchiti con l’aggiunta di diversi livelli di annotazione. I corpora linguistici possono contenere sia annotazioni che descrivono le proprietà acustiche del linguaggio parlato (intonazione, fonemi, ecc) sia annotazioni che riguardano l’analisi testuale proveniente dalla trascrizione del segnale (sintassi, pragmatica, morfosintassi, ecc). Le annotazioni provenienti dall’analisi sulle trascrizioni sono indipendenti dal tempo ed etichettano sequenze di unità linguistiche (che di solito sono rappresentate dalle parole).

Sulla base di quanto detto è possibile classificare tutti i tipi di annotazioni linguistiche in due categorie: annotazioni time-aligned (TMA) ed annotazioni text-aligned (TXA). Per annotazioni TMA intendiamo tutti i tipi di annotazione che descrivono fenomeni di segmentazione (sillabe, morfemi, parole, ecc), intonazione, disfluenze, ecc. Per annotazioni TXA, invece, intendiamo analisi testuali quali semantiche, sintattiche, morfosintattiche, ecc.

Inoltre, l’arricchimento dei corpora linguistici mediante l’aggiunta di nuovi di livelli di annotazione ha comportato un arricchimento delle gerarchie. In particolare abbiamo due tipi di gerarchie: le gerarchie inter-livello e quelle intra-livello. Le prime sono le classiche gerarchie che si instaurano tra diversi livelli di un corpus; le seconde, invece, sono tipi di gerarchie che si instaurano all’interno di un unico livello (ad esempio, si consideri l’analisi testuale in cui una frase è composta da frasi più piccole che a loro volta sono composte da parole).

La produzione di corpora linguistici è in continua evoluzione. Molto spesso accade che assieme ad essi vengono prodotti dei sistemi software creati ad hoc per permetterne la gestione. La mancanza di uno standard internazionale che determini il formato di storage dei corpora ne impedisce il riutilizzo e, di conseguenza, ne accresce lo sforzo prodotto per convertirli in formati differenti e spesso anche lo sforzo per adattare i tool ai diversi formati.

In questo contesto nasce l’esigenza di avere sistemi general purpose in grado di gestire i corpora linguistici indipendentemente dal loro formato. Vi sono diversi sistemi di questo tipo ma sono ancora numerose le limitazioni che impongono. Un esempio di tali limitazioni sono: la definizione a priori di gerarchie inter-livello, mancata gestione di gerarchie inter-livello, scarsa gestione di gerarchie che comprendono contemporaneamente livelli TXA e TMA.

In questo lavoro presentiamo SpLaSH (Spoken Language Search Hawk), un sistema general purpose per la gestione di corpora linguistici multilivello. SpLaSH accetta dati provenienti da differenti corpora ed è capace di legare tra loro annotazioni aventi differenti formati. Questo sistema fornisce la possibilità gestire annotazioni TMA ed annotazioni TXA. Le prime sono codificate attraverso i grafi di annotazione [1] rappresentati in XML

(AG-XML). I grafi di annotazione presentano una struttura logica in grado di accomunare i formati di annotazioni TMA esistenti (Timit [2], Partitur “<http://www.fon.hum.uva.nl/praat/>”, Praat TextGrid [3], ecc). I grafi di annotazione sono strutture dati in cui i nodi rappresentano “ancore” al segnale e gli archi sono etichettati con il livello e il valore delle annotazioni. Le relazioni di precedenza sono definite attraverso le etichette temporali che risiedono sui nodi mentre le relazioni di inclusione, coincidenza e sovrapposizione sono espresse dalla posizione degli archi all’interno dei grafi di annotazione. Le annotazioni TXA, invece, sono codificate in XML in quanto questo rappresenta un valido strumento per rappresentare la struttura gerarchica e ricorsiva di tali annotazioni. Di conseguenza, la struttura logica che rappresenta i dati TMA è un albero etichettato dove le foglie contengono i valori testuali delle annotazioni.

Una delle principali caratteristiche del sistema SpLaSH è quella di riuscire ad integrare in un unica struttura le annotazioni TMA e TXA. In questo modo possono essere definite relazioni temporali anche tra i livelli TXA in quanto, nella struttura integrata, questi ereditano tali riferimenti dai dati TMA.

Infine, SpLaSH fornisce la possibilità di effettuare ricerche avanzate sui dati contenuti all’interno dei corpora. Per soddisfare tale esigenza, il sistema presenta un insieme di interfacce che aiutano l’utente nella composizione delle interrogazioni. Vi sono tre categorie di interrogazione possibili: simple query (possibili su annotazioni TMA o TXA), sequence query (possibili solo su annotazioni TMA) e cross query (possibili su annotazioni TMA e TXA integrate). Le simple query su dati TMA permettono di effettuare ricerche in base alle relazioni temporali. Le simple query su dati TXA, permettono di effettuare ricerche tenendo in considerazione la struttura gerarchica e ricorsiva di tali annotazioni. Le sequence query permettono di esprimere la chiusura transitiva sui valori degli archi dei grafi di annotazione. Utilizzando questo tipo di interrogazione è possibile selezionare sottosequenze di annotazioni contigue da un corpus TMA. Infine, le cross query permettono di effettuare ricerche all’interno della struttura che integra le annotazioni TMA con quelle TXA. In tale tipo di interrogazioni è possibile specificare ricerche considerando la struttura gerarchica delle annotazioni TXA e le relazioni temporali che possono esistere tra le annotazioni TMA e quelle TXA ereditate nel processo di integrazione.

Riferimenti bibliografici

Bird, S. and Lieberman, M., “A formal framework for linguistic annotation”, *Speech Commun.* 33, Issues 1-2, Pages 23-60, 2001.

Garofalo, J.S., Larnel, L.F., Fisher, W.M., Fiscus, J.G., Pallett, D.S. and Dahlgren, N.L., “The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus”, Technical Report NISTIR 4930, NIST, 1993.

Schiel, F., Burger, S., Geumann, A., and Weilhammer, K., “The Partitur Format at BAS”, in *Proceedings of the First International Conference on Language Resources and Evaluation*, Pages 1295-1301, 1998.

EVALUATING SPEECH RATE BY REFERRING ON AN INTENSITY-BASED SYLLABLE NUCLEI SEARCH ALGORITHM

Antonio Origlia¹, Vincenzo Galatà²

¹Dipartimento di Scienze Fisiche – gruppo NLP, Università degli Studi di Napoli – “Federico II”,

²Dipartimento di Linguistica – Laboratorio di Fonetica, Università della Calabria

antori@gmail.com, vgalata@libero.it

Calculating the speech rate in a speech file can be a very much time consuming task if performed manually. At the same time the task difficulty increases noteworthy with the amount of data to be processed. To overcome this problem, we propose here a procedure to evaluate speech rate by referring on an intensity-based syllable nuclei search algorithm.

Syllable nuclei search through intensity profile peaks detection has been performed in the past to find syllable nuclei in a speech file and to evaluate speech rate (De Jong & Wempe, 2009), but this promising approach has always been limited by the need of being somewhat prudent when accepting intensity peaks as syllable nuclei. This is because artifact peaks are common and force intensity-based syllable nuclei search algorithms to look for prominent peaks only, at the price of overseeing many unstressed syllables. Voice Activity Detection (VAD) can improve these algorithms performances by allowing to relax some constraints without risking to introduce too many artifacts as syllable nuclei. Although it has been stated that prominent peaks detection is sufficient for evaluating speech rate, we believe that developing an approach capable of detecting secondary syllables without suffering exceedingly from noise and common speech artifacts could give us the opportunity to investigate more in depth complex communication strategies.

The approach we present here employs traditional pitch analysis-based VAD while taking into account Harmonic-Noise Ratio (HNR) to identify syllable nuclei. An octave jumps check is also performed to keep insertion errors under control.

HNR revealed itself to be a powerful feature to distinguish artifact intensity peaks from actual syllable nuclei due to its capability to help identifying areas of the speech signal containing vowels, which are harmonic in nature, to disambiguate a number of situations where an energy peak could have been an actual syllable nucleus or a mere reference point to set a syllable boundary and to recover many energy peaks that would have been discarded by VAD. HNR, however, cannot be used alone because, from what we have seen, while VAD underestimates syllable nuclei, HNR tends to overestimate them. A combination of both is therefore needed. Octave jumps check relies on abnormal pitch peaks detection to identify undependable areas of the signal and to discard intensity peaks falling in these regions.

In addition to this, we profiled a number of common artifacts causing both insertion and deletion errors so we were able to reject and accept candidates of syllable nuclei accordingly using specialized procedures.

Studies on vocal expression of emotions have to date shown many features to be involved and affected: speech rate and intensity are among them. Speech rate has also been proved to be a useful parameter for listeners to differentiate emotions conveyed in speech. In a study by Breitenstein, Van Lancker, Daum (2001) using systematically pitch and rate altered emotional speech stimuli showed that listeners associated a slow rate with sad

labels, while a fast rate was associated to angry, frightened or neutral labels. The results obtained in a study by Kienast, Paeschke & Sendlmeier (1999) suggest, in fact, that accuracy of articulation and durational aspects are useful parameters to characterize, if not all of them, at least some emotions (like for example anger, fear and sadness). For these reasons and for the presence of a wider range of features differences in emotional speech we decided to verify our approach on a small corpus of acted emotional speech recorded from Italian actors and naïf subjects. Our purpose is firstly to test our approach according to the different emotional speech files recorded in the corpus and secondly to verify how much speech rate is affected according to the different emotions expressed. In this paper we will mainly refer to speech conveying emotions but other applications of the algorithm here proposed could be found, for example, to study speaker strategies to draw auditory attention on specific dialogue parts.

References

- Breitenstein, C., Van Lancker, D., Daum, I., (2001), The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample, in: *Cognition & Emotion* 15(1): 57-79.
- de Jong, N. & Wempe, T. (2009), Praat script to detect syllable nuclei and measure speech rate automatically, *Behavior Research Methods*, 41(2): 385-390.
- Kienast, M., Paeschke, A. & Sendlmeier, W. F. (1999), Articulatory reduction in emotional speech, *Proceedings Eurospeech 99*, Budapest, Vol. 1, pp. 117-120.

SESSIONE POSTER

E. Orso, M. Colegato, F. Rapa, C. Zmarich *L'emergere della fonologia nei bambini dai 18 ai 27 mesi: analisi statistica degli errori di sostituzione*

L. Spreafico *I riempitivi nel parlato plurilingue*

L. Righero, M. Danieli *Polirematiche interiettive ed espressione delle emozioni*

A. Origlia, G. Abete *A syllabification algorithm using energy, pitch and harmonicity related features*

S. Giulivi, M. Vayra, C. Zmarich *Dal babbling alla lingua materna. Tendenze universali e aspetti linguo-specifici nello sviluppo fonetico di due bambini italiani*

S. Schmid *La neutralizzazione del contrasto di lunghezza delle fricative nel dialetto di Zurigo*

G. Tisato *NavigAIS - AIS Digital Atlas and Navigation Software*

L. Mori *Mixed Maltese English: fenomeni di interferenza fonetica lungo il continuum Maltese - Inglese nel repertorio linguistico di Malta*

V. Galatà, L. Romito *Gli italiani e il decoding di emozioni vocali in quattro lingue europee*

F. Orletti, A. Paoloni *Identificazione dei segnali*

Marco A. Piccolino-Boniforti *Distinguere pattern acustici con identica trascrizione fonemica ma struttura morfologica differente: un modello computazionale*

L'EMERGERE DELLA FONOLOGIA NEI BAMBINI DAI 18 AI 27 MESI: ANALISI STATISTICA DEGLI ERRORI DI SOSTITUZIONE

°Emanuele Orso, °Michela Calegario, °Federica Rapa, *Serena Bonifacio, °Claudio Zmarich

°ISTC-CNR, sede di Padova, via Martiri della Libertà 2, 35137, Padova (I),

*IRCSS Burlo Garofolo, Trieste

emanuele.orso@gmail.com, zampiericalegario@aliceposta.it, federicarapa@hotmail.com,
logopedia@burlo.trieste.it, claudio.zmarich@pd.istc.cnr.it

Secondo la Fonologia Articolatoria le unità discrete del linguaggio sono i gesti articolatori e non i tratti distintivi o i segmenti (Studdert-Kennedy & Goldstein, 2002). I prodotti di questi gesti sono variabili, ma i gesti in sé sono discreti e indipendenti dal contesto: si realizzano in funzione del bersaglio e corrispondono alla creazione di un certo grado di costrizione in un preciso luogo lungo il tratto vocale, per mezzo di uno tra i sei "organi" articolatori che sono anatomicamente distinti e indipendenti: labbra, corona della lingua, dorso della lingua, radice della lingua, velo palatino, glottide. Poiché questi gesti articolatori sono distinti, possono funzionare come unità contrastive in fonologia (distinguendo per es. /t/ vs /p/). Un altro tipo di contrasto fonologico coinvolge lo stesso "organo" articolatorio ma è specificato da differenze nel grado di costrizione (per es. /t/ vs. /s/). Nello sviluppo fonetico, è stato dimostrato che le varianti lessicali (cioè le produzioni errate) prodotte dai bambini di due anni nei loro vari tentativi di realizzare le parole bersaglio, contengono molti errori in termini segmentali (cioè molti foni scorretti), ma pochi errori in termini degli "organi" da impiegare (Studdert-Kennedy & Goodell, 1995). Come per gli adulti, gran parte degli errori linguistici compiuti dai bambini sono spiegabili dal punto di vista articolatorio: errori di scelta dell'organo (cioè di luogo di articolazione), errori di dimensionamento del gesto articolatorio (cioè di grado di costrizione, equivalente al modo di articolazione della tabella IPA) ed errori nell'organizzazione temporale di gesti articolatori adiacenti (errori di armonia, di inserzione, di cancellazione). Studdert-Kennedy & Goldstein (2002) ipotizzano che l'identificazione dell'"organo" (articolatorio) preceda l'abilità di imitare uno specifico gesto nei suoi dettagli spaziali e temporali, e infatti Goldstein (2003) ha dimostrato che gli errori che coinvolgono il luogo di articolazione si verificano raramente, mentre gli errori dovuti a caratteristiche diverse di modo e sonorità sono molto più frequenti.

In questo studio si vuole testare, sui bambini che apprendono l'italiano, l'ipotesi sperimentale della Fonologia Articolatoria che, negli errori linguistici infantili, le categorie fonetiche legate al luogo d'articolazione (Labiali, Coronal, Dorsali) si mantengono di più rispetto alle categorie legate al modo di articolazione (Occlusive, Fricative, Approssimanti). Per questo sono state analizzate le produzioni lessicali di 13 bambini (7 femmine e 6 maschi), audioregistrati in 4 tappe successive a 18, 21, 24 e 27 mesi (Zmarich e Bonifacio, 2004). Le produzioni errate dei bambini dovevano presentare la stessa struttura sillabica del target adulto. Alla fine sono state confrontate le consonanti errate iniziali di sillaba (con struttura CV e CVC) con le consonanti delle sillabe appartenenti alle parole del lessico adulto corrispondenti (target).

L'analisi statistica delle produzioni infantili ha considerato innanzitutto il possibile effetto della posizione della sillaba nella parola (iniziale, mediana, finale) sulla probabilità di occorrenza di un dato fonema consonantico. Infatti Davis, MacNeilage & Matyear (2002) hanno studiato l'alternarsi delle consonanti nelle prime produzioni infantili (con

struttura CVCV e CVC) evidenziando alcune restrizioni combinatorie giustificate dai pattern motori tipici dei bambini dai 12 ai 25 mesi di età. Queste limitazioni alla combinazione dei fonemi in sequenza, presenti anche nel linguaggio adulto sotto forma di restrizioni fonotattiche, rendono il campione di consonanti esaminato non casuale perché le osservazioni non sono più stocasticamente indipendenti né identicamente distribuite (Chiari & Castagna 2005; Gomez 2007). Per risolvere il problema si sono considerate solo le sillabe iniziali di parola ottenendo un campione di osservazioni identicamente distribuite e, come viene comprovato dall'applicazione del "runs test" (un noto test statistico per verificare l'indipendenza delle osservazioni), stocasticamente indipendenti.

Una strategia alternativa per risolvere il problema della non casualità del campione di consonanti si ispira al test di significatività proposto da Goldstein (2003) che è basato su un particolare tipo di ricampionamento delle osservazioni. Questa particolare procedura è stata inquadrata nel più generale contesto teorico dei test di permutazione, una classe di test statistici non parametrici che godono di due proprietà molto utili nell'analisi dei campioni di linguaggio: permettono inferenze corrette anche nel caso di campionamento non casuale e sono asintoticamente (quando la numerosità del campione tende all'infinito) tanto potenti quanto i corrispondenti test parametrici (Pesarin 1999). Un test di permutazione, in parte simile a quello proposto da Goldstein (2003), è stato applicato al campione di Zmarich & Bonifacio (2004), contenente 429 casi di errori linguistici infantili, per confrontare le proporzioni di accordo fra produzioni infantili e target adulti relative a consonanti sostituite ma che condividono almeno una proprietà del gesto articolatorio. Questo test di significatività ha consentito di indagare quanto le proporzioni di accordo per luogo e per modo di articolazione sono diverse fra loro al netto delle percentuali di condivisione "baseline" che si avrebbero nell'ipotesi di accoppiamento casuale fra produzione infantile e target adulto (ipotesi che i bambini facciano semplici giochi articolatori senza tentare di riprodurre i target adulti). Un'altra ipotesi da considerare è che i bambini si limitino a riprodurre un set ristretto di gesti articolatori del target che già rientrano nelle loro capacità (Schwartz & Leonard, 1982).

Nel campione di produzioni infantili esaminato, in accordo con l'ipotesi sperimentale, la percentuale di condivisione per le categorie fonetiche di luogo è risultata significativamente maggiore di quella relativa alle categorie fonetiche di modo, al netto del sopracitato effetto dovuto alla possibile casualità degli accoppiamenti. La differenza fra ciascuna percentuale di condivisione e la propria "baseline" è stata infine quantificata tramite l'indice statistico "kappa di Cohen" (Siegel & Castellan 1992), che misura il grado di accordo tra le risposte qualitative di due persone.

Riferimenti bibliografici

Chiari I., Castagna S., *La fonotassi statistica dell'italiano e del tedesco: i nessi consonantici*, in De Mauro T., Chiari I. (a cura di), *Parole e numeri*, Roma : Aracne, 67-84.

Davis B.L., MacNeilage P.F. & Matyear C.L. (2002), Acquisition of serial complexity in speech production: a comparison of phonetic and phonological approaches to first word production, *Phonetica*, 59, 75-107.

Goldstein L. (2003), The emergence of discrete gestures. In *Proceedings of the 15th International Congress of Phonetic Science*, Barcelona, 4-9 Agosto 2003, 85-88.

Gomez R. (2007), Statistical learning in infant language development, in Gareth Gaskell M. (ed.), *The Oxford Handbook of Psycholinguistics*, Oxford (UK): Oxford University Press, 601 – 616.

Pesarin F. (1999), *Permutation Testing of Multidimensional Hypotheses by nonparametric combination of dependent tests*, Padova: Cleup Editrice.

Schwartz, R. e Leonard, L. (1982). Do children pick and choose: An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language*, 9, 319-336. Siegel S. & Castellan N.J. (1992), *Statistica Non Parametrica*, Seconda Edizione, Milano: McGraw-Hill.

Studdert-Kennedy M. & Goodell E. (1995) Gestures, features and segments in early child speech. In *Speech and reading* (B. de Gelder & Morais, editors), Hove: Erlbaum, pp. 65-88.

Studdert-Kennedy M. & Goldstein, L. (2002), Launching language: The gestural origin of discrete infinity. In *Language evolution: The states of the art* (M. Christiansen & S. Kirby, editors), Oxford: Oxford University Press.

Zmarich C. & Bonifacio S. (2004) Gli inventari fonetici dai 18 ai 27 mesi d'età: uno studio longitudinale, in *Atti del Convegno Nazionale "Il Parlato italiano"*, Napoli, 13-15/02/2003.

IDENTIFICAZIONE DEI SEGNALE IN UNA SCENA SONORA

Lorenzo Spreafico
Libera Università di Bolzano
lorenzo.spreafico@unibz.it

Premessa

In molte applicazioni giudiziarie viene richiesto al trascrittore e a volte ad un esperto appositamente convocato se un determinato rumore possa essere attribuito ad una causa ben definita: ad esempio è stato chiesto di accertare se un rumore metallico potesse essere attribuito allo scarrellamento di un'arma, se un rumore impulsivo fosse attribuibile a uno sparo, se un altro rumore, di difficile descrizione, potesse essere attribuito ad un frullatore in funzione.

In letteratura esistono numerosi lavori sul riconoscimento e sulla identificazione dei rumori ma la maggior parte di essi ha come obiettivo lo studio di quali siano le modalità con cui l'uomo identifica, quali le caratteristiche acustiche che caratterizzano un suono e quali le differenze tra ascolto diretto e ascolto del segnale registrato in monofonico o in stereo.

Nelle applicazioni giudiziarie il segnale è sempre registrato e quasi sempre di qualità telefonica; in questo ambito inoltre è importante poter valutare l'attendibilità dell'attribuzione o più precisamente il rapporto di verisimiglianza tra l'attribuzione effettuata ed eventuali attribuzioni alternative. Sarebbe di grande interesse poter disporre di una procedura oggettiva atta a rispondere al quesito ora posto, ma sulla base del segnale acustico normalmente disponibile, segnale di scarsa o pessima qualità, è molto difficile, se non impossibile, trovare tale procedura.

Il presente lavoro ha come obiettivo quello di valutare se le attribuzioni soggettive che di norma vengono effettuate sulla base dell'ascolto di un segnale acustico all'interno di una scena sonora siano attendibili almeno nella misura in cui non si ritenga di voler sospettare ad una volontaria imitazione di un determinato suono.

Esperimento

Come detto obiettivo dell'esperimento è di verificare se, sulla base di un semplice ascolto, le persone siano in grado di determinare la natura di un suono, nel senso di indicare quale sia la sua sorgente. L'identificazione viene effettuata sulla base di un segnale di tipo telefonico o meglio ancora su segnale trasmesso da un telefono cellulare. Gli autori hanno pensato di sottoporre all'ascolto di un gruppo di persone non particolarmente addestrate, un certo numero di suoni particolari. Al fine di predisporre l'esperimento è stato necessario individuare tali suoni che debbono soddisfare le condizioni di essere caratteristici, ossia ben distinguibili, comuni, ossia noti a tutti i soggetti che partecipano al test, e di poter essere in qualche misura presenti in uno scenario di tipo forense.

Si è pensato di mescolare a suoni di origine quotidiana, come l'abbaiare di un cane, il tintinnio di un mazzo di chiavi, il rumore di una serratura che si apre, il rumore di un vetro che si rompe, il rumore di un frullatore, suoni di origine meno comune, come lo sparo di un fucile, lo sparo di una pistola, il rumore di una mitragliatrice, il rumore di un'auto, il rumore di un elicottero, ecc.

Inoltre, per verificare quanto un determinato contesto possa influenzare la decisione sulla natura del suono ascoltato, agli osservatori saranno somministrati due diversi test: in un primo test gli ascoltatori dovranno semplicemente scrivere quale sia la sorgente del rumore da loro ascoltato; nel secondo invece verrà loro fornita un'indicazione circa la scena sonora, ovvero in quale contesto si trova la persona che ascolta quel suono. Entrambi i test saranno preceduti da dei suoni filler dei quali verrà immediatamente dichiarata l'origine al fine di addestrare i soggetti alle modalità di test.

Riferimenti

- Carello, C., J.B. Wagman and M.T. Turvey (2005) Acoustic specification of object properties. In: *Moving Image Theory: Ecological Considerations* (J.D. Anderson and B.F. Anderson, Ed.), pp. 79-104. Southern Illinois University Press, Carbondale, IL.
- Gygi, B., Kidd, G. R., & Watson, C. S., (2004). Spectral-temporal factors in the identification of environmental sounds. *J Acoust Soc Am*, 115, 1252–1265.
- Lawrence, D. M., & Banks, W. P. (1973). Accuracy of recognition memory for common sounds. *Bull Psychon Soc*, 1, 298–300.
- Li, X., Logan, R. J., & Pastore, R. E. (1993). Perception of acoustic source characteristics: walking sounds. *J Acoust Soc Am*, 90, 3036–3049.
- Michael Marcell, Maria Malatanos, Connie Leahy, and Cadie Comeaux *College of Charleston, Charleston, South Carolina* Identifying, rating, and remembering environmental sound events *Behavior Research Methods* 2007, 39 (3), 561-569
- Marcell, M. M., Borella, D., & Greene, J., et al. (2000). Confrontation naming of environmental sounds. *J Clin Exp Neuropsychol*, 22, 830–864.
- R. K. Reddy · V. Ramachandra · N. Kumar · Nandini Chatterjee Singh *Categorization of environmental sounds* Biological Cybernetics, 2009 - Springer
- Shafiro, V., & Gygi, B. (2004). How to select stimuli for environmental sound research and where to find them. *Behav Res Methods Instrum Comput*, 36(4), 590–598.
- Shafiro, V. (2008). Identification of environmental sounds with varying spectral resolution. *Ear & Hear*, 29, 401–420.
- Watson, C. S., & Kidd, G. R. (2002). On the lack of association between basic auditory abilities, speech processing and other cognitive skills. *Semin Hear*, 23, 83–93.
- VanDerveer, N.J. (1979). Ecological acoustics: Human perception of environmental sounds. Doctoral dissertation. Cornell University, Ithaca, NY.

POLIREMATICHE INTERIETTIVE ED ESPRESSIONE DELLE EMOZIONI

Laura Righero e Morena Danieli
Loquendo Voice Technologies, Torino
laura.righero@gmail.com, morena.danieli@loquendo.com

Introduzione

La capacità di riconoscere le emozioni dell'interlocutore e di reagire ad esse in modo appropriato è uno tra gli obiettivi di ricerca più interessanti nel campo dei sistemi di interazione vocale tra persona e macchina. La piena realizzazione di quest'obiettivo richiede non solo l'abilità dell'agente artificiale di capire lo stato emozionale dell'interlocutore, ma anche la generazione di una voce naturale ed espressiva, e infine la consonanza tra lo stile comunicativo del sistema e quello dell'utente. Proprio alla luce di tali considerazioni, molto lavoro è stato svolto per migliorare tutti questi aspetti ed in particolare la naturalezza della produzione sintetica della voce. In questo campo una grande attenzione è tradizionalmente dedicata alle tematiche connesse al ritmo e all'intonazione della voce, mentre gli aspetti relativi alle scelte linguistiche dell'agente hanno ricevuto finora un'attenzione minore.

Lo scopo di questo lavoro è contribuire allo studio di fenomeni collocati all'interfaccia tra scelta lessicale e realizzazione fonetica. Ci siamo proposte di osservare in che misura la manifestazione dell'esperienza emotiva influenzi la scelta lessicale. Nella lingua, infatti, una parte del lessico è spesso definita "lessico emotivo": con tale dizione si intendono le parole che, nella loro rappresentazione semantica, contengono un'indicazione dello stato emotivo del parlante. Il lessico di una lingua, nel nostro caso il lessico della lingua italiana, comprende un certo numero di parole che esprimono stati emotivi: ad esempio nomi (rabbia, paura, gratitudine), verbi (ammiro, mi vergogno, mi arrabbio), aggettivi (felice, allegro, triste), avverbi (con piacere, tristemente, con gratitudine), interiezioni (òoh, perbacco, mamma mia). Nell'esaminare la letteratura linguistica sul lessico emotivo si è scelto di concentrare l'attenzione sulle interiezioni composte da almeno due unità grafiche, come ad esempio mamma mia!, altro che!, spesso definite polirematiche interietive (Voghera 2004). Tali espressioni sono state oggetto di un certo numero di studi sia linguistici che di matrice psicolinguistica (De Mauro & Voghera 1996; Poggi 1995; tra gli altri). Lo sfondo comune di queste ricerche è costituito dall'attenzione dedicata all'indagine dei processi di formazione delle parole, alla struttura del lessico delle lingue naturali e ai meccanismi cognitivi alla base della scelta e dell'inserzione lessicale nella produzione linguistica orale. In particolare si è visto che le polirematiche interietive costituiscono un fenomeno linguistico complesso e variegato, collocabile in un'area del lessico mentale posta idealmente ai confini tra morfologia e sintassi, pur condividendo alcune proprietà con le forme idiomatiche, quali l'alta frequenza d'uso nella produzione orale spontanea, alcune proprietà distribuzionali, la funzione di marcatori pragmatici del discorso e, non ultima, la loro probabile rappresentazione mentale in termini di unità lessicalizzate.

L'esame di queste caratteristiche ci ha condotto a ritenere che nello studio di queste espressioni linguistiche sia utile integrare dati provenienti dall'analisi di corpora di parlato con quelli ricavabili dalle intuizioni linguistiche dei parlanti, indagate secondo le metodologie della ricerca psicolinguistica. Per questo ci siamo proposte di verificare in che misura un insieme di 60 polirematiche interietive, ricavate dallo spoglio di corpora di

parlato, siano riconosciute e utilizzate dai parlanti per comunicare, nel discorso, una propria emozione. Nel seguito forniamo gli elementi essenziali della metodologia e dei risultati della nostra ricerca.

Descrizione dell'esperimento

Per reperire l'insieme di polirematiche interiettive studiate nel nostro esperimento abbiamo analizzato due corpora di trascrizioni di italiano parlato: il corpus di parlato conversazionale del LABLITA (Cresti 2000) e un corpus di trascrizioni di interviste di commento a fotografie raccolto per una tesi di dottorato (Senn 2006). Il materiale reperito da questo spoglio è stato ampliato ricorrendo alla consultazione del LIP, del GRADIT e, limitatamente, all'introspezione. Si è così individuato un elenco di 60 polirematiche interiettive (Alla faccia! Neanche per idea!) che, secondo la nostra ipotesi di partenza, erano particolarmente adatte ad essere scelte dai parlanti per esprimere il loro stato emozionale. In base alla nostra ipotesi, un campione opportunamente selezionato di parlanti avrebbe dovuto preferire, per esprimere stati emozionali, questo insieme di espressioni rispetto ad un insieme di sintagmi, tratti dall'analisi degli stessi corpora, che però non presentavano caratteristiche adatte all'espressione dello stato emotivo (ad esempio: aula uno, sabato sera, vari negozi). Si è quindi scelto di mescolare l'insieme di interiezioni con un numero corrispondente di sintagmi di vario tipo e pareggiati per lunghezza alle locuzioni interiettive, sottoponendo così al giudizio del campione di parlanti un insieme di 120 espressioni, costituite per metà da polirematiche interiettive e per metà da sintagmi. Il giudizio dei parlanti è stato raccolto sottoponendo loro un questionario in cui si chiedeva di indicare, su una scala di giudizio tipo-Likert a sette valori, il loro accordo o disaccordo ad utilizzare l'espressione proposta per comunicare, nel parlato, una propria emozione.

Il questionario è stato somministrato a 120 soggetti, ovvero a 60 uomini e a 60 donne, tutti di madrelingua italiana, suddivisi per diverse fasce di età (dai 18 ai 65 anni) e livelli di scolarizzazione (licenza media, diploma, laurea). Il questionario è stato strutturato in modo tale da somministrare a tutti i soggetti lo stesso numero e lo stesso tipo di stimoli. La compilazione è avvenuta in modo individuale e in forma anonima.

Analisi dei dati sperimentali e discussione dei risultati

Le statistiche descrittive, costituite dalle medie dei punteggi che i soggetti hanno attribuito al gruppo delle polirematiche e a quello dei sintagmi, rivelano che l'insieme delle interiezioni ha conseguito dei valori mediamente più alti rispetto a quelli ottenuti dai sintagmi. Da un punto di vista linguistico, la differenza osservata suggerisce che i soggetti si sono dimostrati d'accordo a riconoscere e a scegliere le polirematiche interiettive come voci lessicali in grado di comunicare, nel parlato, un'emozione.

Per quanto concerne l'analisi relativa al solo gruppo delle interiezioni, la differenza tra le medie delle singole polirematiche non è così saliente. È verosimile ritenere che tale risultato possa essere una conseguenza della natura di questo tipo di esperimento, in cui la necessità di limitare il numero di variabili oggetto di analisi sperimentale ha richiesto di escludere riferimenti al contesto linguistico ed extralinguistico. Tale mancanza può aver indotto i soggetti a riconoscere le interiezioni come espressioni "generalmente emozionali", ma non ha dato la possibilità di individuare una scalarità significativa all'interno di tale raggruppamento.

Abbiamo verificato se la differenza tra i due insiemi di espressioni, osservata a livello descrittivo, fosse significativa. Il risultato dell'applicazione del test T di Student alle medie dei punteggi ottenuti ($t=8.2978$, $df=119$, $p\text{-value}=0.000243$) non ha consentito di poter affermare che sia probabile, in termini statisticamente significativi, che la differenza delle medie delle polirematiche e dei sintagmi non sia dovuta al caso o ad altre variabili che non è stato possibile includere nella ricerca. Riteniamo, come già osservato poco sopra, che il ruolo del contesto di enunciazione, sia linguistico che extralinguistico, e dei suoi effetti sulla decisione lessicale del parlante siano determinanti. Inoltre è ipotizzabile che la modalità di presentazione degli item lessicali in forma scritta abbia contribuito ad attenuare le differenze tra sintagmi e polirematiche interietive: riteniamo verosimile che la scelta lessicale per questa classe di espressioni non possa essere disgiunta dalla realizzazione prosodica. Infatti pensiamo che i risultati di questo esperimento costituiscano un'evidenza indiretta della forza della connessione tra significazione e realizzazione prosodica nei contesti enunciativi reali.

Riferimenti bibliografici

Albano Leoni F. (2009), *Dei Suoni e dei Sensi. Il volto fonico delle parole*, Bologna, Il Mulino

Bazzanella C. (1995), "I segnali discorsivi", in Renzi L., Salvi G., Cardinaletti A. (Eds), *Grande Grammatica Italiana di Consultazione*, vol. III, Bologna, Il Mulino, pp. 225-257

Cresti E. (Ed.), (2000), *Corpus di Italiano Parlato*, vol. I-II, Firenze, Accademia della Crusca

Danieli M. (2008), "Verso una sintesi da testo espressiva", in Magno Caldognetto E., Cavicchio F. (Eds), *Aspetti Emotivi e Relazionali dell'E-learning*, Firenze, Florence University Press, pp. 113-126

De Mauro T., Mancini F., Vedovelli M., Voghera M. (Eds), (1993), *Lessico di Frequenza dell'Italiano Parlato (LIP)*, Milano, Etaslibri

De Mauro T., Voghera M. (1996), "Scala mobile: un punto di vista sui lessemi complessi", in Benincà P., Cinque G., De Mauro T., Vincent N. (Eds), *Italiano e Dialetti nel Tempo: Saggi di Grammatica per Giulio C. Lepschy*, Roma, Bulzoni, pp. 99-131

De Mauro T. (1999), *Grande Dizionario Italiano dell'Uso (GRADIT)*, Torino, UTET

Poggi I. (1981), *Le Interiezioni: Studio del Linguaggio e Analisi della Mente*, Torino, Editore Boringhieri

Poggi I. (1994), "Le interiezioni: le voci della mente", in Salza P. L. (Ed.), *Atti delle IV Giornate di Studio del G.F.S. (A.I.A.)*, Torino, 11-12 novembre 1993, Roma, Esagrafica, pp. 155-161

Poggi I. (1995), "Le interiezioni", in Renzi L. (Ed.), *Grande Grammatica Italiana di Consultazione*, vol. III, Bologna, Il Mulino, pp. 403-425

Poggi I., Caldognetto E. M. (2004), "Il parlato emotivo. Aspetti cognitivi, linguistici e fonetici", in Albano Leoni F., Cutugno F., Pettorino M., Savy R. (Eds), *Atti del convegno "Italiano parlato"*, Napoli, D'Auria Editore, CD-Rom

Senn M. (2006), *Venise, une Évidence Symbolique*, http://venezia.objectis.net/dea_html

Voghera M. (1994), "Lessemi complessi: percorsi di lessicalizzazione", in *Lingua e Stile*, a. XXIX, n. 2, pp. 185-214

Voghera M. (2004), "Polirematiche", in Grossman M., Rainer F. (Eds), *La Formazione delle Parole in Italiano*, Niemeyer, Tübingen, pp. 56-69

A SYLLABIFICATION ALGORITHM USING ENERGY, PITCH AND HARMONICITY RELATED FEATURES

Antonio Origlia, Giovanni Abete

Natural Language Processing Group - Università di Napoli "Federico II"

antori@gmail.com, giovanni.abete@libero.it

Researching speech often leads to facing an important issue: giving speech a reliable basic unit to build higher level representations of the speech signal. This is mainly needed to investigate human capability to convey through speech a number of informations going beyond what is actually being said. Emotional informations, questioning/sentencing attitude, intentional attention drawing on specific dialogue parts are important examples of the subtleties humans are capable to perform when producing speech. These problems, however, being intrinsically elusive and difficult to describe, force researchers to look for easy to use and yet powerful building blocks to assemble complex speech representations with. Describing and exploiting features that these representations can bring in the foreground is the key to solve many of the issues speech research still presents.

Even though phonemes have been proposed to cover this important role, syllables are much more intuitive and easy to deal with at conceptual level (Fugimura 1975; Greenberg & Kingsbury 1997; Segui et al. 1990). Giving rhythm and structure to speech, they are a very reliable and solid ground to build on. Through syllables, the segmental level of a speech signal is easy to describe and investigate. Higher layers can be reached and studied by building on this pragmatic and tangible starting point.

We will refer to syllables as a basic rhythmical unit as it is perceived by the human ear so grammatical meanings of the term are to be discarded. This is because it is not always true that a speaker accurately pronounces all the grammatically expected syllables in a word. Context and speech rate can influence a speaker's fluency or articulation accuracy thus introducing relevant differences between what should have been pronounced and what has actually been pronounced (Lindblom 1990). Detecting syllables can be a difficult task: firstly it is necessary to find their nucleus and next, boundaries are to be set (De Jong & Wempe 2009; Jittiwarakul et al. 1998; Mermelstein 1975; Nagarajan et al. 2003; Pfizinger et al. 1996). It is well known that informations needed to perform these two steps are mainly held by the signal's energy profile: syllable nuclei are mostly signaled by peaks in the energy profile while boundaries are found in valleys preceding syllable nuclei. This, however, is not sufficient to perform good syllabification.

Concerning syllable nuclei detection, we know that peaks in the energy profile can be produced by a number of different causes other than syllable nucleus occurring like background noise, breathing and tongue clicking. Plus, speech itself can produce artifact peaks that are not to be considered syllable nuclei. In this work, specialized procedures are employed to detect and remove artifacts that are easy to describe: in particular we observed that small peaks showing on high slope rising regions of the energy profile are not to be considered syllable nuclei. Pitch based voice activity detection and harmonicity analysis are used to further narrow syllable nuclei candidates. Harmonicity (Boersma 1993) also helps to recover syllable nuclei that happen not to have a particularly strong energy peak and to disambiguate some peculiarities. These have the shape of small peaks showing on high slope descending regions of the energy profile. We observed that these situations, as opposed to their counterparts showing on high slope rising regions of the energy profile, are not always artifacts and can play the role of a syllable nucleus or act as a reference point to

set syllable boundaries. Concerning the task of setting syllable boundaries, consonants, especially fricatives, can influence the behaviour of the energy profile causing boundaries to shift from valleys to adjacent regions. To deal with this problem, we filter the signal cutting off the frequencies where fricative sounds lie and compare this filtered signal with the original one. Whenever a strong difference between the two signals is detected, the syllable boundary is shifted from its original position to the left. When shifting, the interval between the valley and its preceding nucleus is inspected and the syllable boundary is placed where the first derivative's value reaches its minimum.

To evaluate the accuracy of our algorithm we employed a system based on Petek et al. (1996). We tested the automatic segmentation against a manual one, performed by human experts, and achieved promising results. We hope to further develop this approach to obtain a strong, reliable segmental representation of the speech signal that can be used for a number of tasks ranging from speech recognition to prosody conveyed informations extraction.

References

- Boersma, P. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Institute of Phonetic Sciences, University of Amsterdam, *Proceedings* 17, 97-110.
- De Jong, N. & Wempe, T. 2009. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, 41(2), 385-390.
- Fujimura, O. 1975. Syllable as a unit of speech recognition. In *IEEE Trans. Acoust. Speech Signal Process*, 23 (1), 82-87.
- Greenberg, S. & Kingsbury, B. E. 1997. The modulation spectrogram: in pursuit of an invariant representation of speech. In *Proceedings of ICASSP '97*. Munich, Germany, 1647-1650.
- Jittiwarangkul, N., Jitapunkul, S., Luksaneeyanavin, S., Ahkuputra, V. & Wutuwiwatchai, C. 1998. Thai syllable segmentation for connected speech based on energy. In *Proceedings of the IEEE Asia Pacific Conference on Circuits and Systems (APCCA'98)*. Chiangmai, Thailand, 169-172.
- Lindblom, B. 1990. Exploring phonetic variation: a sketch of the H&H theory. In Hardcastle, W. J. & Marchall, A. (eds), *Speech production and speech modelling*. Dordrecht, Kluwer Academic Publications, 403-439.
- Mermelstein, P. 1975. Automatic segmentation of speech into syllabic units. *Journal of Acoustical Society of America* 54, 4, 880-883.
- Nagarajan, T., Murthy, H. A. & Hedge, R. M. 2003. Segmentation of speech into syllable-like units. In *Eurospeech 2003*. Geneva, 2893-2896.
- Petek, B., Andersen, O. & Dalsgaard, P. 1996. On the robust automatic segmentation of spontaneous speech. In *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP)*. Philadelphia, 913-916.
- Pfizinger, H., Burger, S. and Heid, S. 1996. Syllable detection in read and spontaneous speech. In *Proceedings of the Fourth International Conference on Spoken Language (ICSLP)*. Philadelphia, 1261-1264.
- Segui, J., Dupoux, E. & Mehler, J. 1990. The role of the syllable in speech segmentation, phoneme identification and lexical access. In Altmann, G. T. M. (ed.), *Cognitive models of speech processing: psycholinguistic and computational perspectives*. Cambridge (MA), MIT Press, 263-280.

DAL *BABBLING* ALLA LINGUA MATERNA. TENDENZE UNIVERSALI E ASPETTI LINGUO-SPECIFICI NELLO SVILUPPO FONETICO DI DUE BAMBINI ITALIANI

Sara Giulivi¹; Mario Vayra² Claudio Zmarich³

¹ Dipartimento di Linguistica, Università di Firenze, ² Dipartimento di Studi Linguistici e Orientali, Università di Bologna, ³ Istituto di Scienze e Tecnologie della Cognizione del C.N.R., Sede di Padova,

sara.giulivi@unibo.it, claudio.zmarich@pd.istc.cnr.it, mario.vayra@unibo.it

Esiste in letteratura un sostanziale accordo sul rapporto di continuità tra *babbling* e linguaggio adulto (numerosi sono infatti le caratteristiche fonetiche e fonotattiche comuni), e sul fatto che molti aspetti fonetici del *babbling* e delle prime parole abbiano carattere universale. Con il termine *babbling* si intende la produzione di sequenze di sillabe di tipo consonante-vocale (C-V) dotate di un'organizzazione temporale e ritmica simile a quella del parlato adulto e caratterizzate da transizioni formantiche relativamente rapide tra elementi consonantici e vocalici (Vihman, 1996).

La ricerca sul campo ha mostrato che le sillabe dei primi stadi di *babbling* si manifestano generalmente con un alto grado di coarticolazione, osservano un principio di minimo sforzo, e sono influenzate da un vincolo biomeccanico che lega al movimento ritmico di tipo oscillatorio della mandibola, organo articolatore già molto mobile, la postura di labbra e lingua, articolatori ancora poco mobili. Secondo MacNeilage (1998), proprio il movimento di elevazione e abbassamento della mandibola, legato in ontogenesi e in filogenesi all'attività neuromuscolare della nutrizione, è responsabile del fatto che, nelle prime fasi dello sviluppo linguistico, le vocalizzazioni del bambino presentino una preferenza per determinate combinazioni consonante-vocale piuttosto che per altre. In particolare, le consonanti labiali tendono ad associarsi più frequentemente con vocali centrali, le consonanti alveolari con vocali anteriori e le consonanti velari con vocali posteriori (cfr. Davis & MacNeilage, 1995). La *Articulatory Phonology* di Browman e Goldstein (Goldstein, Byrd & Saltzman, 2006) offre di questi dati una diversa interpretazione. Essa sostiene infatti che certe restrizioni di tipo neurofisiologico non siano sufficienti a rendere conto dei fenomeni individuati da MacNeilage come caratterizzanti le prime fasi dello sviluppo linguistico. Secondo la *Articulatory Phonology*, il parlato può essere descritto in termini di *gesti articolatori*, unità minime che catturano proprietà sia fisiche (fonetiche), sia cognitive (fonologiche) del segnale vocale e che tendono a coordinarsi tra loro secondo modalità stabili proprie dei sistemi dinamici. In questa prospettiva, fatti fonetico-fonologici come la preferenza per certe combinazioni consonante-vocale (ad es. [ba], [di], [gu]) o per determinati tipi sillabici (ad esempio CV), sarebbero riconducibili ad un principio di 'facilità' articolatoria. Tali combinazioni C-V e tali tipi sillabici, infatti, richiedono l'attivazione di gesti che non devono 'competere' tra loro per essere prodotti. Essi possono pertanto essere sincronizzati per quanto concerne il momento dell'attacco. Ciò darebbe origine ad una massima stabilità articolatoria, ottenuta con il minimo dispendio di energia.

Quale che sia l'interpretazione prescelta di questo insieme di fatti, è indubbio che certe caratteristiche fonetiche e fonotattiche del *babbling* presentano carattere universale; e ciò nonostante l'elevata sensibilità percettiva alla lingua materna di cui il bambino è già in possesso in questa prima fase dello sviluppo linguistico (cfr. Curtin & Werker, 2007). I tempi e le modalità con cui inizia nel bambino il processo di progressiva sintonizzazione

sul sistema fonetico-fonologico della lingua materna costituiscono invece un aspetto ancora aperto alla verifica sperimentale. Secondo Vihman & de Boysson-Bardies (1994), la lingua materna esercita un'influenza 'positiva' sulle produzioni linguistiche del bambino intorno ai 9-10 mesi di vita. In questa fase si assiste infatti ad un incremento dei suoni nativi. Intorno ai 12 mesi, invece, l'influenza della lingua materna è di tipo 'negativo', in quanto porta ad una diminuzione dei foni non nativi. Per quanto riguarda l'italiano, gli studi finora disponibili sulle caratteristiche fonetiche del *babbling* e delle prime parole analizzano dati raccolti a partire dai 10 mesi (cfr. Zmarich & Ferrero, 1999; Zmarich & Miotti, 2003; Zmarich, Stocco, Minozzi & Bonifacio, 2005), età in cui la soglia di inizio delle influenze linguo-specifiche potrebbe già esser stata superata (cfr. de Boysson-Bardies, Sagart & Durand, 1984; de Boysson-Bardies, Hallé, Sagart & Durand, 1989).

Il presente lavoro intende indagare l'emergere dell'influenza linguo-specifica entro un arco temporale che va dai 6 ai 18 mesi di vita, cioè dall'emergere nel bambino delle prime produzioni di *babbling* fino al raggiungimento di un vocabolario di circa 50 parole. I soggetti presi in esame sono due bambini italiani, un maschio e una femmina, audioregistrati e trascritti a intervalli regolari di due mesi.

Lo studio consiste in un'analisi segmentale delle produzioni dei soggetti, preliminarmente distinte in produzioni classificabili come *babbling*, e in produzioni classificabili come 'parole', sulla base del criterio proposto in Vihman & McCune (1994), secondo cui possono essere definite 'parole' le associazioni stabili di forma e significato.

I risultati sono stati analizzati per stadi di età cronologica, in particolare a 6, 8, 10, 12, 14, 16 e 18 mesi.

Su tutte le produzioni sono state effettuate alcune analisi statistiche di tipo descrittivo, allo scopo di rendere conto della frequenza di:

- i) suoni consonantici;
- ii) suoni vocalici;
- iii) gruppi consonantici;
- iv) classi fonologiche di modo, luogo e sonorità consonantica;
- v) classi fonologiche di luogo e altezza vocalica;
- vi) tipi sillabici;
- vii) co-occorrenze di consonanti e vocali all'interno di ciascuna sillaba.

I calcoli frequenziali del presente studio si sono basati sulla trascrizione fonetica del materiale audio secondo l'IPA (*International Phonetic Alphabet*). Nonostante sia ancora poco chiaro, allo stato attuale delle ricerche, se le vocalizzazioni prelinguistiche del bambino possano considerarsi effettivamente organizzate in termini di "segmenti", e fino a che punto i suoni prodotti dal bambino possano essere interpretati in modo oggettivo attraverso le categorie segmentali dell'adulto, le trascrizioni delle produzioni di *babbling* in simboli IPA sono da ritenersi affidabili. Infatti, la maturità anatomofisiologica dell'apparato pneumo-fono-articolatorio che il bambino raggiunge intorno ai 6-7 mesi di vita (Kent & Vorperian, 1995), età d'inizio del *babbling*, e l'emergere della struttura sillabica, facilmente riconoscibile come tale dall'adulto, fanno sì che i suoni consonantici e vocalici possano essere descritti in modo affidabile secondo i parametri fonetici di modo, luogo e sonorità, incorporati nei simboli IPA (cfr. *The International Phonetic Association*, 1999).

Nel presente lavoro, i tempi e le modalità dell'emergere dell'influenza linguo-specifica sono stati esaminati alla luce dei risultati ottenuti per l'italiano in studi di frequenza sulle occorrenze di parlato spontaneo adulto. In particolare, per quanto riguarda i tipi sillabici, le produzioni dei soggetti presi in esame sono state confrontate con i dati riportati in Mancini e Voghera (1994); per quanto riguarda l'analisi segmentale, i risultati del presente studio

sono stati comparati con quelli presentati in Bortolini, Degan, Minnaja, Paccagnella & Zilli (1978). Un confronto è stato fatto inoltre con i risultati ottenuti da uno studio condotto su un gruppo di bambini di età compresa tra i 10 e i 27 mesi, riportato in Zmarich *et al.* (2005).

Riferimenti bibliografici

- Boysson-Bardies de, B., Vihman, M. M., Roug-Hellichius, L., Durand, C.; Landberg, I., Arao, F., (1992), *Material evidence of infant selection from target language: A crosslinguistic phonetic study*. In C. A. Ferguson, L. Menn, C. Stoel-Gammon (eds.), *Phonological Development. Models, Research, Implications*, Timonium, York Press, 369-391.
- Bortolini, U., Degan, F., Minnaja, C., Paccagnella, L. & Zilli, G., (1978), Statistics for a Stochastic Model of Spoken Italian, in *Proceedings of the 12th International Congress of Linguists*, 580-586.
- Curtin S. & Werker J.F., (2007), *The perceptual foundations of phonological development*. In M. G. Gaskell (ed.), *The Oxford Handbook of Psycholinguistics*, Oxford, Oxford University Press, 579-599.
- Davis B.L., MacNeilage P.F. (1995), The articulatory basis of babbling, *Journal of Speech Hearing Research*, 38, 1199-1211.
- Goldstein L.M, Byrd, D. & Saltzman E, (2006), The role of vocal tract gestural action units in understanding the evolution of phonology. In M. Arbib (eds.), *Action to language via the mirror neuron system*, Cambridge, Cambridge University Press, 215-249.
- MacNeilage P.F. (1998), The frame/content theory of evolution of speech production, *Behavioral and Brain Sciences*, 21, 499-546.
- Kent, R.D. & Vorperian, H. K. (1995), *Development of the craniofacial-oral-laryngeal anatomy*. San Diego, Singular Publishing Group.
- Mancini, F. & Voghera, M., (1994), Lunghezza, tipi di sillabe e accento in italiano. In T. De Mauro (a c. d.), *Come parlano gli italiani*. Firenze, La Nuova Italia.
- The International Phonetic Association, (1999), *Handbook of the International Phonetic Association*, Cambridge, Cambridge University Press.
- Vihman, M. M., 1996. *Phonological development*. Oxford, Blackwell Publishers.
- Vihman, M. M. & de Boysson-Bardies B., (1994), The Nature and Origin of Ambient Language Influence on Infant Vocal Production and Early Words, *Phonetica*, 51, 159-169.
- Vihman & McCune (1994), When is a word a word?, *Journal of Child Language*, 21, 61- 82.
- Zmarich C. & Ferrero F., (1999), A phonetic and acoustic study on the early speech development of two Italian children, in J. Ohala *et al.* (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences*, 3, 2165-2168.
- Zmarich C., Miotti R. (2003), The frequency of consonants and vowels and their cooccurrences in the babbling and early speech Italian children, in Solé N. J. *et al.*, *Proceedings of the 15th International Congress of Phonetic Sciences*, 2, 1947-1950.
- Zmarich C., Stocco D., Minozzi M., Bonifacio S., (2005), La frequenza di occorrenza di consonanti e vocali e delle loro combinazioni nella sillabe del babbling e delle prime parole dai 10 ai 27 mesi di età, in P. Così (a c. d.), *Misura dei parametri, Atti del 1° Convegno Nazionale dell'Associazione Italiana di Scienze della Voce*, EDK Editore, Brescia, 2005, 481-510.

LA NEUTRALIZZAZIONE DEL CONTRASTO DI LUNGHEZZA DELLE FRICATIVE NEL DIALETTO DI ZURIGO

Stephan Schmid
Università di Zurigo
schmidst@pholab.unizh.ch

Una delle differenze più salienti tra il tedesco standard e i dialetti alemanni della Svizzera riguarda la correlazione di sonorità. Nel tedesco standard, il tratto [+sonoro] può essere implementato articolatoriamente per tutte le ostruenti attraverso la vibrazione delle corde vocaliche: ciò vale almeno per il contesto intervocalico, mentre nella coda sillabica si applica il noto processo della *Auslautverhärtung* che porta alla desonorizzazione delle ostruenti sonore; all'inizio di parola vigono condizioni fonologiche e fonetiche specifiche per i due modi di articolazione delle occlusive e delle fricative (v. De Dominicis 1999).

A differenza del tedesco standard, i dialetti della Svizzera germanofona sono caratterizzati dall'assenza categorica di ostruenti sonore. Ciononostante esiste un contrasto fonologico tra due serie di occlusive e fricative che viene descritto – secondo una tradizione che risale a Winteler (1876) – in termini di forza articolatoria, ovvero come opposizione tra consonanti ‘forti’ e ‘leni’ (cfr. Fleischer & Schmid 2006). Sul piano acustico, come correlato principale di tale opposizione è stata individuata la durata (cfr. Willi 1996, Nocchi & Schmid 2006); tant'è vero che il contrasto fra ostruenti brevi e lunghe è stato reinterpretato anche come opposizione tra consonanti scempie e geminate (Kraehenmann 2003).

Un'altra differenza rispetto al tedesco standard riguarda i contesti di neutralizzazione. Le analisi di Kraehenmann (2003) mostrano infatti come nel dialetto del canton Turgovia l'opposizione tra consonanti brevi e lunghe resta intatti anche in posizione finale. La letteratura dialettologica indica invece altri due contesti fonotattici dove il contrasto tra ostruenti forti e leni viene neutralizzato. Secondo la ‘legge di Heusler’ (1888), nelle sequenze di due ostruenti (non importa se esse si trovino al confine di parola o in posizione interna) le consonanti verrebbero realizzate con una “intensità intermedia”, in questo caso si parla quindi anche di consonanti ‘semi-forti’ (Moulton 1986). La stessa neutralizzazione colpirebbe le occlusive in posizione iniziale assoluta, contesto in cui le fricative conservano invece il loro carattere di leni (Dieth 1950: 177); all'inizio di parola le fricative forti sono comunque escluse da una restrizione fonotattica.

La presente ricerca è infatti limitata alla categoria delle fricative e intende verificare tre ipotesi, ovvero 1) la mancanza di neutralizzazione del contrasto tra forti e leni in posizione prepausale, 2) l'effettiva neutralizzazione dello stesso contrasto in seguito al contatto tra due ostruenti, 3) la non-rilevanza della quantità vocalica per la lunghezza della consonante successiva (la quantità vocalica ha valore distintivo nello zurighese; cfr. Schmid 2004). A tale scopo è stato analizzato un *corpus* di parlato di laboratorio contenente 270 realizzazioni. Tre locutrici, parlanti native del dialetto di Zurigo, hanno letto due serie di frasi cornice, in cui erano inserite 15 parole terminanti in fricativa; nella prima serie la fricativa in questione occorreva in posizione prepausale, nella seconda in posizione preconsonantica. Ogni frase cornice è stata ripetuta tre volte.

Le misurazioni delle durate, effettuate manualmente mediante *Praat*, mostrano una completa neutralizzazione del contrasto tra fricative forti e leni nel contesto preconsonantico: dopo vocale breve le durate medie sono di 117 ms per le forti e di 113 ms per

le leni, dopo vocale lunga le durate medie sono invece di 90 ms per le forti e di 91 ms per le leni. Viene quindi confermata pienamente l'ipotesi 2), mentre l'ipotesi 3) è confermata solo parzialmente; si verifica in effetti un lieve effetto di allungamento compensatorio dopo le vocali brevi. In fine enunciato, com'è da aspettarsi in base al processo universale di allungamento prepausale, le durate delle fricative mostrano valori più elevati: dopo vocale breve le medie sono di 232 ms per le forti e di 188 ms per le leni, dopo vocale lunga le medie sono invece di 199 ms per le forti e di 159 ms per le leni. Viene quindi riconfermata anche per il contesto prepausale la lieve compensazione intrasillabica tra vocale e consonante, per cui la terza ipotesi non può essere sostenuta in modo categorico.

Inoltre, esiste sì una differenza di durata tra forti e leni in posizione di fine enunciato, ma tale differenza si rivela essere piuttosto ridotta (232 ms vs. 188 ms e 199 ms vs. 159 ms). Allo stato attuale emerge quindi la necessità di un ulteriore confronto con il contesto intervocalico; se tra due vocali la differenza di durata tra forti e leni si dovesse mostrare più elevata (e questa infatti è l'ipotesi), allora bisognerebbe concludere che in posizione prepausale le fricative sono sottoposte ad un processo di neutralizzazione 'parziale'.

Riferimenti bibliografici

- De Dominicis, A. (1999). *Fonologia comparata delle principali lingue europee moderne*, Bologna: CLUEB.
- Dieth, E. (1950). *Vademekum der Phonetik*, Bern: Francke.
- Fleischer, J. & Schmid, S. (2006). Zurich German, *Journal of the International Phonetic Association*, 36, 243-253.
- Heusler, A. (1888). *Der alemannische Consonantismus in der Mundart von Baselstadt*. Strassburg: Trübner.
- Kraehenmann, A. (2003). *Quantity and prosodic asymmetries in Alemannic. Synchronic and diachronic perspectives*, Berlin: Mouton de Gruyter.
- Moulton, W. (1986). Sandhi in Swiss German dialects, in *Sandhi phenomena in the languages of Europe* (H. Andersen, editor), Berlin/New York: de Gruyter, 385-392.
- Nocchi, N. & Schmid, S. (2006). Le consonanti labiodentali dello svizzero tedesco, in *Analisi prosodica. Teorie, modelli e sistemi di annotazione. Atti del 2° Convegno nazionale AISV* (R. Savy & C. Crocco, editors), Torriana: EDK editore, 528-537.
- Schmid, S. (2004). Zur Vokalquantität in der Mundart der Stadt Zürich, *Linguistik online*, 20, 93-116. http://www.linguistik-online.de/20_04/
- Willi, U. (1996). *Die segmentale Dauer als phonetischer Parameter von 'fortis' und 'lenis' bei Plosiven im Zürichdeutschen. Eine akustische und perzeptorische Untersuchung*, Stuttgart: Steiner.
- Winteler, J. (1876). *Die Kerenzer Mundart des Kantons Glarus in ihren Grundzügen dargestellt*, Leipzig: Winter.

NAVIGAIS - AIS DIGITAL ATLAS AND NAVIGATION SOFTWARE

Graziano Tisato

ISTC - Institute of Cognitive Sciences and Technologies – CNR

Via Martiri della Libertà 2 – Padova – Italy

tisato@pd.istc.cnr.it

Notwithstanding the venerable old age, the Sprach- und Sachatlas Italiens und der Südschweiz (AIS) (Linguistic and Ethnographic Atlas of Italy and Southern Switzerland), is still one of the most useful instrument in the linguistic geography, and in the study of Romance and Italian dialectology. Moreover, the atlas turns out to be an inexhaustible mine of ethnographic information.

The AIS was conceived by the Swiss linguists Karl Jaberg and Jakob Jud, and published in eight volumes between 1928 and 1940, with an introductory book, and followed only in 1960 by an index, rather difficult to be found today. On the field, the data collection was carried out, between 1919 and 1925, in 306 places of southern Switzerland, and north and central Italy, by Paul Scheuermeier, in 81 places of southern Italy by Gerhard Rohlf, and in 20 Sardinian places by Max Leopold Wagner. In the following years Scheuermeier did continue the work in north Italy in different stages until 1935, to improve the AIS ethnographic aspects.

There are some good reasons to want a digital acquisition and navigation of AIS. First of all, the obvious necessity to exploit adequately the enormous quantity of information contained in the atlas. The index volume published in 1960 doesn't allow to search a lemma, or part of a lemma, in a convenient way. Indeed, for the normal time and space publication limits, it could list only some prototypal dialectal forms, but not all the 1 M words present in the AIS. Moreover the AIS maps suffer from a lack of legibility, because the words are only indexed with an identifier number without any place name. The retrieval of a single lemma in 8 volumes and 1705 pages could require an eternity in comparison with a digital search engine.

The paper presents NavigAIS, the first high resolution digital version of the AIS. The digital atlas is provided with a search and navigation software, indispensable to explore in a quick and comfortable way the 1705 maps contained in the atlas.

The software, written in Matlab®, was planned as a stand alone application, but also as a preliminary tool for a current project, aiming to acquire all the about 1 M dialectal words in the atlas.

The NavigAIS package will be shortly downloadable.

References

- [1] Jaberg, K. and Jud, J., “Sprach- und Sachatlas Italiens und der Südschweiz”, Vol.1-8, Zofingen, Bern, 1928-1940. [Reprint edition by: Kraus Reprint, Nendeln, Liechtenstein, 1971-1981. Kraus Reprint, New York].
- [2] Jaberg, K. and Jud, J., “Der Sprachatlas als Forschungsinstrument. Kritische Grundlegung und Einführung in den Sprach- und Sachatlas Italiens und der Südschweiz”, Halle, 1928. [It. translation: “Ais - Atlante Linguistico ed Etnografico dell'Italia e della

Svizzera Meridionale. Vol. 1: Fondamenti Critici e Introduzione. Vol. 2: Scelta di Carte Commentate”, Unicopli, Milano, 1988.

[3] Jabert, K., “Index zum Sprach- und Sachatlas Italiens und der Südschweiz: Ein propädeutisches etymologisches Wörterterbuch der italienischen Mundarten”, Bern, Stämpfli, 1960.

[4] Scheuermeier, P., “Bauernwerk in Italien, der italienischen und rätoromanischen Schweiz”, Vol. 1-2, Erlenbach-Zürich, Eugen Rentsch Verlag, 1943, [It. Translation: Scheuermeier P., “Il lavoro dei contadini”, Longanesi, Milano, 1980.

‘MIXED MALTESE ENGLISH’: FENOMENI DI INTERFERENZA FONETICA LUNGO IL *CONTINUUM* MALTESE – INGLESE NEL REPERTORIO LINGUISTICO DI MALTA

Laura Mori
Università della Tuscia, Viterbo
laura.mori@luspicio.it

Nel 1964 la Costituzione della *Repubblika ta' Malta* riconobbe maltese e inglese come lingue ufficiali; un regime bilingue che formalmente avrebbe dovuto determinare le condizioni per il passaggio da una secolare condizione diglossica ad una di bilinguismo sociale. Tuttavia ancora oggi, la situazione sociolinguistica di Malta non prevede una eguale distribuzione delle due lingue, inglese e maltese, come codici interscambiabili nella prassi comunicativa in qualsiasi dominio. Ciò che si riscontra è piuttosto una realtà bilingue caratterizzata da una distribuzione complementare che vede l'utilizzazione dell'inglese in contesti formali e ufficiali, in quanto lingua percepita come più prestigiosa, là dove il maltese rimane ancora circoscritto alla sfera di vita privata degli individui.

Ad un bilinguismo ufficiale non corrisponde un'accettazione diffusa del maltese il cui uso risulta ridotto rispetto all'inglese che continua ad essere considerato uno strumento indispensabile per l'ascesa sociale.

Nel repertorio linguistico della comunità maltese l'inglese si configura come il codice deputato ad attribuire un maggior grado di formalità alla situazione comunicativa e come il mezzo indispensabile per veicolare contenuti tecnico-scientifici, fungendo così da 'indicatore', contemporaneamente di stile e del profilo socio-culturale dei parlanti.

Persino all'interno del sistema scolastico, maltese e inglese sono entrambi utilizzati largamente maltese e in alcune scuole è stato dimostrato che le lezioni si svolgono normalmente in modalità bilingue dando origine a forme di comunicazione intermedie in cui si genera commutazione di codice a diversi gradi (cfr. Camilleri, 1991).

Questo panorama sociolinguistico crea le condizioni perché si verifichino casi di commutazione e sovrapposizione tra inglese e maltese al punto tale da convergere nella formazione di ciò che è stato definito come '*Mixed Maltese English*', una varietà di contatto caratterizzata da commistione di inglese e maltese. La presenza dell'inglese all'interno del panorama sociolinguistico maltese si colloca lungo un *continuum* che si configura così, secondo una scala che considera la dimensione diatopica e diafasica:

1. dialetti maltesi (*malti djalettali*)⁷, 2. maltese standard (*malti pulit*)⁸, 3. maltese misto (*Malti mhallat/Mixed Maltese English*) 4. inglese di Malta (*Malti Ingliz/Maltese English*).

Per coloro che hanno una competenza bilingue in maltese e in inglese, la compresenza all'interno del loro repertorio di entrambi i codici determina numerose situazioni in cui la

⁷ In linea con tale riflessione è il dato che la dimensione di variazione che risulta particolarmente vitale è correlata con la variabile geografica: sull'isola di Malta si registra una presenza massiccia di dialetti che si differenziano tra loro non solo per varianti lessicali, ma anche per inventario fonologico e caratteristiche sintattiche.

⁸ Fino agli anni Settanta era in uso una varietà di maltese codificato nella letteratura e noto come *malti safi* (maltese puro) caratterizzata da un uso diffuso delle forme semitiche, seppur desuete, a livello lessicale e di morfosintassi e da una eliminazione di neologismi di origine italiana o inglese (cfr. Chircop, 1997).

commistione diventa sistematica producendo fenomeni di interferenza fonetica regolari che caratterizzano il *Mixed Maltese English*. Più che ad una varietà i cui confini sono nettamente individuabili all'interno dello spazio sociolinguistico maltese, ci si trova piuttosto di fronte ad 'addensamenti' – per utilizzare una nozione proposta da Berruto (1987) – tra due poli, inglese e maltese, con diversi gradi di commistione: dalla presenza isolata di elementi stranieri nelle varietà degli strati sociali più bassi fino a varietà caratterizzate da continui fenomeni di commutazione e mescolanza di codice nella produzione di parlanti appartenenti agli strati sociali più alti (il cosiddetto *Malti tal-pepè*, snob)⁹.

Obiettivo del presente studio è l'investigazione dei fenomeni di interferenza sul piano fonetico-fonologico che caratterizzano il continuo linguistico maltese al punto da divenire tratti della varietà dell'inglese di Malta (*Maltese English*), in misura più o meno massiccia, presenti anche nel maltese misto (*Mixed Maltese English*).

L'ipotesi di ricerca parte dall'analisi delle possibili aree di interferenza che emergono dal confronto tra gli inventari fonologici delle due lingue in esame; a questo proposito verrà dedicata una particolare attenzione alle aree che differenziano i due sistemi. Tra i possibili fenomeni di interferenza verranno considerati come plausibili i seguenti: l'ipodifferenziazione della laterale /l/, la neutralizzazione delle distinzioni di sonorità delle occlusive in posizione finale di parola, la sostituzione delle fricative interdentali presenti in inglese con le occlusive alveo-dentali /t/ e /d/, la sovrapposizione delle vocali nello spazio centrale del trapezio vocalico vista la non fonematicità dello *schwa* in maltese e la diminuzione dei fenomeni di riduzione di durata delle vocali non accentate – caratteristica dell'inglese - con conseguenti differenze sul piano ritmico.

Una particolare attenzione verrà rivolta all'interferenza fonetica, andando a verificare il grado di aspirazione delle occlusive sorde in base all'ipotesi che tali consonanti in maltese presentino dei valori diversi di aspirazione che verranno investigati acusticamente mediante la misurazione del tempo di attacco della sonorità (*Voice Onset Time*). Complessivamente l'analisi spettro-acustica ci consentirà di descrivere le variabili fonologiche distribuite lungo il continuo linguistico della comunità maltese esaminandone le proprietà fonetico-acustiche.

Riferimenti bibliografici

Azzopardi M. *The Phonetics of Maltese: some areas relevant to the deaf*. Unpublished Ph.D. thesis. University of Edinburgh. 1981

Bonnici, L.M. Maltese English. History of use, structural variation and sociolinguistic status, in B. Comrie/R. Fabri/E. Hume/M.Mifsud/T. Stolz/M. Vanhove, *Introducing Maltese Linguistics*, John Benjamins Company: Amsterdam-Philadelphia, pp.393-414

Borg A./Azzopardi-Alexander M. (1997) *Maltese*. London & New York. Routledge

Brincat, J. (2003) *Malta. Una storia linguistica*, Genova: Le Mani

Camilleri, A.

(1991) *Crosslinguistic influence in a bilingual classroom: the example of Maltese and English*,

Edinburgh Working Papers in Linguistics, 2, p101-111

(1995) *Bilingualism in education: the Maltese experience*. Heidelberg:Julius Groos

9 Cfr. Chircop, 1997

Camilleri-Grima, A. (2009) *Diglossia: Variation on a theme*, in B. Comrie/R. Fabri/E. Hume/M.Mifsud/T. Stolz/M. Vanhove, *Introducing Maltese Linguistics*, John Benjamins Company: Amsterdam-Philadelphia, pp. 379-391

Chircop, J. (1997) *Il lessico italiano della giustizia a Malta*. Tesi di laurea non pubblicata in Sociolinguistica, Facoltà di Lettere e Filosofia, Università degli studi di Firenze.

Mazzon, Gabriella (1993) English in Malta. English World-Wide, 14, pp. 171-208

Sciriha, L./Vassallo, M. (2006) *Living Languages in Malta*. Malta: IT Printing

GLI ITALIANI E IL DECODING DI EMOZIONI VOCALI IN QUATTRO LINGUE EUROPEE

Vincenzo Galatà, Luciano Romito

Dipartimento di Linguistica – Laboratorio di Fonetica, Università della Calabria

vgalata@libero.it, luciano.romito@unical.it

La comunicazione tra gli esseri umani è per definizione un processo multimodale e ogni qualvolta comunichiamo con gli altri individui della nostra specie poniamo in essere un meccanismo di codifica e decodifica di informazioni trasmesse dai singoli canali, un meccanismo che avviene a vari livelli senza grossi ostacoli e con una rapidità che spesso ci sorprende.

Sebbene la comunicazione sia da intendere come una continua integrazione di diversi canali di espressione (come ad es. quello del volto, dei gesti e non ultima della voce) siamo anche in grado di utilizzarli singolarmente, come avviene, ad esempio, nel caso estremo di soggetti sordomuti che, non potendo fare ricorso al canale vocale e uditivo, fanno ricorso a quello gestuale e visivo; o ancora a quello di soggetti privi della vista che fanno esclusivo riferimento al canale vocale e uditivo, oltre che a quello tattile. Si tratta ovviamente di casi particolari dove determinati canali di comunicazione vengono sviluppati più di altri per sopperire alle carenze o mancanze a carico di determinati organi sensoriali. Sulla scia degli esempi appena presentati, non è necessario andare molto lontano per rendersi conto di quanto accade per la comunicazione verbale in situazioni normali, alla quale ci rivolgeremo in questa sede. Come noto, attraverso la comunicazione verbale trasmettiamo ai nostri interlocutori molteplici informazioni, informazioni che Magno Caldognetto et al. (2008) hanno definito come informazioni di tipo linguistico, para-linguistico ed extra-linguistico. Ad esempio, nella realtà quotidiana e nel caso di soggetti normodotati, in assenza di contatto visivo con il nostro interlocutore e grazie a tali informazioni trasmesse attraverso il canale vocale, oltre a decodificare il messaggio verbale, siamo anche in grado di inferire e determinare tutta una serie di informazioni legate al nostro interlocutore, come ad esempio età, sesso, provenienza. Allo stesso modo, attraverso la voce, comunichiamo le nostre intenzioni emotive: durante una discussione con un altro individuo, siamo infatti in grado di capire, condividendo o meno tutta una serie di informazioni, se il nostro interlocutore sia triste, arrabbiato o altro, semplicemente facendo riferimento a “come” egli dice determinate cose.

Che gli esseri umani siano in grado di decodificare correttamente produzioni emotive con probabilità che vanno ben oltre la pura casualità è un dato oramai consolidato che emerge da numerosi studi sull’argomento anche nel caso in cui si affronti la tematica del parlato emotivo in lingue diverse dalla propria, come si evince dalle ricerche condotte sino ad oggi in contesti cross-linguistici e cross-culturali (si vedano proprio su quest’ultimo aspetto Kramer, 1964; Beier & Zautra, 1972; Solomon & Ali, 1975; McCluskey et al., 1975; Albas et al., 1976; McCluskey & Albas, 1981; Van Bezooijen, Otto, Henan, 1983; Magno Caldognetto & Kori, 1983; Rintell, 1984; Roseberry-McKibbin, Brice, 1999; Tickle, 1999 e 2000; Erickson, Maekawa, 2001; Scherer, Banse, Wallbott, 2001; Graham, Hamblin, Feldstein, 2001; Nakamichi, Jogan, Usami, Erickson, 2002; Shigeno, 1998 e 2003; Dromey, Silveira & Sandor, 2005; Sawamura et al., 2007; Shochi, Aubergé & Rilliard, 2007; Pell et al. 2009).

Stando alla meta-analisi condotta da Laukka (2004) emerge, infatti, come il *decoding*

cross-culturale sia inferiore al *decoding* intra-culturale del 7% facendo inoltre rilevare la presenza di quello che viene definito *in-group advantage* nel riconoscimento di emozioni nella propria lingua.

Il dato riportato da Laukka (*op. cit.*) confermerebbe quindi l'idea di Scherer, Banse & Wallbott (2001: 78) secondo i quali "it seems reasonable to assume that the recognition of vocal emotion expressions might work across language and culture boundaries".

Dagli studi sopra menzionati emerge come l'aspetto del *decoding* cross-linguistico e cross-culturale sia stato investigato in varie forme: ad esempio prendendo in esame contesti culturali nettamente contrapposti e spesso agli antipodi (come americani vs. giapponesi o ancora cinesi); utilizzando stimoli vocali composti da frasi *meaningless*, semplici parole (sia in versione originale che in versione filtrata e/o modificata), stimoli che in ciascun esperimento non sono stati raccolti in più di tre lingue. Pochissimi tra questi, infine, si sono occupati esclusivamente di lingue europee e ancora meno sono gli studi su stimoli in lingua italiana o su ascoltatori italiani.

In questo studio ci proponiamo di verificare l'abilità da parte di soggetti di madrelingua italiana nel decodificare emozioni vocali "posate" prodotte da soggetti attori e *naïf* in quattro lingue europee (italiano, francese, inglese e tedesco).

Obiettivi della presente proposta sono pertanto quelli di:

1. somministrare degli esperimenti percettivi per appurare la capacità di soggetti di madrelingua italiana a decodificare espressioni emotive nella propria lingua e lingue differenti dalla propria (inglese, francese e tedesco);
2. stabilire in che misura il grado di conoscenza e competenza della lingua possa influire sul riconoscimento delle produzioni emotive presentate verificando, al contempo, quanto ipotizzato da Scherer, Banse, Wallbott (2001: 88): "It seems that as soon as vocal expressions, other than pure nonlinguistic affect bursts, are used, segmental and suprasegmental aspects of language affect encoding and decoding of emotion."

Riferimenti bibliografici

Albas, D. C., McCluskey, K.W., Albas, C. A. (1976), Perception of the emotional content of speech: A comparison of two Canadian groups. *Journal of Cross-Cultural Psychology*, 7, pp. 481-489.

Beier, E. G., & Zautra, A. J. (1972), Identification of vocal communication of emotions across cultures, *Journal of Consulting and Clinical Psychology*, 39, 166.

Dromey, C., Silveira, J., Sandor, P. (2005), Recognition of affective prosody by speakers of English as a first or foreign language, *Speech Communication*, 47(3): 351-359.

Erickson, D. & Maekawa, K. (2001), Perception of American English emotion by Japanese listeners, *Acoustical Society of Japan, Spring Meeting*, pp. 333-334.

Graham, C.R., Hamblin, A., Feldstein, S. (2001), Recognition of emotion in English voices by speakers of Japanese, Spanish and English, *IRAL*, 39: 19-37.

Kramer, E. (1964), Elimination of verbal cues in judgments of emotion from voice, *Journal of Abnormal and Social Psychology*, 68, pp. 390-396.

Laukka, P. (2004), Vocal expression of emotion: discrete-emotions and dimensional accounts, *Ph.D thesis*, Uppsala University.

Magno Caldognetto E., Kori S. (1983), Intercultural judgment of emotions expressed through voice, in: *Quaderni del Centro di Studio per le Ricerche di Fonetica*, 2,

pp. 339-363.

Magno Caldognetto, E., Cavicchio, F., Cosi, P. (2008), La faccia e la voce delle emozioni, in: Poggi I. (a cura di), *La mente del cuore. Scienze cognitive ed emozioni*, Roma: Armando, pp. 178-212.

McCluskey, K.W., Albas, D. C. (1981), Perception of the emotional content of speech by Canadian and Mexican children, adolescents and adults, *International Journal of Psychology*, 16, pp. 119-132.

McCluskey, K.W., Albas, D. C., Niemi, R. R., Cuevas, C., Ferrer, C. A. (1975), Cross-cultural differences in the perception of emotional content of speech: A study of the development of sensitivity in Canadian and Mexican children, *Developmental Psychology*, 11: 15-21.

Nakamichi, A., Jogan, A., Usami, M. & Erickson, D. (2002), Perception by native and non-native listeners of vocal emotion in a bilingual movie, in: *Gifu City Women's College Research Bulletin*, 52: 87-91.

Pell, M.D., Monetta, L., Paulmann, S. & Kotz, S.A. (2009), Recognizing emotions in a foreign language, in: *Journal of Nonverbal Behavior*, 33: 107-120.

Rintell, E. (1984), "But how did you FEEL about that?": The learner's perception of emotion in speech, *Applied Linguistics* 5(3): 255-264.

Roseberry-McKibbin, C., Brice, A. (1999), The perception of vocal cues of emotion by Spanish-speaking limited English proficient children, *Journal of Children's Communication Development*, 20(2): 19-25.

Sawamura, K., Dang, J., Akagi, M., Erickson, D., et al. (2007), Common factors in emotion perception among different cultures, *Proceedings of ICPHS 2007*, Saarbrücken, Germany, pp. 2113-2116.

Scherer, K. R., Banse, R., Wallbott, H.G. (2001), Emotion Inferences from Vocal Expression Correlate across Languages and Cultures, *Journal of Cross-Cultural Psychology* 32(1), pp. 76-92.

Shigeno, S. (1998), Cultural similarities and differences in the recognition of audio-visual stimuli, 5th *International Conference on Spoken Language Processing*, paper 1057, 149-152.

Shigeno, S. (2003), Recognition of emotion transmitted by vocal and facial expression: Comparison between the Japanese and the American, in: *The AGU Journal of Psychology*, 3: 1-8.

Shochi, T., Aubergé, V. & Rilliard, A. (2007), Cross-Listening of Japanese, English and French social affect: about universals, false friends and unknown attitudes, *Proceedings of ICPHS 2007*, Saarbrücken, Germany, pp. 2097-2100.

Solomon, D., Ali, F. A. (1975), Influence of verbal content and intonation on meaning attributions of first-and-second language speakers, *Journal of Social Psychology*, 95: 3-8.

Tickle, A. (1999), Cross-language vocalisation of emotion: methodological issues, *Proceedings of the 14th International Conference of Phonetic Sciences*, San Francisco, USA, pp. 305-308.

Tickle, A. (2000), English and Japanese speakers' emotion vocalization and recognition: A comparison highlighting vowel quality, *Speech and Emotion*, 104-109.

van Bezooijen, R., Otto, S. A., Heenan, T. A. (1983), Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics, *Journal of Cross-Cultural Psychology*, 14, pp. 387-406.

IDENTIFICAZIONE DEI SEGNALI IN UNA SCENA SONORA

Franca Orletti¹, A. Paoloni

¹Università di Roma Tre

orletti@uniroma3.it, pao@fub.it

In molte applicazioni giudiziarie viene richiesto al trascrittore e a volte ad un esperto appositamente convocato se un determinato rumore possa essere attribuito ad una causa ben definita: ad esempio è stato chiesto di accertare se un rumore metallico potesse essere attribuito allo scarrellamento di un'arma, se un rumore impulsivo fosse attribuibile a uno sparo, se un altro rumore, di difficile descrizione, potesse essere attribuito ad un frullatore in funzione.

In letteratura esistono numerosi lavori sul riconoscimento e sulla identificazione dei rumori ma la maggior parte di essi ha come obiettivo lo studio di quali siano le modalità con cui l'uomo identifica, quali le caratteristiche acustiche che caratterizzano un suono e quali le differenze tra ascolto diretto e ascolto del segnale registrato in monofonico o in stereo.

Nelle applicazioni giudiziarie il segnale è sempre registrato e quasi sempre di qualità telefonica; in questo ambito inoltre è importante poter valutare l'attendibilità dell'attribuzione o più precisamente il rapporto di verisimiglianza tra l'attribuzione effettuata ed eventuali attribuzioni alternative. Sarebbe di grande interesse poter disporre di una procedura oggettiva atta a rispondere al quesito ora posto, ma sulla base del segnale acustico normalmente disponibile, segnale di scarsa o pessima qualità, è molto difficile, se non impossibile, trovare tale procedura.

Il presente lavoro ha come obiettivo quello di valutare se le attribuzioni soggettive che di norma vengono effettuate sulla base dell'ascolto di un segnale acustico all'interno di una scena sonora siano attendibili almeno nella misura in cui non si ritenga di voler sospettare ad una volontaria imitazione di un determinato suono.

Esperimento

Come detto obiettivo dell'esperimento è di verificare se, sulla base di un semplice ascolto, le persone siano in grado di determinare la natura di un suono, nel senso di indicare quale sia la sua sorgente. L'identificazione viene effettuata sulla base di un segnale di tipo telefonico o meglio ancora su segnale trasmesso da un telefono cellulare. Gli autori hanno pensato di sottoporre all'ascolto di un gruppo di persone non particolarmente addestrate, un certo numero di suoni particolari. Al fine di predisporre l'esperimento è stato necessario individuare tali suoni che debbono soddisfare le condizioni di essere caratteristici, ossia ben distinguibili, comuni, ossia noti a tutti i soggetti che partecipano al test, e di poter essere in qualche misura presenti in uno scenario di tipo forense.

Si è pensato di mescolare a suoni di origine quotidiana, come l'abbaiare di un cane, il tintinnio di un mazzo di chiavi, il rumore di una serratura che si apre, il rumore di un vetro che si rompe, il rumore di un frullatore, suoni di origine meno comune, come lo sparo di un fucile, lo sparo di una pistola, il rumore di una mitragliatrice, il rumore di un'auto, il rumore di un elicottero, ecc.

Inoltre, per verificare quanto un determinato contesto possa influenzare la decisione sulla natura del suono ascoltato, agli osservatori saranno somministrati due diversi test: in un primo test gli ascoltatori dovranno semplicemente scrivere quale sia la sorgente del rumore da loro ascoltato; nel secondo invece verrà loro fornita un'indicazione circa la scena sonora, ovvero in quale contesto si trova la persona che ascolta quel suono. Entrambi i test

saranno preceduti da dei suoni filler dei quali verrà immediatamente dichiarata l'origine al fine di addestrare i soggetti alle modalità di test.

Riferimenti bibliografici

Carello, C., J.B. Wagman and M.T. Turvey (2005) Acoustic specification of object properties. In: *Moving Image Theory: Ecological Considerations* (J.D. Anderson and B.F. Anderson, Ed.), pp. 79-104. Southern Illinois University Press, Carbondale, IL.

Gygi, B., Kidd, G. R., & Watson, C. S., (2004). Spectral-temporal factors in the identification of environmental sounds. *J Acoust Soc Am*, 115, 1252–1265.

Lawrence, D. M., & Banks, W. P. (1973). Accuracy of recognition memory for common sounds. *Bull Psychon Soc*, 1, 298–300.

Li, X., Logan, R. J., & Pastore, R. E. (1993). Perception of acoustic source characteristics: walking sounds. *J Acoust Soc Am*, 90, 3036–3049.

Michael Marcell, Maria Malatanos, Connie Leahy, and Cadie Comeaux *College of Charleston, Charleston, South Carolina* Identifying, rating, and remembering environmental sound events

Behavior Research Methods 2007, 39 (3), 561-569

Marcell, M. M., Borella, D., & Greene, J., et al. (2000). Confrontation naming of environmental sounds. *J Clin Exp Neuropsychol*, 22, 830–864.

R. K. Reddy · V. Ramachandra · N. Kumar · Nandini Chatterjee Singh
Categorization of environmental sounds Biological Cybernetics, 2009 - Springer

Shafiro, V., & Gygi, B. (2004). How to select stimuli for environmental sound research and where to find them. *Behav Res Methods Instrum Comput*, 36(4), 590–598.

Shafiro, V. (2008). Identification of environmental sounds with varying spectral resolution. *Ear & Hear*, 29, 401–420.

Watson, C. S., & Kidd, G. R. (2002). On the lack of association between basic auditory abilities, speech processing and other cognitive skills. *Semin Hear*, 23, 83–93.

VanDerveer, N.J. (1979). Ecological acoustics: Human perception of environmental sounds. Doctoral dissertation. Cornell University, Ithaca, NY.

DISTINGUERE PATTERN ACUSTICI CON IDENTICA TRASCRIZIONE FONEMICA MA STRUTTURA MORFOLOGICA DIFFERENTE: UN MODELLO COMPUTAZIONALE

Marco A. Piccolino-Boniforti
University of Cambridge
map55@cam.ac.uk

L'obiettivo della mia ricerca è la realizzazione di un modello computazionale della percezione del parlato nell'uomo che sia in grado di predire la struttura grammaticale di una sequenza acustica sfruttando il dettaglio fonetico contenuto in essa.

Presenterò un *proof-of-concept* in grado di distinguere prefissi veri e propri da cosiddetti pseudoprefissi in inglese britannico, in parole come "discolour" (dove "dis" è un prefisso) e "discover" (dove "dis" è uno pseudoprefisso). Entrambe le parole hanno i primi quattro fonemi in comune (/dɪsk/), ma analisi fonetiche e linguistiche (Ogden *et al.* 2000) hanno mostrato che gli pseudoprefissi tendono ad avere peso ritmico più debole rispetto ai prefissi. Alcune delle dimensioni acustiche che causano tali differenze ritmiche sono state individuate (Baker *et al.* 2007a; Baker 2008 dissertazione). Descriverò i risultati ottenuti in tre simulazioni mirate alla valutazione di singoli componenti del modello: nello specifico, la rappresentazione della struttura ritmica basata sull'*auditory primal sketch* (Todd 1994) e il sistema di classificazione basato sulla *relevance vector machine* (Tipping 2001). Presenterò infine alcune estensioni su cui sto lavorando che dovrebbero permettere la comparazione del modello con i risultati di un esperimento di ascolto che sfrutta la tecnica dell'*eyetracking*.

SESSIONE FONETICO-FONOLOGICA

A. Stella *L'influenza della L1 nell'intonazione della L2: l'allineamento tonale in parlanti con differenti livelli di competenza*

C. Pisciotta, S. Bernardini, E. Agazzi, N. Crivelli, F. Manni, R. Perosa, D. Stocco, C. Zmarich *Indici fonetici predittivi di balbuzie cronica in età prescolare: studio di un caso*

G. Schirru *La pendenza spettrale come indice acustico della tensione laringea*

B. Gili Fivela, S. D'Apolito, A. Stella, F. Sigona *Domain initial strengthening: dati acustici ed articolatori relativi a due varietà di italiano*

M. C. Felloni, D. Avesani *La percezione della interrogativa globale nell'italiano regionale di Parma*

V. Retaro *L'abbassamento di [ɛ] nelle varietà dialettali nord-vesuviane: un'analisi socio fonetica*

L'INFLUENZA DELLA L1 NELL'INTONAZIONE DELLA L2: L'ALLINEAMENTO TONALE IN PARLANTI CON DIFFERENTI LIVELLI DI COMPETENZA

Antonio Stella
Università del Salento
antonio.stella@unisalento.it

Le produzioni in una lingua straniera (L2) sono comunemente influenzate dalle caratteristiche segmentali e suprasegmentali del sistema fonetico/fonologico della lingua madre (L1). Per quanto riguarda il livello segmentale, i suoni di una L2 che sono simili ad altri di una L1 sono profondamente influenzati dal sistema nativo; tali influenze sono

percepiti come accento straniero ma possono anche essere accettabili dal punto di vista dell'intelligibilità e non portano necessariamente ad una riduzione della comprensione (Munro e Derwing, 1995). Lo stesso non si può dire per i foni di una L2 che non sono simili ai quelli di una L1, così come per le caratteristiche suprasegmentali, le quali possono addirittura influenzare il significato pragmatico dell'enunciato e portare a incomprensioni: Anderson-Hsieh *et al.* (1992) mostrano infatti che tra il livello segmentale, la struttura sillabica e la prosodia, quest'ultima esercita un'influenza più forte sull'intelligibilità delle produzioni in una lingua straniera.

L'influenza del sistema fonetico/fonologico nativo nella prosodia di una L2 è rilevata anche in parlanti con un alto livello di competenza della lingua straniera. Mennen (2007), comparando differenti studi riguardanti l'intonazione nelle produzioni in lingua straniera, identifica una serie di caratteristiche influenzate da una L1, che divide in due gruppi: errori fonologici (uso di una categoria invece di un'altra; ad esempio, un'ascesa invece di una discesa) ed errori fonetici (differenti implementazioni della stessa categoria; per esempio, una differente inclinazione della curva intonativa per una stesso accento ascendente).

Una frequente sorgente di errori nella produzione della prosodia di una L2 è rappresentata dall'allineamento tonale, il quale si riferisce alla relazione temporale tra i target tonali (alti (H) e bassi (L)) e la stringa segmentale. Un esempio di errore fonetico commesso in una produzione in lingua straniera è dovuto a differenze nell'allineamento tonale è riportato da Mennen (2004) in un esperimento sulla prosodia del greco L2 prodotto da parlanti nativi dell'olandese; sia il greco che l'olandese condividono infatti un accento prenucleare (L+H*) che è allineato più a destra nel greco rispetto all'olandese; l'autrice mostra che quattro su cinque parlanti nativi dell'olandese con un alto livello di competenza del greco trasferiscono le proprietà fonetiche di questo *pitch accent* dalla loro lingua madre nelle produzioni in greco, allineando in anticipo il picco di F0 sulla sillaba target.

Diversi studi sull'allineamento degli accenti ascendenti di varie lingue mostrano che l'allineamento del target basso L è “ancorato” all'*onset* della sillaba, mentre il target H risulta essere più variabile e sembra non “ancorarsi” ad uno specifico punto nella sillaba (*inter alia*: per l'olandese, Caspers e Van Heuven, 1993; per lo spagnolo messicano, Prieto

et al., 1995; per il greco, Arvaniti *et al.*, 1998; per l'inglese, Ladd *et al.*, 2000). Per esempio, Prieto e Torreira (2007) analizzano l'allineamento dell'accento prenucleare L+H* nello spagnolo e mostrano che il target basso dell'ascesa prenucleare è allineata con l'*onset* sillabico in maniera stabile, mentre il target alto è profondamente influenzato dalla struttura sillabica e dalla velocità di elocuzione. Prieto (2009) ripete le stesse misure in un altro esperimento con l'accento nucleare discendente H+L*, il quale è utilizzato in catalano nelle

domande polari, rilevando un pattern d'allineamento simile a quello utilizzato nell'accento prenucleare ascendente già menzionato: il target alto H è infatti strettamente sincronizzato con l'onset della sillaba accentata, mentre il target basso L mostra un allineamento più variabile ed è influenzato dalla struttura sillabica e dalla velocità di elocuzione. È stato quindi proposto che l'asimmetria dei pattern d'allineamento in relazione alla posizione nella sillaba accentata sia da attribuire alle proprietà della coordinazione intergestuale (Pietro e Torreira, 2007; Prieto, 2009). Osservare la sincronia tra i gesti sopralaringali e la struttura sillabica può essere utile per comprendere i meccanismi che sono alla base dell'allineamento dei bersagli tonali con il materiale segmentale. Sono stati già condotti alcuni studi sull'allineamento dei target tonali dal punto di vista articolatorio su varie lingue (D'Imperio *et al.*, 2003; D'Imperio *et al.*, 2007; Prieto *et al.*, 2007) ed hanno mostrato infatti che l'allineamento dei toni rispetto ai gesti articolatori è più stabile dell'allineamento con punti di riferimento acustici.

Il lavoro presentato in questa sede mira ad esplorare l'allineamento dei target tonali dal punto di vista acustico e articolatorio nella produzione di differenti tipi di focus in italiano e in tedesco prodotti da parlanti di area leccese con differenti livelli di competenza del tedesco come L2. Numerosi studi hanno infatti dimostrato che il grado di influenza del sistema nativo a livello sia fonologico che fonetico può essere realmente correlato al grado di competenza della L2 (Ueyama, 1997; Ueyama e Jun, 1998; Jun e Oh, 2000).

I contorni intonativi sono stati analizzati seguendo la teoria Autosegmentale-Metrica (vedi Ladd, 1996 per una descrizione generale) e descritti utilizzando due versioni modificate del sistema ToBI (Beckman e Ayers, 1997): per l'italiano è stata utilizzata una versione del sistema originale adattata per la varietà di Lecce (Stella e Gili Fivela, in stampa); per il tedesco è stato utilizzato GToBI (Grice *et al.*, 2005).

Dati acustici e cinematici delle produzioni di soggetti con tre differenti livelli di competenza (basso, medio e alto) sono stati acquisiti con l'articulografo AG500 presso il centro di ricerca CRIL dell'Università del Salento. Per esplorare l'allineamento dei bersagli tonali in maniera coerente, cioè confrontando tra loro le produzioni in entrambe le lingue dai punti di vista acustico e cinematico, sono state acquisite produzioni nelle quali il focus è prodotto su parole contenenti le sillabe /ma/ e /man/ in posizione accentata. Infatti, quando queste sillabe sono accentate, vengono implementate alla stessa maniera a livello segmentale sia in italiano che in tedesco (cfr. Di Meola, 2007) e sono perciò direttamente comparabili tra loro. Per ogni lingua sono state scelte sette parole con le sillabe target in posizione accentata, composte da due o tre sillabe e con l'accento lessicale sulla terzultima, penultima o ultima sillaba. Le parole target sono state utilizzate per creare due corpora di frasi (uno in italiano ed uno in tedesco) che sono stati letti dai soggetti per almeno dieci volte a due differenti velocità di elocuzione (normale e veloce).

L'ipotesi è che vi siano sostanziali differenze tra soggetti con differente livello di competenza nella produzione di accenti tonali prenucleari e nucleari in entrambe le lingue, a causa dell'influenza del sistema fonologico nativo; ciò è particolarmente evidente quando i soggetti producono enunciati contenenti un focus stretto-contrastivo usato per correggere una precedente asserzione del partner, perché il tedesco standard e l'italiano parlato a Lecce utilizzano due differenti categorie in tali condizioni di focus: infatti, seguendo GToBI, il tedesco standard utilizza un accento ascendente L+H* per esprimere contrasto (Grice *et al.*, 2005; Baumann *et al.*, 2007; evidenze riguardanti l'uso di un accento ascendente in temi contrastivi provengono anche da Braun e Ladd, 2003), mentre l'italiano parlato a Lecce usa un accento discendente, analizzato come H*+L (Stella e Gili Fivela, in stampa). In questo studio, la condizione di focus stretto-contrastivo è stata ricreata utilizzando una

coppia domanda-risposta, nella quale la domanda è stata prodotta con la focalizzazione in posizione iniziale di enunciato.

In generale, ci si aspetta che i parlanti con un basso livello di competenza utilizzino un accento discendente (o una modificazione di esso) per esprimere questo tipo di contrasto in tedesco, a causa del *transfer* dalla L1; al contrario, in parlanti con un alto livello di competenza del tedesco ci si aspetta l'uso della categoria corretta, cioè un accento ascendente (o, ancora, una modificazione di esso). Ovviamente, non si presenteranno differenze nette tra i livelli di competenza, data la grande variabilità nelle produzioni; tuttavia, comparando le produzioni dal punto di vista del livello di competenza, si potranno probabilmente rilevare dei *trend* gradualmente nella modificazione dell'allineamento.

Nel presente contributo saranno discussi solo i dati acustici; tali risultati forniranno un'analisi delle variazioni nell'allineamento dei target in relazione al livello di competenza, alla struttura sillabica e alla velocità d'eloquio; questo studio costituirà poi il punto di partenza per un'indagine articolatoria sulla sincronia tra i gesti che compongono la sillaba target e la loro interazione con i target alti e bassi dei *pitch accent* associati.

Riferimenti bibliografici

Anderson-Hsieh, J., Johnson, R. and K. Koehler (1992), The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure, in *Language Learning*, 42, 529–555.

Arvaniti, A., Ladd, D. R. and I. Mennen (1998), Stability of tonal alignment: the case of Greek prenuclear accents, in *Journal of Phonetics*, 26, 3–25.

Baumann, S., Becker, J., Grice, M. and D. Mücke (2007), Tonal and articulatory marking of Focus in German, in *Proceedings of the XVI International Congress of Phonetic Sciences* (J. Tourvain and W. J. Barry, editors), Pirrot GmbH, Dudweiler, 1029–1032.

Beckman, M. and G. Ayers (1997), Guidelines for ToBI labelling, Version 3.0, Ms. And accompanying speech materials, OSU, Ohio.

Braun, B. and D. R. Ladd (2003), Prosodic correlates of contrastive and noncontrastive themes in German, in *Proceedings of 8th European Conference on Speech Communication and Technology*, Geneva, 789–792.

Caspers, J., and V. J. van Heuven (1993), Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall, in *Phonetica*, 50, 161–171.

D'Imperio, M., Nguyen, N. and K. G. Munhall (2003), An articulatory hypothesis for the alignment of tonal targets in Italian, in *Proceedings of 15th International Congress of Phonetic Sciences*, Barcelona, 253–256.

D'Imperio, M., Espesser, R., Loevenbruck, H., Menezes, C., Nguyen, N. and P. Welby (2007), Are tones aligned with articulatory events? Evidence from Italian and French, in *Papers in Laboratory Phonology IX* (J. Cole, J. I. Hualde, editors), Mouton de Gruyter, 577–608.

Di Meola, C. (2007), *La linguistica tedesca. Un'introduzione con esercizi e bibliografia ragionata*, Roma, Bulzoni.

Grice, M., Baumann, S. and R. Benz Müller (2005), German Intonation in Autosegmental-Metrical Phonology, in S.-A. Jun (ed.) *Prosodic Typology and Transcription. A Unified Approach*, Oxford, OUP, 55–83.

Jun, S.-A. and M. Oh (2000), Acquisition of second language intonation, in *Proceedings of International Conference on Spoken Language Processing*, 4, 76–79.

Ladd, D. R. (1996), *Intonational Phonology*, Cambridge, CUP.

- Ladd, D. R., Mennen, I., Schepman, A. (2000), Phonological conditioning of peak alignment in rising pitch accents in Dutch, in *Journal of Acoustical Society of America*, 107, 2685-2696.
- Mennen, I. (2004), Bi-directional interference in the intonation of Dutch speakers of Greek, in *Journal of Phonetics*, 32, 543-563.
- Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. In J. Trouvain and U. Gut (eds), *Non-native Prosody: Phonetic Descriptions and Teaching Practice*, The Hague: Mouton De Gruyter, 53-76.
- Munro, M. J. and Derwing, T. M. (1995), Foreign accent, comprehensibility and intelligibility in the speech of second language learners, in *Language Learning*, 45, 73-97.
- Prieto, P., van Santen, J. and J. Hirschberg (1995), Tonal alignment patterns in Spanish, in *Journal of Phonetics*, 23, 429-451.
- Prieto, P., Mücke, D., Becker, J. and M. Grice (2007), Coordination patterns between pitch movements and oral gestures in Catalan, in *Proceedings of the XVI International Congress of Phonetic Sciences* (J. Tourvain and W. J. Barry, editors), Pirrot GmbH, Dudweiler, 989-992.
- Prieto, P. and F. Torreira (2007), The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish, in *Journal of Phonetics*, 35, 473-500.
- Prieto, P. (2009), Tonal alignment patterns in Catalan nuclear falls, in *Lingua*, 119, 865-880.
- Stella, A. and B. Gili Fivela (in stampa), L'intonazione nell'italiano dell'area leccese: prime osservazioni dal punto di vista autosegmentale-metrico, in *Atti del Convegno AISV 2007, 2-3 December 2007, Università della Calabria, Arcavacata di Rende (CS), Italy*.
- Ueyama, M. (1997), The phonology and phonetics of second language intonation: the case of "Japanese English", in *Proceedings of the 5th European Speech Conference*, 2411-2414.
- Ueyama, M. and S.-A. Jun, (1998), Focus realization in Japanese English and Korean English intonation, in *UCLA Working Papers in Phonetics*, 94.

INDICI FONETICI PREDITTIVI DI BALBUZIE CRONICA IN ETA' PREScolare: STUDIO DI UN CASO

°Caterina Pisciotto, #Simona Bernardini, °Eleonora Agazzi, °Noemi Crivelli, ° Francesca Manni , #Roberta Perosa, #Debora Stocco, °Claudio Zmarich

°Istituto di Scienze e Tecnologie della Cognizione (ISTC), C.N.R., Sede di Padova;
#Centro Medico di Foniatria, Padova

caterina.pisciotta@pd.istc.cnr.it, sbernardini@centrofoniatra.it, conteddracula@live.it,
noemicrivelli@libero.it, francesca.manni@gmail.com, rperosa@centrofoniatra.it,
dstocco@centrofoniatra.it, claudio.zmarich@pd.istc.cnr.it

La balbuzie è un disturbo della parola che affligge l'1% della popolazione mondiale (tasso di prevalenza), percentuale che sale al 5% (tasso di incidenza) se si considerano i casi in cui il disturbo scompare spontaneamente o tramite terapia e che può avvenire fino a 3 anni dopo l'insorgenza, che recenti ricerche situano a un'età media di 33 mesi (Yairi & Ambrose, 2005).

Tra tutti i bambini che iniziano a balbettare, quelli che cronicizzeranno (20%), probabilmente su base genetica (Bloodstein & Bernstein Ratner, 2008), devono essere individuati e trattati il prima possibile per ottenere una remissione che per molti specialisti potrebbe essere completa e priva di ricadute (Yairi & Ambrose, 2005).

Infatti l'intervento precoce previene, in molti casi, la balbuzie cronica (per es. Curlee, 1999; Yaruss et al., 2006) e permette, nella maggior parte dei casi, un intervento di durata più breve.

Il progetto di ricerca scientifica "Indici fonetici predittivi della balbuzie in età prescolare", di cui è responsabile il Dr. Claudio Zmarich dell'ISTC-CNR, sede di Padova, è finanziato dal CNR (cod. RSTL 995, 2007) e supportato dal Centro Medico di Foniatria di Padova, ed è volto all'individuazione di indici clinici predittivi, di tipo percettivo ed acustico, che consentano in fase di valutazione di identificare già nei primi mesi dopo l'esordio la balbuzie cronica da quella remissiva in bambini a rischio genetico.

Poiché ci sono evidenze scientifiche che la predisposizione alla balbuzie viene trasmessa per via genetica, e dato che in letteratura esistono prove di tipo fonetico (analisi percettiva delle disfluenze, analisi acustica della coarticolazione intrasillabica Consonante-Vocale) che consentirebbero di individuare precocemente su base probabilistica i bambini a rischio di balbuzie cronica, lo scopo del presente progetto mira a:

1. individuare gli indici predittivi di balbuzie in una popolazione di bambini genericamente a rischio per predisposizione genetica, ma non ancora balbuzienti;
2. individuare gli indici predittivi che consentano di differenziare la balbuzie cronica da quella remissiva in bambini, sempre a rischio genetico, ma che hanno appena cominciato a balbettare.

Il soggetto sperimentale del presente studio è una bambina che chiameremo Lucia, che ha 4 anni nel momento della stesura dell'abstract, è figlia di genitore balbuziente (padre), balbuziente lei stessa dai 30 mesi d'età, già facente parte di un progetto pilota, precedente al progetto RSTL, per il quale avevamo incominciato a registrarla con sedute mensili dal compimento del sesto mese di età. Lucia è stata sottoposta, a 12 mesi dall'insorgenza della balbuzie, a trattamento logopedico con il *Lidcombe Program* (Onslow, Packman, Harrison, 2003).

Saranno analizzate 3 registrazioni, in ambiente domestico, dell'eloquio della bambina in un compito di denominazione e descrizione di figure: la prima effettuata un mese prima dall'insorgenza della balbuzie, la seconda registrata in concomitanza all'insorgenza, la terza a 12 mesi dall'insorgenza, ossia nel periodo immediatamente precedente all'inizio del trattamento.

Le tre trascrizioni sono state effettuate in simboli IPA (1999) e codificate con il programma *open-source* Phon. Si è proceduto pertanto alla compilazione degli inventari fonetici delle parole balbettate e non balbettate.

Sulla base di queste codifiche verrà qui presentato il profilo delle disfluenze, secondo la classificazione di Yairi (Yairi & Ambrose, 2005), e un'analisi della complessità fonetica delle parole balbettate secondo il modello EXPLAN di Howell & Au-Yeung (2002), che combinano le ipotesi su problemi nella pianificazione centrale di tipo fonetico o lessicale con una dissincronia tra pianificazione ed esecuzione dell'articolazione.

Si effettuerà inoltre l'analisi acustica utilizzando la tecnica delle equazioni di luogo per quantificare il grado di coarticolazione nella sillaba C-V (Sussman et al., 1999; Zmarich & Marchiori, 2005) e la misurazione del V.O.T.

Ci aspettiamo che i risultati, per quanto riguarda il profilo delle disfluenze, l'analisi acustica ed il V.O.T, siano assimilabili a quelli emersi dal calcolo del profilo delle disfluenze dell'*Illinois Longitudinal Study* (Yairi & Ambrose, 2005), ossia che il numero di *Stuttering-Like Disfluencies* (SLD) decresca considerevolmente già dai 7 ai 12 mesi dall'inizio della balbuzie, nel caso di bambini con remissione spontanea.

Per quanto riguarda il modello EXPLAN, ci aspettiamo che, nella produzione fluente, la pianificazione ed esecuzione siano indipendenti e procedano in parallelo (cfr. produzione incrementale di Levelt, 1989), ma che nella produzione disfluente dei balbuzienti la produzione incrementale subisca un arresto, generando una dissincronia tra pianificazione ed esecuzione, a causa di difficoltà nel pianificare ed eseguire i gesti fonetici di parole complesse dal punto di vista fonetico (per es. più lunghe, con nessi consonantici, fonemi poco frequenti ecc.), e nei bambini anche sintattico (all'inizio di enunciato).

Riferimenti bibliografici

AA. VV., *Handbook of the International Phonetic Association*, (1999), Cambridge: Cambridge University Press.

Bloodstein, O. & Bernstein Ratner, N. (2008), *A Handbook on Stuttering*, Thomson Delmar Learning.

Curlee, R.F. (1999), *Stuttering and Related Disorders of Fluency* (2nd ed.). New York: Thieme Medical Publishers.

Howell, P. & Au-Yeung, J. (2002) The EXPLAN theory of fluency control and the diagnosis of stuttering. In Fava E. (Ed.), *Pathology and therapy of speech disorders*, John Benjamins: Amsterdam, 75-94.

Levelt, W.J.M. (1989), *Speaking*, MIT Press, Cambridge, USA.

Onslow, M., Packman, A., Harrison, E. (2003), *The Lidcombe Program of early stuttering intervention: a clinician's guide*, Austin TX: Pro-Ed.

Phon: <http://phon.ling.mun.ca/phontrac>

Sussman, H.M., Duder, C., Dalston, E., Cacciatore, A. (1999), An acoustic analysis of the development of CV coarticulation: a case study, *Journal of Speech, Language, and Hearing Research*, 42.

Yairi, E. & Ambrose, N.G (2005), *Early Childhood Stuttering*, Pro-Ed, Austin-

USA.

Yaruss, JS., Coleman, C., Hammer, D. (2006), Treating Preschool Children Who Stutter: Description and Preliminary Evaluation of a Family Focused Treatment Approach, *Lang., Speech Hear. Services in Schools*, 37, 118-136.

Zmarich C., Marchiori M., (2005), L'influenza del focus contrastivo sulla coarticolazione anticipatoria di sillabe "CV" prodotte fluentemente da balbuzienti e non balbuzienti, in *Misura dei parametri*, Atti del 1° Convegno Nazionale dell'Associazione Italiana di Scienze della Voce (AISV) (P. Cosi, a cura di), Padova, 2-4 dicembre, 2004, EDK Editore, Brescia, 2005, 231-250.

LA PENDENZA SPETTRALE COME INDICE ACUSTICO DELLA TENSIONE LARINGEA

Giancarlo Schirru
Università di Cassino
g.schirru@unicas.it

La teoria fonologica ha fatto uso della distinzione tra consonanti tese e consonanti rilassate raggruppando tra loro fenomeni che spesso, nella tradizione, erano stati considerati distinti: per esempio la differenza tra consonanti sorde e sonore, tra forti e leni, o consonanti comportanti un diverso grado di tensione laringea (p. es. cricchiate, modali, mormorate, ecc.). Sotto il profilo formale, tale uso si è manifestato nell'applicazione al consonantismo del tratto distintivo [\pm teso] (Jakobson – Fant – Halle 1952) o nell'introduzione della coppia di tratti [\pm corde vocali rigide] [\pm corde vocali rilassate] (cfr. la discussione in Stevens 1977; Fant 1977; Kenstowicz 1994; Stevens 2000).

Il presente contributo si propone di indagare la differenza di tensione laringea nel consonantismo, e di confrontare questa dimensione con la differenza tra consonanti forti e consonanti leni.

A questo fine verrà indagata la pendenza spettrale come indicatore della tensione vocale: secondo i dati offerti dalla letteratura (Cao – Maddieson 1993; Ladefoged – Maddieson 1996: 53-66; Ni Chasaide – Gobl 1997; Gordon – Ladefoged 2001; Hanson *et alii* 2001), la pendenza spettrale può essere misurata quantitativamente prendendo in esame i seguenti parametri:

- g) H2-H1
- h) A1-H1
- i) A3-H1

dove:

- H1 = intensità della prima armonica
- H2 = intensità della seconda armonica
- A1 = intensità della prima formante
- A3 = intensità della terza formante

Mediante tali parametri verrà analizzata la differenza tra la seconda e la terza serie di occlusive dell'armeno orientale (con dati raccolti dall'autore in un'inchiesta sul terreno – svolta in quattro diverse località e coinvolgente un totale di 27 parlanti – sulla base di una lista di coppie minime in cui le consonanti rilevanti compaiono sempre in posizione iniziale): più in particolare sarà sottoposta ad analisi la porzione di segnale vocalico corrispondente ai primi 50 msec successivi al rilascio della consonante. Si mostrerà quindi come la distinzione tra queste due serie, generalmente considerata come una differenza tra sorde non aspirate e sonore (Ladefoged – Maddieson 1996: 66-68; Vaux 1998: 12-13; Hacopian 2003), sia più realisticamente caratterizzabile in base al diverso grado di tensione laringea, e ascrivibile quindi piuttosto a una differenza tra consonanti tese e consonanti rilassate, simile a quanto già noto per lingue dell'Estremo Oriente (ad es. coreano, giavanesse, cinese cantonese). E' inoltre da sottolineare come la presenza di un grado consonantico teso (causa tra l'altro molto spesso di un innalzamento del pitch nella vocale successiva; vd. Hombert – Ohala – Ewan 1979) in tali sistemi linguistici comporti una

discussione della classificazione dei tipi di emissione vocale generalmente in uso (per una sintesi vd. ad es. Laver 1994: 184-201) in cui l'aumento della tensione laringea è associato alla glottidalizzazione, e quindi a una discesa del pitch (una riflessione su questo tema è sviluppata in Henderson 1977; Gordon – Ladefoged 2001).

Si proporrà poi un confronto tra la situazione del consonantismo armeno e il processo di lenizione delle occlusive sorde intervocaliche caratterizzante il consonantismo dell'italiano centro-meridionale (per uno studio acustico sulla varietà romanesca vd. ora Marotta 2005), col fine di indagare le somiglianze e le differenze, rilevabili sul piano quantitativo ed esprimibili mediante parametri acustici, tra fenomeni imputabili alla tensione laringea e fenomeni tradizionalmente ascritti alla forza consonantica. In generale si mostrerà come una differenza consonantica presente in due sistemi linguistici, per la quale nelle rispettive tradizioni grammaticali si fa uso di coppie di termini che rimandano al significato di 'sordo' e 'sonoro', sia implementata, dal punto di vista fonetico, lungo dimensioni diverse tra loro.

Riferimenti bibliografici

- Cao, J. - Maddieson, I. (1993), "An exploration of phonation types in Wu dialects of Chinese". *Journal of Phonetics*, 20: 77-92.
- Fant, G. (1977), "Introductory remarks – Laryngeal mechanism and features". *Phonetica*, 34: 252-55.
- Gordon, M. – Ladefoged, P. (2001), "Phonation types: a cross-linguistic overview". *Journal of Phonetics*, 29: 383-406.
- Hacopian, N. (2003), "A three-way VOT contrast in final position: data from Armenian". *Journal of International Phonetic Association*, 33: 51-80.
- Hanson, H.M. – Stevens, K.N. – Kuo, H.-K.J. – Chen, M.Y. – Slifka, J. (2001), "Towards models of phonation". *Journal of Phonetics*, 29: 451-80.
- Henderson, E.J.A. (1977), "The larynx and language: a missing dimension?". *Phonetica*, 34: 256-63.
- Hombert, J.-M. – Ohala, J.J. – Ewan, W.G. (1979), "Phonetic explanations for the development of tones". *Language*, 55: 37-58.
- Jakobson, R. – Fant, G. – Halle, M. (1952), *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates*. Cambridge: MIT Press.
- Kenstowicz, M. (1994), *Phonology in Generative Grammar*. Cambridge (Ma): Blackwell.
- Ladefoged, P. – Maddieson, I. (1996), *The Sounds of the World Languages*. Malden (Ma): Blackwell.
- Laver, J. (1994), *Principles of Phonetics*, Cambridge, Cambridge University Press.
- Marotta, G. (2005), "Il consonantismo romano. Processi fonologici e aspetti acustici". In *Italiano parlato. Analisi di un dialogo*, a cura di F. Albano Leoni, R. Giordano. Napoli: Liguori: 1-24.
- Ní Chasaide, A. – Gobl, C. (1997), "Voice source variation". In Hardcastle, W.J. – Laver, J. (eds.) *The Handbook of Phonetic Sciences*. Oxford: Blackwell: 427-61.
- Stevens, K.N. (1977), "Physics and laryngeal behaviour and larynx modes". *Phonetica*, 34: 264-79.
- Stevens, K.N. (2000), *Acoustic Phonetics*. Cambridge (Ma): MIT Press.
- Vaux, B. (1998), *The Phonology of Armenian*. Oxford: Clarendon Press.

‘DOMAIN INITIAL STRENGTHENING’: DATI ACUSTICI ED ARTICOLATORI RELATIVI A DUE VARIETA’ DI ITALIANO

Barbara Gili Fivela, Sonia D'Apolito, Antonio Stella, Francesco Sigona
Università del Salento
{barbara.gili, sonia.dapolito, antonio.stella, francesco.sigona}@unisalento.it

In italiano la durata consonantica svolge una funzione contrastiva, benché diversi studi evidenzino l'esistenza di differenze diatopiche relative alla durata delle consonanti. Nell'italiano standard e in molte varietà l'opposizione di lunghezza svolge un ruolo distintivo [Lepschy, 1964]. Tuttavia alcune varietà parlate nella parte settentrionale della penisola non mostrano l'opposizione di lunghezza per le consonanti (cfr. Devoto [1972]; Loporcaro [1992]). A parte il fatto che i parlanti di queste regioni possono utilizzare il contrasto di lunghezza quando parlano italiano, seppur nella varietà regionale, piuttosto che il dialetto [Zmarich & Gili Fivela, 2005; Gili Fivela *et al.* 2007], si riscontrano altre peculiarità. In alcune varietà, soprattutto nelle regioni centrali e meridionali, le consonanti sono lunghe anche in contesti nei quali altri parlanti italofoeni normalmente realizzano consonanti scempie, come ad esempio in posizione iniziale di parola. Inoltre nelle regioni meridionali, l'uso pervasivo di segmenti lunghi in posizione iniziale di parola si osserva in particolare per consonanti specifiche, come l'occlusiva bilabiale sonora /b/ [Romano, 2003].

La lunghezza consonantica è qui considerata come esempio di rafforzamento, sia dal punto di vista acustico che dal punto di vista articolatorio. Precedenti lavori sulle consonanti geminate in posizione interna di parola hanno evidenziato che la loro durata acustica è circa il doppio della durata delle scempie e che il loro gesto di apertura ha una durata ed un'ampiezza maggiore rispetto al gesto di apertura delle consonanti scempie [Zmarich & Gili Fivela, 2005; Gili Fivela & Zmarich, 2005]. Le consonanti possono anche essere 'rafforzate' in base a fattori come la posizione nella frase e nella struttura prosodica. Molti lavori in letteratura hanno mostrato che le consonanti all'inizio dei domini prosodici hanno caratteristiche acustiche ed articolatorie che indicano la presenza di rafforzamento [Fougeron & Keating, 1997; Byrd *et al.* 2000]. Obiettivo della nostra ricerca è osservare cosa accade in termini di *domain initial strengthening* (rafforzamento all'inizio del dominio) per parlanti di varietà che utilizzano diversamente la durata consonantica.

Obiettivo del presente lavoro è studiare questi aspetti per mezzo di analisi acustiche ed articolatorie relative alla consonante bilabiale sonora /b/ prodotta all'inizio di domini di vari livelli da quattro parlanti di italiano: due parlanti leccesi, i quali dovrebbero produrre consonanti lunghe in posizione iniziale di parola, e due parlanti torinesi, i quali, nella stessa posizione, dovrebbero invece produrre consonanti scempie. Le consonanti bersaglio sono inserite all'inizio di domini prosodici di varia lunghezza (parola, sintagma intonativo, enunciato, ed enunciato in posizione iniziale di segmento di discorso), ottenute modificando la struttura sintattica (ad es., ramificata vs non ramificata), il peso prosodico dei costituenti (ad es., numero di parole), e la posizione nella struttura informativa (ad es., posizione all'interno del capoverso/segmento di discorso).

I soggetti sono stati registrati mentre leggevano ad alta voce due *corpora*. Nel primo corpus, la consonante bersaglio era inserita in frasi, secondo le seguenti condizioni:

3. Condizione di controllo: la consonante era in posizione intervocalica

in frasi atte ad indurre i soggetti a realizzare una consonante geminata o una scempia - es., *a battere* vs. *abbattere* (una coppia di pseudoparole è stata inserita per permettere il confronto con il resto del materiale – ad es., *ibà* vs *ibbà*).

4. La consonante era in posizione iniziale di parola e di frase (#CV, con C=/b/ e V=/a/; ad es. *battere*).
5. La consonante era in un contesto V#CV (/i/#/ba/) all'interno di un sintagma in cui l'unico confine presente era quello di parola.
6. La consonante era in un contesto V#CV (/i/#/ba/) in cui il confine di parola corrispondeva ad un confine sintattico SN/SV che ci si aspettava fosse realizzato come confine di sintagma prosodico intermedio.
7. La consonante era inserita in un contesto V#CV (/i/#/ba/) in cui il confine di parola corrispondeva ad un confine tra principale e subordinata, normalmente realizzato come confine intonativo.

Il secondo corpus era costituito da un discorso composto da tre capoversi; ognuno di essi conteneva enunciati che avevano la consonante bersaglio in posizione iniziale (v. punto 2) ed erano inseriti in diversi punti del capoverso.. Infatti, all'interno di un discorso è possibile rilevare una struttura di argomenti e alcuni lavori in letteratura mostrano come, nella produzione del parlato, il grado di 'sforzo', che varia in base alla posizione dell'enunciato nel discorso [Gussenhoven, 2002], abbia un impatto sui gesti articolatori. In questo modo si è quindi avuta la possibilità di osservare:

6) la produzione della consonante in posizione iniziale di enunciato, in frasi che si trovino in posizione iniziale mediana e finale di un segmento di discorso (quindi in contesti #CV, V#CV e V#CV).

Ai parlanti è stato richiesto di leggere i *corpora* per 10 volte a velocità di eloquio normale. I dati acustici ed articolatori sono stati raccolti presso il centro di ricerca CRIL dell'Università del Salento, con un sistema che permette l'acquisizione simultanea di materiali acustici e di dati ricavati per mezzo di un articolografo (AG500-EMA) e di un ecografo (Toshiba Aplio). In questa sede, saranno discussi principalmente i dati acustici, con alcuni riferimenti ai dati EMA ricavati in corrispondenza del sensore posto sul dorso della lingua, sull'asse sagittale mediano, e sul labbro inferiore [Gili Fivela *et al.*, 2008; in revisione].

I materiali sono stati analizzati uditivamente da tre degli autori, al fine di verificare la reale presenza ed entità del confine prosodico, e poi sono stati ulteriormente studiati, effettuando le seguenti misurazioni:

- durate acustiche: consonante bersaglio, sillaba che la include, vocale/i adiacente/i.
- dati articolatori ricavati con EMA: i sensori posti sul dorso della lingua e sul labbro inferiore sono stati considerati per misurare la durata del gesto consonantico, la relazione di fase tra la consonante e la vocale seguente e, in caso di contesti V#CV, la relazione di fase tra le vocali adiacenti alla consonante bersaglio.

I risultati preliminari (si veda anche [Gili Fivela *et al.*, 2008, in revisione]) mostrano che, dal punto di vista articolatorio, i parlanti realizzano diversi livelli di *initial strengthening* in base alla presenza dei confini prosodici considerati e che le differenze osservate a seconda del parlante non sembrano indicare strategie che dipendano dalla varietà di italiano parlata.

Riferimenti bibliografici

Byrd D., Kaun A., Narayanan S., Saltzman E. (2000). Phrasal signatures in articulation. In Broe, M. B., Pierrehumbert, J. B. (ed.). *Papers in Laboratory Phonology V*, Cambridge: Cambridge University Press, pp. 70 - 87.

Devoto G. (1972). *Scritti minori III*. Le Monnier, Firenze.

Fougeron, C., Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. In *Journal of Acoustical Society of America*, 101 (6), pp. 3728 - 3740.

Gili Fivela B., Zmarich C., Perrier P., Savariaux C., Tisato G. (2007). Acoustic and kinematic correlates of phonological length contrast in Italian consonants. In *Proceedings of the XVI International Conference of Phonetic Sciences (ICPhS'07)* ISBN:978-3-9811535-0-7, Saarbrücken, Germany, 6-10 agosto 2007, pp.469-472.

Gili Fivela B., Stella A., D'Apolito S., Sigona F. (in revisione). The timing of articulatory gestures at prosodic boundaries. In *Motor Control*, Special Issue on Speech motor control, edited by S.Fuchs, R.Sock, Y.Laprie.

Gili Fivela B., D'Apolito S., Stella A., Sigona F. (2008). Domain Initial Strengthening in sentences and paragraphs: preliminary findings on the production of voiced bilabial plosives in two varieties of Italian. In Sock R., Fuchs S., Laprie Y. (ed.), *Proceedings of the International Seminar on Speech Production*, Strasbourg France, 8-12 December, ISBN 2-905267-63-1, pp.205-208.

Gussenhoven C. (2002). Intonation and interpretation: Phonetics and phonology. In B.Bel, I.Marlien (ed.), *Proceedings of the Speech Prosody 2002 Conference*, Aix-en-Provence: Laboratoire Parole et Langage, pp. 47-57.

Lepschy G. C. (1964). Note sulla fonemica italiana. In *L'Italia Dialettale*, 27, pp.53-67 (1978, *Saggi di Linguistica Italiana* Bologna, Il Mulino, pp.63-75)

Loporcaro, M. (1992). On the analysis of geminates in Standard Italian and Italian dialects. In Hurch, B., Rhodes, R. A. (ed.), *Natural Phonology: The State of the Art*, Trends in Linguistics, Mouton de Guyter, Berlin, pp.153-187.

Romano A. (2003). Indici acustici di alcune geminate iniziali salentine. In Marotta G., Nocchi N. (ed.), *Atti delle Giornate di Studio del Gruppo di Fonetica Sperimentale*, pp.233-241.

Zmarich C., Gili Fivela B. (2005). Consonanti scempie e geminate in italiano: studio cinematico e percettivo dell'articolazione bilabiale e labiodentale. In *Atti del Convegno Nazionale AISV "Misura dei parametri"*, Padova, dicembre 2004, pp. 429-448.

LA PERCEZIONE DELLA INTERROGATIVA GLOBALE NELL'ITALIANO REGIONALE DI PARMA

Maria Chiara Felloni e Daniele Avesani
Univesrità di Modena e Reggio Emilia
mcfelloni@libero.it; daniele.avesani@yahoo.com

All'interno del quadro metodologico messo a punto nell'ambito del progetto AMPER (*Atlas Multimédia Prosodique de L'Espace Roman*, cfr. Lai *et al.* 1997, Contini *et al.* 2002a e b) è stata condotta un'indagine sulla (produzione e) percezione della interrogativa globale nell'italiano regionale di Parma. Il metodo utilizzato, di approccio teorico sovrapposizionale (Rossi 2001), prevede la definizione di un prototipo dell'intonazione affermativa e interrogativa globale della varietà di lingua indagata attraverso la raccolta di parlato di laboratorio basato su un corpus fisso di frasi fuori contesto (costruite secondo restrizioni di ordine fonetico-accentuale e sintattico), elicitate tramite richiesta indiretta e per immagini. Essendo il metodo AMPER (o modello grenobliano) un modello ibrido acustico-percettivo, esso prevede la validazione del prototipo con test di identificazione (nonché la validazione sul parlato spontaneo). Per quanto riguarda il profilo melodico della interrogativa globale nella varietà di italiano regionale di Parma, è stato individuato un intonema di tipo ascendente-discendente (con *turnig point* in corrispondenza dell'ultima vocale accentata) che si estende in corrispondenza della frequenza media del parlante, terminandone al di sotto; l'ultima vocale accentata presenta una spezzatura bitonale, di tipo discendente-ascendente, con tre allotoni principali, ugualmente riconosciuti dai parlanti come tipici della varietà (spezzatura "ascendente", "a gomito" e "a v"). Il profilo del primo sintagma nominale, invece, presenta un movimento ascendente-discendente quasi topicalizzato, dato dal raggiungimento dei valori di F0 minimi di frase al confine di sintagma, indipendentemente dalla struttura accentuale. La configurazione tipica individuata rimane tale indipendentemente sia dalla struttura accentuale a cui si sovrappone (anche se di essa risente in parte, poiché si allunga nei proparossitoni e si tronca negli ossitoni), che dalla presenza di espansioni aggettivali nel secondo sintagma nominale. Dall'osservazione oggettiva si nota che la strategia dell'espressione della modalità utilizzata dalla varietà in oggetto si concentra non solo sul contorno terminale di modalità (CTM), ma anche sul primo sintagma intonativo (SN1), mentre le due curve si sovrappongono in corrispondenza del gruppo verbale. Per quanto riguarda il parametro della durata, è emerso che anch'esso collabora alla realizzazione dell'intonazione, e che resta il principale responsabile della realizzazione dell'accento di parola e di frase. La batteria di test percettivi messi a punto per la valutazione della modalità mostra, tuttavia, risultati in parte discordanti con l'osservazione oggettiva dei dati: se gli uditori riconoscono nella quasi totalità dei casi la modalità interrogativa in frasi sintetiche anche di varietà linguistiche diverse dalla propria, e la parte determinante per il riconoscimento della modalità resta essenzialmente il contorno terminale, solo una bassa percentuale di uditori è stata in grado di identificare correttamente il profilo di SN1 pronunciato in isolamento (cfr. Romano 1998a e b). I test sulla modalità, messi a punto con Praat (Boersma/Weenink 2005), sono composti da tre fasi: una fase di addestramento (in cui all'uditore è chiesto di riconoscere la modalità affermativa o interrogativa di enunciati sintetici di sola prosodia di diverse varietà di italiano) seguita da un test di identificazione con stimoli ibridi (creati a partire da una base di una determinata modalità sulla quale è stato innestato o il contorno terminale o l'intero sintagma nominale, SN2, cioè le ultime tre sillabe, dell'altra modalità) e da un ultimo test

identificativo in cui all'uditore è chiesto di riconoscere una parte di enunciato come l'inizio o la fine di una frase affermativa o interrogativa. I test percettivi condotti invece sulla diatopia (cfr. Interlandi 2004) hanno mostrato, in un'elevata percentuale dei casi, la capacità degli uditori di riconoscere il modello prototipico dell'interrogativa globale individuato per la varietà in questione per tutte le strutture sintattiche, discriminandolo da altre varietà di italiano (a conferma della validità del metodo d'indagine utilizzato). Inoltre, dai risultati dei test si evince che la durata dell'ultima vocale accentata è ritenuta dai parlanti un importante indice geoprosodico, insieme alla struttura prosodica complessa realizzata sul contorno terminale di modalità (calcolata in semitoni su millisecondi). I test percettivi sulla diatopia si compongono di due fasi: una fase di addestramento (in cui l'uditore discrimina tra stimoli sintetici di diverse varietà di italiano) e una fase in cui l'uditore deve riconoscere come appartenenti alla comunità linguistica in oggetto stimoli modificati dal ricercatore. Questi ultimi, aventi la finalità di rilevare l'importanza dei singoli parametri (F0 e durata) e dei punti salienti della curva nel veicolare le informazioni diatopiche, sono stati creati a partire dalla base di uno stimolo interrogativo globale prototipico della varietà parmigiana sul quale, di volta in volta, sono stati innestati valori di durata e di F0 (prima singolarmente e poi insieme), su ciascun segmento vocalico di SN2 (prima singolarmente e poi sull'intero sintagma), tipici di altre varietà di italiano, sia adiacenti (Reggio Emilia, Modena e Piacenza) che geograficamente distanti (Torino). Il confronto tra la varietà di Parma e quella di Torino è risultato di grande interesse, poiché esse presentano, per l'interrogativa globale, profili intonativi di F0 opposti lungo tutta la curva e sovrapposti nel contorno terminale di modalità, in cui normalmente le diverse varietà di italiano massimizzano le differenze. Dalla somministrazione delle produzioni sintetizzate di tutti i 15 parlanti del campione differenziato sociolinguisticamente (per età, genere, grado d'istruzione e rete sociale) a un campione di uditori differenziato anch'esso (per età e per genere), è emerso, invece, che esiste un diverso prototipo intonativo per il gruppo dei giovani (18-30 anni) e per quello degli adulti (60-80 anni), e che, nella comunità in oggetto, la variabile diatopica si sovrappone a quella diastratica, come confermato dal correlato sociolinguistico di genere. Da questa ultima batteria di test, infatti, è emerso che la durata dell'ultima vocale accentata (con soglie che variano in base ai raggruppamenti sociolinguistici), la configurazione del contorno terminale di modalità e la struttura prosodica complessa possono essere ritenuti anche indici sociofonetici. Inoltre, è stato rilevato che parlanti con reti sociali strette mostrano, anche per il livello dell'intonazione, comportamenti conservativi, mentre i parlanti a reti larghe adottano modelli innovativi, che si sottraggono, con strategie diverse, al prototipo individuato per la varietà in oggetto. Per l'elaborazione degli stimoli sintetici dei test di percezione ci si è serviti di alcuni script per MATLAB® (The MathWorks, Inc.) e PRAAT (Boersma/Weenink 2005), frutto di sviluppo da zero (con inserimenti di porzioni di codice ad opera di Romano 1998a). Tra questi, lo strumento principale è costituito dallo script `Stim_Creator_V2.m`, il quale elabora i files di testo contenenti i valori numerici dei correlati acustici estratti dalle registrazioni audio, conformi al protocollo AMPER. In particolare, esso permette all'utilizzatore di modificare la F0 in ciascuno dei tre punti considerati per ogni segmento vocalico (secondo il modello grenobliano), direttamente in semitoni, secondo due diversi criteri (sostituendo i valori originari con altri arbitrari oppure variandone l'altezza di una quantità di semitoni scelta dall'utente) e la durata, anche in questo caso secondo due diverse modalità (sostituendo la durata originaria con un valore calcolato in percentuale sulla durata media totale dell'enunciato, o attraverso una percentuale di aumento o diminuzione rispetto al valore effettivo della durata del segmento scelto). Le modifiche della F0 e della durata possono

avvenire singolarmente o nel medesimo tempo. Lo script in oggetto permette inoltre di normalizzare gli stimoli sintetici per il confronto tra locutori e per la creazione dei test di percezione, riallineando la F0 media dello stimolo stesso su un valore scelto dall'utente e ricalcolandone i rapporti di variazione nell'enunciato. Stim_Creator_V2.m produce un file sonoro di sola prosodia (in formato wave, e quindi utilizzabile da altri applicativi per la composizione delle batterie di test, nonché di tipo "ton", come elaborato per la prima volta da Romano 1998a), e due files di testo, contenenti i valori di durata e quelli di frequenza della F0, questi ultimi rispettivamente in semitoni e in Hertz (nel formato conforme al protocollo AMPER e quindi riutilizzabili con lo stesso script per ulteriori modifiche). In conclusione, l'analisi sulla percezione della interrogativa globale di Parma ha mostrato la validità del metodo di analisi utilizzato (modello AMPER) e ha permesso di identificare, in parte confermando e in parte superando l'osservazione oggettiva, alcuni indici acustici come importanti veicoli di informazioni diatopiche e diastratiche, localizzate, in particolare, nel contorno terminale di modalità.

Riferimenti bibliografici

Boersma, P. / Weenink, D. (2005), *Praat: doing phonetics by computer*, (Version 4.3.14).

Contini, M./Lai, J. P./Romano, A. (2002a), "La géolinguistique à Grenoble: de l'ALiR à AMPER", in Simoni-Aurembou, M. K. (a cura di), *Nouveaux regards sur la variation diatopique*, Revue belge de Philologie et d'Histoire, vol. 80 n. 3, 931-941.

Contini, M./Lai, J. P./Romano, A./Roullet, S./de Castro Moutinho, L./Coimbra R. L./Pereira Bendiha U./Secca Ruivo S. (2002b), "Un project d'Atlas Multimédia Prosodique de l'Espace Roman", in Bel, B./Marlien, I. (a cura di), *Proceedings of the 1st International Conference on Speech Prosody*, Aix-en-Provence, 11-13 April 2002, Aubenas d'Ardèches, Lienhart, 227-230.

't Hart, J./Collier, R./Cohen, A. (1990), *A perceptual study of intonation. An experimental-phonetic approach to speech melody*, New York, CUP.

Interlandi, G.M. (2004), *L'intonazione delle interrogative polari nell'italiano parlato a Torino: tra varietà regionale e nuova koine*, Tesi di Dottorato, XV ciclo, Dipartimento di Linguistica Teorica e Applicata, Università di Pavia.

Interlandi, G.M. (2005), *L'italiano parlato a Torino. Varietà regionale o nuova koiné? L'intonazione delle interrogative polari*, "Géolinguistique", Hors série n.3, Centre de Dialectologie, Université Stendhal Grenoble 3, 217-240.

Lai, J. P. (2002), *L'Intonation dans le parler de Nuoro*, Tesi di Dottorato in Scienze del Linguaggio, Centre de Dialectologie, Université Stendhal Grenoble 3.

Lai, J.P. (2004), *Le sarde de Nuoro au sein de nouvel Atlas Multimédia Prosodique de l'Espace Roman (AMPER)*, "Geolinguistique", Hors série n.9, Centre de Dialectologie, Université Stendhal Grenoble 3, 22-45.

Lai, J.P./Romano, A./Roullet, S. (1997), *L'analisi prosodica di alcune varietà linguistiche d'Italia: proposta per un approccio metodologico unitario*, "Bollettino dell'ALI", 21, 23-70.

Romano, A. (1998a), *Analyse des structures prosodiques des dialects et de l'italien régional parlés dans le Salento (Italie): approche linguistique et instrumentale*, Tesi di Dottorato in Scienze del Linguaggio, Centre de Dialectologie, Université Stendhal Grenoble 3, 2 voll.

Romano, A. (1998b), *Accent et Intonation des parlers du Salento: une approche théorique et instrumentale*, "Géolinguistique", Hors série n. 7, Centre de Dialectologie, Université Stendhal Grenoble 3, 93-132.

Roullet, S. (1999a), *Accent et Intonation dans deux parlers francoprovençaux de la Vallée d'Aoste (Sarre et Cogne)*, Tesi di Dottorato in Scienze del Linguaggio, Centre de Dialectologie, Université Stendhal Grenoble 3, 2 voll.

Roullet, S. (1999b), *Accent et Intonation dans le parlers de deux villages valdôtains – Sarre et Cogne*, "Géolinguistique", Hors série n. 8, Centre de Dialectologie, Université Stendhal Grenoble 3 197-235.

Zucca, I. (2005), *L'intonazione del cagliaritano: il quartiere di Castello*, Tesi di Dottorato, XVII ciclo, Dipartimento di Linguistica Teorica e Applicata, Università di Pavia.

L'ABBASSAMENTO DI [ɛ] NELLE VARIETÀ DIALETTALI NORD-VESUVIANE: UN'ANALISI SOCIOFONETICA

Valentina Rètaro

Università di Napoli "Federico II"

valentinaretaro@virgilio.it

Il fenomeno dell'abbassamento di [ɛ] tonica è diffuso a macchia di leopardo in tutta la Campania, ad esclusione della città di Napoli, dove appare soltanto in forma latente (si veda Maturi e Sornicola 1990 e Sornicola e Maturi 1993). A differenza però di altri fenomeni dialettali che tendono a scomparire nelle produzioni più orientate verso il polo alto del continuum dialetto-lingua, l'abbassamento compare anche in enunciati italianizzanti. Il tratto è inoltre tra i più stigmatizzati in Campania. In questo studio si intende presentare un'analisi fonetica dell'abbassamento di [ɛ] tonica in cinque varietà dialettali campane di area peri-urbana. L'obiettivo che ci si pone è verificare la distribuzione areale degli esiti più aperti all'interno di cinque località dell'area nord-vesuviana, attraverso un confronto interlinguistico, e far emergere tratti di somiglianza o dissimilarità tra le varietà indagate.

Nell'area nord-vesuviana l'abbassamento di [ɛ] appare particolarmente diffuso e vitale, ma nella percezione dei parlanti esso è associato soprattutto agli abitanti di Somma. Si è voluto dunque verificare se il tratto fosse davvero presente soltanto in questa località e se presentasse caratteristiche del tutto peculiari che giustificassero in qualche modo la percezione dei parlanti. Ad una prima analisi impressionistica, inoltre, è emerso come il tratto sia del tutto assente nei parlanti di Ottaviano e San Giuseppe. Per comprovare in chiave acustica l'assenza del fenomeno in queste due ultime località e verificare se e in che misura le vocali fossero interessate anche da fenomeni di diversa natura rispetto alla prevista maggiore apertura (cfr. Sornicola e Maturi 1993) si è proceduto all'analisi di tutte le occorrenze del segmento vocalico in questione e non solo delle varianti che impressionisticamente risultavano più aperte.

Metodologia

L'area oggetto di questo studio è formata da cinque centri del versante interno del Vesuvio: Pollena Trocchia, Sant'Anastasia, Somma, Ottaviano e San Giuseppe. I dati provengono da un corpus di interviste di parlato dialettale, effettuate a 30 informatori di età superiore ai 45 anni (Rètaro 2009). Per questa analisi sono stati selezionati 10 informatori, 5 uomini e 5 donne. Dopo aver valutato i contesti potenziali di occorrenza del fenomeno, si è scelto di selezionare le sole parole che contenessero la vocale /ɛ/ seguita da laterale lunga (ad esempio parole come *bello*, *fratello*, *carrello*, *gemello*, *carrettella*) e da -nt- (es. avverbi in -mente) a prescindere dalla resa fonetica della vocale. In totale, sono state etichettate 200 parole e dunque 200 segmenti vocalici (40 per ciascun punto di inchiesta). Poiché il campione è costituito sia da donne che da uomini e poiché l'obiettivo è comparare i risultati ottenuti tra i 5 punti di inchiesta a prescindere dal sesso dell'informatore, si è reso necessario l'utilizzo di tecniche di normalizzazione dei dati di tipo *vowel-intrinsic* (cfr. Adank, Smits e van Hout 2004 e Calamai 2005). Tali procedure fanno uso di trasformazioni non lineari della scala di frequenza (come quella qui utilizzata in Bark) e di trasformazioni basate su combinazioni di frequenze formantiche, come quella qui utilizzata, basata sul calcolo della differenza F2-F1 (cfr. Calamai 2004). Inoltre, poiché la frequenza fondamentale sembra agire come fattore di normalizzazione, influenzando la percezione del timbro vocalico (cfr. Traunmüller 1981 e Johnson 1990, ma anche Di Benedetto 1991), si è

utilizzato anche quest'altro parametro. Di ciascuna parola è stato dunque selezionato il solo segmento vocalico tonico e sono stati misurati i valori di F1, F2 ed F3. È stata misurata poi la frequenza fondamentale (valore medio di F0 per ciascuna vocale) e la durata dell'intero segmento. Sono state poi calcolate le differenze F2-F1 in Hz e in Bark e F1-F0 in Hz e in Bark per ciascuna vocale, le medie di tali differenze e le deviazioni standard delle stesse. Il campione è stato mantenuto distinto sulla base della provenienza e del sesso. Per valutare la significatività statistica delle differenze riscontrate tra le 5 comunità nelle medie di F1-F0 è stata effettuata un'analisi della varianza (ANOVA).

Risultati

L'analisi acustica ha smentito l'impressione uditiva dell'assenza del fenomeno nei due centri di Ottaviano e San Giuseppe, evidenziando anche per queste località fenomeni di abbassamento di [ɛ]. Ciò che apparso più significativo nel confronto fra i cinque punti è la qualità fonetica raggiunta di volta in volta dalla vocale, più stabile nelle realizzazioni di Ottaviano e San Giuseppe e più variabile nelle altre tre località indagate. Tra tutti i centri, Somma mostra il maggior grado di abbassamento raggiunto dalla vocale. In questa località il tratto appare decisamente più frequente e macroscopico.

Differenze tra le cinque varietà concernono però non solo la resa fonetica del segmento, ma anche la distribuzione degli esiti nei diversi contesti e il rapporto tra posizione prosodica e durata. I dati hanno mostrato che la posizione del segmento vocalico all'interno dell'enunciato incide significativamente sulla sua resa fonetica soltanto per il punto di Somma, dove le vocali in fine di frase sono più aperte rispetto a quelle poste all'interno. A Ottaviano e Sant'Anastasia, dove "le forze che hanno prodotto un certo effetto hanno cessato di operare, lasciando però una "mutazione" stabile" (caso *b* di Sornicola e Maturi 1993: 93-95), l'abbassamento non correla con una durata maggiore, contrariamente a quanto accade a Somma, dove il fenomeno è presente in maniera vistosa ed è sensibile alla durata (caso *a* di Sornicola e Maturi 1993). Per gli altri due punti, San Giuseppe e Pollena, invece, il condizionamento della durata e della posizione è minimo, sebbene ancora presente.

Ciò che sembra dunque più interessante è non tanto la diversità degli esiti nelle realizzazioni fonetiche riscontrabile tra i cinque centri di indagine, quanto piuttosto la variabilità insita nel fenomeno stesso dell'abbassamento. Oscillazioni nelle realizzazioni delle vocali caratterizzano tutti i parlanti del corpus analizzato, ma in particolar modo esse sono presenti nelle vocali dei parlanti di Somma e di Pollena: a tendenze maggiori all'abbassamento sembrano pertanto corrispondere anche maggiori oscillazioni. Di contro, nei punti di Sant'Anastasia, San Giuseppe e Ottaviano, pur essendo presente una certa tendenza all'apertura della vocale, il fenomeno mostra una minore variabilità e un minore dinamismo interno.

Riferimenti bibliografici

Adank, P., Smits, R. e R. van Hout (2004), "A comparison of vowel normalization procedures for language variation research", in *Journal of the Acoustical Society of America*, 116, pp. 3099-3107.

Calamai, S. (2004), "Vocali fiorentine e vocali pisane a confronto", in *Il parlato Italiano, Atti del Convegno Nazionale* (acd) F. Albano Leoni, F. Cutugno, M. Pettorino, R. Savy, Napoli, 13-15 febbraio 2003, Napoli: D'Auria, CD-rom, E02, 1-25.

Calamai, S. (2005), "Intrinsic Vowel Normalization: Comparing Different Procedures (Data from Tuscan Italian)", in *Italian Journal of Linguistics/Rivista di Linguistica*, 17.2, pp. 211-270.

Di Benedetto, M. G. (1991), "Complex relation between F1 and F0 in determining vowel height: acoustic and perceptual evidence", in *Studi Italiani di Linguistica Teorica ed Applicata*, 20, pp. 579-603.

Johnson, K. (1990), "Contrast and normalization in vowel perception", in *Journal of Phonetics*, 18, pp. 229-254.

Maturi, P. e R. Sornicola (1990), "Un indicatore sociolinguistico in area napoletana", in E. Banfi e P. Cordin (acd), *Storia dell'italiano e forme dell'italianizzazione*, Atti del XXIII Congresso della SLI (Trento-Rovereto 18-20 maggio 1989), pp. 411-425, Roma: Bulzoni.

Rètaro, V. 2009, *Dinamiche dialettali in area nord-vesuviana*, Tesi di dottorato presentata e discussa presso l'Università degli Studi di Torino, a.a. 2005-2008.

Sornicola, R. e P. Maturi (1993), "Un modello epidemiologico del cambiamento linguistico: dinamica di una micro-variazione fonetica in Campania", in G. Holtus, e E. Radtke (1993) (acd), *Sprachprognostik und das Italiano di domani*, pp. 59-98, Tübingen: Gunter Narr.

Traunmüller, H. (1981), "Perceptual dimension of openness in vowels", in *Journal of the Acoustical Society of America*, 69, pp. 1465-1475.

WORKSHOP “GRUPPO DI FONOLOGIA E LINGUISTICA”

P. M. Bertinetto, I. Ricci, Z. Na *Le nasali sorde dell'ayoreo: prime prospezioni*

G. Bocci, C. Avesani *Givenness, deaccentazione e il ruolo di L* nell'italiano di Toscana*

M. Grimaldi, A. Calabrese, F. Sigona, B. Sisinni, L. Garrapa *La variazione micro-parametrica nei processi di armonizzazione del Salento meridionale. Studio articolatorio tramite Articolografo Elettromagnetico 3D ed Ultrasuoni*

F. Cangemi, R. Delucchi, M. Loporcaro, S. Schmid *Vocalismo finale atono "toscano" nei dialetti del Vallo di Diano (Salerno)*

L. Mereu *Variazione, universali linguistici e contributo dell'intonazione*

A. Romano, F. Gambino *Cacuminali calabresi: modi e luoghi d'articolazione alla luce di misurazioni acustiche e indagini per risonanza magnetica (IRM)*

B. Sisinni *Metodologie negli studi cross-linguistici. Riflessioni sulla relazione fra dati acustici e percettivi*

A. Uguzzoni *Invarianza e variabilità, ieri e oggi. Spunti per una discussione a più voci*

P. Mattana, A. Paoloni, M. Grimaldi *Variazione dei parametri neurolinguistici nell'unspoken speech*

A. De Dominicis *Variazione nelle categorie intonative*

LE NASALI SORDE DELL'AYOREO: PRIME PROSPEZIONI

Pier Marco Bertinetto[^], Irene Ricci[^], Zhi Na

[^]Scuola Normale Superiore

PierMarco.Bertinetto@sns.it, i.ricci@sns.it

L'ayoreo è parlato in un territorio che sta a cavallo fra Bolivia e Paraguay.

Appartiene alla famiglia zamuco, di cui fa parte una sola altra lingua, il chamacoco.

Entrambe esibiscono una serie di consonanti nasali sorde. Nello studio che si intende presentare, vengono analizzate le realizzazioni di parlanti ayoreo; i materiali chamacoco, recentemente acquisiti, saranno oggetto di successive indagini.

La presenza di una serie di fonemi affini è segnalata anche in resigaro, lingua arawak parlata in Perù (e gravemente minacciata di scomparsa). Casi analoghi sono segnalati anche in lingue austronesiane, come il drehu (Isole Loyauté) e lo iaai (Nuova Caledonia). Anche i materiali relativi al resigaro ed al drehu (ottenuti tramite la collaborazione di colleghi stranieri) saranno oggetto di successivi studi. Al momento non è dunque possibile stabilire il grado di similarità circa le modalità di realizzazione di tali suoni nelle lingue citate: questo sarà appunto lo scopo dell'indagine comparativa che si intende condurre. In effetti, benché tutti questi suoni siano interpretabili, fonologicamente, come nasali sorde, foneticamente essi potrebbero differire in maniera anche piuttosto marcata.

Nella grafia, i suoni in analisi sono convenzionalmente resi come <jn>, <jm> e <jñ> in ayoreo, <hn>, <hm> e <hñ> in chamacoco. Benché ciò faccia pensare ad una prima fase consistente in una fricativa laringale, è più ragionevole interpretarla come una fase di desonorizzazione della nasale. Resta tuttavia da capire se la doppia natura di questi suoni costituisce un gesto coscientemente mirato – una sorta di “affricata nasale” al rovescio, tanto per intenderci – ovvero se si tratta di un effetto secondario non intenzionale, meramente dovuto alla mancata coordinazione dei gesti articolatori. Se così fosse, si tratterebbe semplicemente della comparsa anticipata dell'attività laringale in rapporto al momento di rilascio della nasale, ossia di una precoce attivazione delle corde vocali in preparazione della vocale seguente.

Le analisi attualmente in corso hanno evidenziato l'esistenza di almeno due strategie di realizzazione di questi suoni in ayoreo. In certi casi, la desonorizzazione riguarda l'intera durata della nasale; in altri, si possono nettamente distinguere due fasi. Ciascuno dei due tipi potrebbe costituire il gesto “bersaglio”, con l'altro a fungere da variante allofonica. Ma ovviamente, l'interpretazione complessiva differirebbe, a seconda che il bersaglio sia costituito da nasali sorde o da nasali predesonorizzate. Un possibile criterio per orientarsi è costituito dalla durata della fase sonora all'interno della consonante nasale. Ove questa sia (in prevalenza) di breve durata, sarebbe lecito supporre che si tratti di un occasionale imperfetto coordinamento motorio fra articolatori orali e laringali; ove invece la durata sia sufficientemente ampia, non potrebbe che trattarsi di un gesto consapevole.

Le analisi vengono condotte tenendo conto di vari fattori, al fine di valutarne

l'influsso comparato: identità del locutore, parola isolata vs. parola in contesto, punto di articolazione (bilabiale, alveolare, palatale), posizione nella parola (iniziale vs. noniniziale).

GIVENNESS, DEACCENTAZIONE E IL RUOLO DI L* NELL'ITALIANO DI TOSCANA

Giuliano Bocci* e Cinzia Avesani°

Università di Siena*, ISTC-CNR°

giulianobocci@gmail.com, cinzia.avesani@pd.istc.cnr.it

La nozione di *Givenness* formulata da Schwarzschild (1999) ha avuto una crescente fortuna nei recenti modelli della struttura prosodica. Gli approcci radicalmente riduzionistici come quello di Bader (2001) propongono che gli effetti prosodici collegati al focus siano in realtà indotti da *Givenness*, negando la legittimità stessa della nozione di focus (si veda anche Sauerland 2004). Al contrario, Féry e Samek-Lodovici (2004) e Selkirk (2008) propongono che sia la nozione di focus che quella di *Givenness* interagiscano con le regole di conversione sintassi-prosodia di *default* nel determinare la struttura prosodica e la distribuzione dei accenti intonativi (PA).

Se, da un lato, la deaccentazione di materiale dato (*given*) è massicciamente attestata nelle lingue germaniche (cfr. Bolinger, 1968; Ladd, 1996; German et al., 2004), dall'altro è stato ripetutamente osservato che le lingue romanze non esibiscono un consistente processo di deaccentazione dell'informazione data (tra gli altri, Ladd 1996, Zubizarreta 1998). In un esperimento comparativo, Swerts et al. (2002) mostrano che mentre in olandese l'informazione data all'interno di sintagmi nominali deve essere deaccentata, in italiano è sistematicamente accentata. A corroborare questo risultato sperimentale, Avesani e Vayra (2005) mostrano che in un corpus di italiano parlato (semi) spontaneo, elicitato tramite la tecnica del Map Task, l'accentazione di materiale dato è un fenomeno pervasivo a tutti i livelli di costituenza.

Tuttavia, l'italiano sembra apparentemente richiedere deaccentazione per taluni elementi: elementi emarginati (nel senso di Antinucci e Cinque, 1997), topic dislocati a destra e background in frasi con focus ex situ. In questo lavoro, concentrandoci sulla varietà di Italiano parlato a Siena, forniremo un'analisi di questi casi mostrando che 1) la rappresentazione fonologica di questi apparenti casi di deaccentazione deve essere trattata come associazione con un accento intonativo L*, piuttosto che come assenza di prominenza intonativa e metrica; 2) L* non è assegnato come il correlato intonativo di *Givenness*, ma l'associazione di L* è non-localmente determinata da regole strutturali sulla rappresentazione prosodica.

Presenteremo i risultati di un esperimento di produzione progettato per identificare il ruolo di *Givenness* e del focus nella rappresentazione prosodica dell'enunciato. I dialoghi (46 dialoghi letti da 4 soggetti senesi x 5-8 ripetizioni) sono stati strutturati in modo da indurre differenti strutture focali (focus di nuova informazione e contrastivo) in combinazione con topic dislocati a destra e a sinistra in differenti ordini lineari.

I risultati indicano chiaramente che ciò che precede il focus è accentato indipendentemente dalla nozione di *Givenness*. Gli accenti intonativi su materiale dato non possono essere analizzati quali accenti prenucleari di tipo "eufonico" (cfr. Selkirk, 2008), poiché possono configurarsi anche come accenti nucleari. Dopo l'ultimo accento intonativo associato con il costituente in focus, tuttavia, il contorno intonativo è invariabilmente basso e piatto fino alla fine dell'enunciato, indipendentemente dal numero, dalla funzione e dalla lunghezza dei costituenti coinvolti. Sebbene il contorno intonativo in contesto post-focale sia del tutto piatto, il materiale post-focale non può essere considerato come extrametrico.

Bocci (2009) mostra ad esempio che il materiale post-focale è scandito in costituenti prosodici (e soggetto ad allungamenti pre-confine) ed è portatore di prominenze metriche di livello frasale, la cui presenza induce effetti di iper-articolazione localizzata (effetti di durata, traiettorie formantiche più estreme) ed enfasi spettrale. Questo quadro contrasta crucialmente con quanto proposto per l'inglese, tra gli altri, da Féry & Samek-Lodovici (2004), secondo cui il materiale dato e deaccentato non può essere metricamente prominente.

Per rendere conto della rappresentazione fonologica del materiale post-focale nell'italiano di toscana, proponiamo che ad esso sia associato un accento intonativo L*, la cui assegnazione in questa varietà è limitata ai contesti post-focali (per una diversa interpretazione di L*, si veda Marotta e Sardelli 2003). Questa proposta permette innanzitutto di mantenere una piena congruenza tra rappresentazione metrica e rappresentazione intonativa. La congruenza tra i due livelli andrebbe infatti persa se le prominenze frasali metriche post-focali rimanessero senza un'interpretazione intonativa.

Assumere L* permette inoltre un'analisi unitaria delle diverse varietà romanze. Mentre in portoghese europeo (Frota 2000) e in molte varietà di Italiano (D'Imperio, 2000; Grice et al. 2005) il contorno post-focale è caratterizzato da speciali accenti compressi, in toscano tale contorno è invariabilmente basso e piatto. La nostra proposta permette di mantenere una comune rappresentazione sottostante tra differenti varietà italiane e romanze, riducendo le differenze superficiali ad una scelta linguo-specifica del tipo di accento da associare in contesto post-focale.

Così come il fenomeno della deaccentazione in Inglese e quello dell'associazione di L* in toscano differiscono sul piano della rappresentazione fonologica, allo stesso modo, proponiamo che differiscano per il loro ruolo nella grammatica. Mentre la deaccentazione in inglese è correlato prosodico di *Givenness*, L* in italiano di toscana non è, secondo la nostra analisi, il correlato di alcuna proprietà informazionale, sia essa *Givenness*, *background* o "*tail*" (Vallduvì, 1992). Mostreremo infatti che L* è assegnato unicamente in funzione della posizione superficiale del focus. Secondo la nostra proposta, L* è impiegato come una marca per identificare il lato destro del costituente focale definendo il materiale alla sua destra come non focale, in ottemperanza con la "regola di identificazione del focus" ("FDR", Bocci e Avesani, 2008; Bocci, 2009). Tale regola rappresentazionale richiede, in breve, che l'accento associato al focus sia l'accento intonativo più a destra capace di esprimere focus (dove L* è inadatto ad esprimere focus in toscano così come !H+L* lo è in portoghese europeo).

Ad esempio, coppie minime di frasi con focus in situ ed ex situ (elicitate nello stesso contesto) presumibilmente condividono la stessa articolazione in focus-background. Tuttavia, le proprietà prosodiche associate al background radicalmente differiscono nei due casi: il materiale pre-focale nel primo caso non si associa mai con L*, mentre L* è necessariamente assegnato al *background* che segue il costituente in focus. Allo stesso modo, mostreremo che i topic dislocati a destra possono genuinamente precedere l'occorrenza del costituente in focus. Sebbene coerentemente caratterizzati dalle proprietà di *tail*, i topic dislocati a destra in questa condizione, non possono associarsi con L*, ma al contrario diventano portatori di accenti nucleari prominenti (H+L*).

Discuteremo infine alcune implicazioni della nostra analisi. In primo luogo, se la proprietà di *Givenness* non è attiva nel sistema prosodico dell'Italiano, non è plausibile concludere che la nozione di focus possa essere rimpiazzata da quella di *Givenness*. In secondo luogo, secondo la nostra proposta, la selezione di L* non è localmente guidata dalle proprietà degli elementi a cui si associa (*contra* Pierrehumbert e Hirschberg, 1990),

ma dipende da requisiti strutturali sulla rappresentazione prosodica. Ne segue quindi l'associazione/interpretazione degli accenti intonativi non può essere diretta e immediata, ma deve far riferimento alla rappresentazione strutturale in cui sono inseriti.

Riferimenti bibliografici

- Antinucci, F. e G. Cinque (1977) "Sull'ordine delle Parole in Italiano: L'emarginazione," *Studi di grammatica italiana* 6, 121-146.
- Avesani, C. e Vayra, M. (2005). Accenting, deaccenting and information structure in Italian dialogues. In L. Dybkjaer & Minker W. (eds), 6th SIGDial Workshop, 19-24.
- Avesani, C. and Vayra, M. (2003). "Broad, narrow and contrastive focus in Florentine Italian". In Solé M.J et. al. (eds.), *Proceedings of the 15th ICPHSS*, Barcelona, 1803-1806.
- Bader, C. (2001) *Givenness, Focus, and Prosody*, Ph.D. dissertation, MIT.
- Bocci, G. (2009). *On syntax and prosody in Italian*, Ph.D. dissertation, University of Siena.
- Calabrese, A. (1992) "Some informal remarks on Focus and Logical Structures in Italian." *Harvard Working Papers in Linguistics*, vol. I, 91-127.
- D'Imperio, M.P. (2000). *The Role of Perception in Defining Tonal Targets and their Alignment*, PhD dissertation, Ohio State University
- Féry, C. and Samek-Lodovici, V. (2006). "Focus Projection and Prosodic Prominence in Nested Foci", *Language*, Vol. 82, 1.
- Féry, C. and V. Samek-Lodovici, (2006) "Focus Projection and Prosodic Prominence in Nested Foci," *Language* 82.
- Frota, S. (2000). *Prosody and Focus in European Portuguese*, New York, Garland.
- German, J., Pierrehumbert, J. e Kaufmann, S. (2006). Evidence for phonological constraints on nuclear accent placement, *Language* 82(1), 151-168.
- Grice M., D'Imperio M.P., Savino M. e Avesani C. (2005). "Towards a strategy for ToBI labelling varieties of Italian". In Sun-Ah Jun (ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford, OUP, 362-389.
- Ladd R. (1996). *Intonational Phonology*, Cambridge, Cambridge University Press.
- Marotta, G., e Sardelli E. (2003). Sulla prosodia della domanda con soggetto postverbale in due varietà di italiano toscano (pisano e senese). In P. Cosi et al. (eds), *Voce, canto, parlato. Studi in onore di F. Ferrero*. 205-212. Padova: Unipress.
- Pierrehumbert J. e Hirschberg J. (1990). "The meaning of intonational contours in the interpretation of discourse". In P.R. Cohen, J. Morgan and M.E. Pollack (eds.), *Intentions in communication*, Cambridge, Mass, MIT Press, 271-311.
- Sauerland, U. (2005) "Don't Interpret Focus: Why a Presuppositional Account of Focus Fails, and How a Presuppositional Account of Givenness Works," *Proceedings of Sinn und Bedeutung* 9.
- Schwarzschild, R. (1999). "Givenness, avoid F and other constraints on the placement of accent", *Natural Language Semantics*, 7, 141-177.
- Selkirk, E. O. (2007) "Contrastive Focus, Givenness and Phrase Stress," unpublished ms., University of Massachusetts, Amherst.

- Swerts, M., Krahmer E. e Avesani C. (2002). "Prosodic marking of information status in Dutch and Italian: a comparative analysis", *Journal of Phonetics*, 30, 4, 629-654.
- Vallduví, E. (1992). *The informational component*, Garland, New York.
- Zubizarreta, M. L. (1998) *Prosody, Focus, and Word Order*, MIT Press, Cambridge, MA.

LA VARIAZIONE MICRO-PARAMETRICA NEI PROCESSI DI ARMONIZZAZIONE DEL SALENTO MERIDIONALE. STUDIO ARTICOLATORIO TRAMITE ARTICULOGRAFO ELETTROMAGNETICO 3D ED ULTRASUONI

Mirko Grimaldi¹, Andrea Calabrese^{1,2}, Francesco Sigona¹, Bianca Sisinni¹, Luigia Garrapa^{1,3}

¹Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL) – Università del Salento, ²Department of Linguistics – University of Connecticut, ³Fachbereich Sprachwissenschaft – Universität Konstanz
mirko.grimaldi@ateneo.unile.it, francesco.sigona@unisalento.it, bianca.sisinni@unisalento.it,
luigia.garrapa@ateneo.unile.it

Un gruppo di varietà del Salentino meridionale è caratterizzato da processi di assimilazione vocalica che, in linea generale, coinvolgono i fonemi medi /E/ ed /O/ quando sono seguiti dalle vocali alte atone -i ed -u. Il fenomeno di assimilazione comporta l'innalzamento di queste vocali nelle rispettive medio-alte [e] ed [o] (cfr. Grimaldi, 2003).

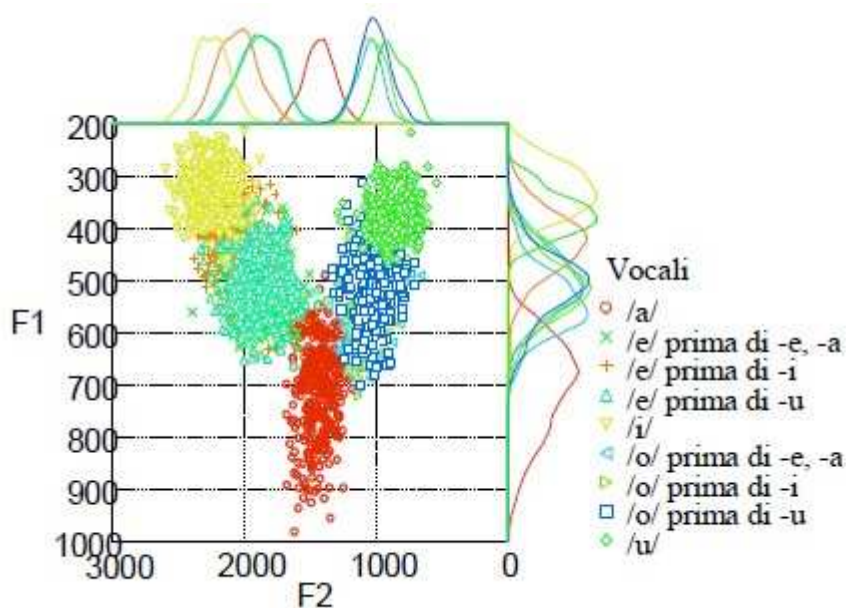


Figura 1: Diagramma F1/F2 su scala lineare in Hz delle aree di esistenza delle 5 vocali del Salentino meridionale prodotte da 36 parlanti di sesso maschile (un parlante per ogni località indagata). I valori in Hz sono associati alla curva di Kernel: la misura più appropriata per campioni di dati che manifestano un fenomeno le cui dinamiche sono poco conosciute. Le vocali medie /E/, /O/ sono differenziate in funzione del contesto metafonetico.

Sulla base, fra gli altri, di Calabrese (1998) e Walker (2005), Grimaldi (2009) propone che l'assimilazione del Salentino meridionale implichi un processo di diffusione del tratto [+ATR] dalla vocale finale alta alla vocale media tonica che la precede. Tuttavia, l'uso del tratto [ATR] per distinguere contrasti di altezza nelle vocali medie romanze (e germaniche)

è stato criticato da Ladefoged & Maddieson (1996), i quali propongono che i contrasti di altezza in questione siano semplicemente dovuti a differenti posizioni del corpo della lingua e non al coinvolgimento della radice della lingua.

In questo studio si cercherà di stabilire se la proposta di Ladefoged & Maddieson (1996) sia valida per le varietà del Salentino meridionale. Per raggiungere lo scopo abbiamo acquisito dati articolatori della lingua tramite l'Articulografo Elettromagnetico 3D (EMA AG500) ed un sistema ad Ultrasuoni (US) sincronizzati. La prima domanda cui intendiamo fornire una risposta è la seguente: il tratto [ATR], che fa riferimento alla adice della lingua, risulta effettivamente coinvolto nel processo di assimilazione vocalica del Salentino meridionale?

È interessante osservare inoltre che – come evidenziato da due ANOVA univariate, un test LSD Posthoc e una serie di t-test a due code per campioni appaiati – le varietà del Salentino in questione possiedono una variazione micro-parametrica del processo di assimilazione vocalica. In effetti, in alcune località /E/ viene innalzata solo prima di -i, mentre /O/ può subire innalzamento a volte prima di -i, altre prima di -u. In un altro gruppo di località, invece, solo /E/ viene innalzata per -i. L'innalzamento di /E/, quindi, avviene prima di -i ed -u in talune varietà, ma solo davanti ad -i in altre varietà, come meglio esemplificato nella Tabella 1, e già messo in evidenza dalle curve di Kernel correlate ai valori di F1 nella Figura 1:

Vocali		Differente azione delle vocali alte atone							
V ₁	V ₂	-i	-u	-i	-u	-i	-u	-i	-u
/E/ →		[e]	[E]	[e]	[E]	[e]	[E]	[e]	[E]
/O/ →		[O]	[O]	[o]	[O]	[o]	[o]	[O]	[o]
Aree		Area A 11 parlanti		Area A1 2 parlanti		Area A2 3 parlanti		Area A3 3 parlanti	

Tabella 1: Differente azione delle vocali alte atone (V2) sulle vocali medie toniche (V1). Si noti che i processi di assimilazione sono stati evidenziati solo in 19 località delle 36 indagate.

Ricorrendo all'utilizzo sincronizzato dell'EMA ed US intendiamo anche esplorare le dinamiche articolatorie di questo processo di assimilazione per determinare il motivo per cui le caratteristiche

articolatorie delle vocali finali influenzino in modo differenziato le vocali toniche precedenti e, quindi, diano luogo alla micro-variazione appena descritta.

Abbiamo registrato i movimenti articolatori della lingua di 3 parlanti di sesso maschile che, dopo una analisi acustica preliminare, hanno dimostrato di possedere i fenomeni di micro-variazione esemplificati. I sensori EMA sono stati posizionati su punta, corpo e dorso della lingua. Come stimoli sono state usate parole inserite nella frase cornice *I dico ____ ora* (10 ripetizioni per ogni tipo vocalico: 5 in sillaba aperta e 5 in sillaba chiusa). Il contesto consonantico che circonda le vocali bersaglio di assimilazione è stato scelto in modo tale da evitare che il processo analizzato sia oscurato da eventuali effetti consonantici. Per quanto possibile sono state quindi preferite consonanti labiali e labiodentali, poiché influenzano in misura minore la posizione del corpo della lingua. Sullo

schermo di un computer abbiamo poi visivamente somministrato in modo randomizzato le frasi da elicitare. Infine come stimoli di controllo abbiamo anche utilizzato delle pseudoparole inserite nello stesso contesto frasale delle parole.

Sulla base dei dati raccolti, nella nostra comunicazione affronteremo anche alcune questioni metodologiche che di recente stanno animando la discussione nel campo degli studi articolatori (cfr. Davidson 2005; Benus, Gafos 2007, Zharkova, Hewlett 2008), come di seguito sinteticamente esemplificato:

1. Come è possibile quantificare i dati degli ultrasuoni al fine di ottenere informazioni affidabili sulle caratteristiche e sulle dinamiche dei processi di armonizzazione vocalica?
2. Come si possono comparare lingue dalla forma diversa utilizzando le immagini degli ultrasuoni?
3. Come si possono comparare i dati EMA con quelli ottenuti tramite US?

Riferimenti bibliografici

- Benus, S. & Gafos A.I., 2007, Articulatory characteristics of Hungarian 'transparent' vowels, *Journal of Phonetics*, 35, 3: 271-300.
- Calabrese, A. 1998, Metaphony revisited, *Rivista di Linguistica*, 10:1.7-68.
- Davidson, L. 2005, Addressing phonological questions with ultrasound, *Clinical Linguistics and Phonetics*, 19, 619-633.
- Grimaldi, M. 2003, *Nuove ricerche sul vocalismo tonico del Salento meridionale. Analisi acustica e trattamento fonologico dei dati*. Alessandria: Edizioni dell'Orso.
- Grimaldi, M. 2009, Acoustic correlates of phonological micro-variations. The case of unsuspected highly diversified metaphonetic processes in a small area of Southern Salento (Apulia), in Danièle Tock and W. Leo Wetzels (eds.): *Romance Languages and Linguistic Theory 2006*. Amsterdam / Philadelphia: John Benjamins. ISBN: 9789027290922.
- Ladefoged, E. I. & Maddieson P. 1996, *The Sounds of the World's Languages*. Oxford, Backwell.
- Walker, Rachel. 2005, Weak triggers in vowel harmony, *Natural Language & Linguistic Theory*, 23.917-989.
- Zharkova N. & Hewlett N., 2009, Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and />/, *Journal of Phonetics*, 37, 2: 248-256.

VOCALISMO FINALE ATONO «TOSCANO» NEI DIALETTI DEL VALLO DI DIANO (SALERNO)

Francesco Cangemi**, Rachele Delucchi*, Michele Loporcaro*, e Stephan Schmid*

*Università di Zurigo, **Laboratoire Parole et Langage & Aix-Marseille I
francesco.cangemi@lpl-aix.fr, delucchi@rom.uzh.ch, loporcar@rom.uzh.ch,
schmidst@pholab.uzh.ch

Mentre il napoletano odierno ha neutralizzato in /ə/ le vocali atone finali proto-romanze (v. Maturi & Schmid 2002), i dialetti dell'estremo meridionale della Campania presentano condizioni più conservative. Rohlfs (1937:86) riporta per il basso Cilento un vocalismo finale atono di tipo siciliano, preservante /i/ (< -E = -I) ≠ /a/ ≠ /o/ (< -O = -U), osservando però che, ad es. a Laurito o Morigerati, «la generazione più giovane pronuncia per lo più già *e* (ə) e *o*». Quest'ultima situazione egli descrive come già stabilizzata nei dialetti dell'alto Cilento (Omignano, Castel S. Lorenzo ecc.). Su questa scia, la bibliografia dialettologica successiva ha riconosciuto come primariamente conservativi in quest'area solo i sistemi finali atoni trivocalici (v. anche Radtke, 1997:63), ascrivendo invece la ricorrenza di -[o] ed -[e] finali a sviluppi secondari più recenti: ad es. De Blasi e Fanciullo (2002:632) spiegano il sistema atono finale quadrivocalico /i e a o/ del dialetto di Cicerale (sempre nell'alto Cilento) come dovuto a «riaggiustamento [...] a partire da una situazione di vocalismo atono più ridotto».

Nel presente saggio consideriamo alcune varietà (San Pietro al Tanagro, Sanza, Polla) parlate immediatamente alle spalle dell'area cilentana, nel vallo di Diano. I dati, raccolti sul campo nel 2006 e già oggetto dello studio sperimentale sul vocalismo tonico di Cangemi (2007), mostrano che in questa zona si ha compattamente un sistema /i e a o/, di cui nessun indizio interno permette di mettere in dubbio l'originarietà (v. Loporcaro, 2009:146s). In altre parole, questi dialetti sono tuttora allo stesso stadio di evoluzione che il toscano, l'unica neutralizzazione ormai compiuta essendo quella di -O e -U in /o/. Per il resto, i movimenti centripeti nello spazio vocalico sono rimasti sincronicamente allofonici, senza incidere sul sistema soggiacente. Il primo fonema a subire centralizzazione è /e/, che pure rimane distinto anche da /i/, come mostra il diagramma in figura 1 (relativo a proparossitoni e parossitoni con /a/ tonica pronunciati isolatamente nel dialetto di San Pietro al Tanagro; l'analisi è stata condotta attraverso l'estrazione automatizzata del Long Term Average Spectrum, utilizzando degli script appositamente elaborati in Praat, Tcl/Tk ed R. Le porzioni di segnale in esame sono state analizzate a 512 punti FFT, con metodo LPC a valori variabili di poles e di preenfasi).

L'asimmetria diacronica (centralizzazione solo di /e/) – che produce una simmetria sincronica – dimostra che siamo di fronte ad una centralizzazione «di sistema» (in modo simile a quanto illustrato per il calabrese centro-settentrionale in Romito *et alii*, 1997), non ascrivibile a ipoarticolazione: è una prima avvisaglia della tendenza alla riduzione del vocalismo atono che in effetti ha mosso altrove il primo passo attraverso la confluenza in /ə/ di /e/ ed /i/ (così nel napoletano tre-quattrocentesco, per il quale si ricostruisce un vocalismo finale trimembre /ə a o/, v. Formentin, 1998:187).

La dimostrazione sperimentale qui condotta per i dialetti del Vallo di Diano impone di riscrivere un capitolo di fonetica storica italiana centro-meridionale.

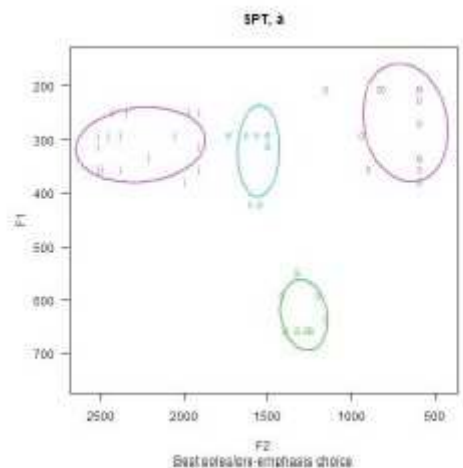


Figura 1. Dialecto di San Pietro al Tanagro: aree di esistenza delle vocali finali in proparossitoni e parossitoni con /a/ tonica (pronunciati in isolamento)

Riferimenti bibliografici

- Cangemi, F. (2007), *Vocalismi tonici nel Vallo di Diano: dinamiche metafonetiche e dittonghi anfitoni*, MThesis, University of Napoli «Federico II», Italy.
- Cortelazzo, M., Marcato, C., De Blasi, N. & Clivio, G., editors (2002), *I dialetti italiani. Storia struttura uso*, Torino: UTET.
- De Blasi, N. & Fanciullo, F. (2002), La Campania, in: Cortelazzo *et al.* (2002), 628-678.
- Formentin, V., editor (1998), *Loise de Rosa. Ricordi*, Roma: Salerno.
- Loporcaro, M. (2009), *Profilo linguistico dei dialetti italiani*, Roma-Bari: Laterza.
- Maturi, P. & Schmid, S. (2002), Dialettologia e fonetica acustica. Una ricerca in Campania, in *La fonetica acustica come strumento di analisi della variazione linguistica in Italia vocali* (Atti delle VII Giornate del Gruppo di Fonetica Sperimentale) (A. Regnicoli, editor), Roma, Tip. Esagrafica, 23-28
- Radtke, E. (1997), *I dialetti della Campania*, Roma: Il Calamo.
- Rohlf, G. (1988), *Studi linguistici sulla Lucania e sul Cilento*, Galatina: Congedo.
- Romito, L., Turano, T., Loporcaro, M. & Mendicino, A. (1997), Micro- e macrofenomeni di centralizzazione vocalica nella variazione diafasica: rilevanza dei dati acustici per il quadro dialettologico del calabrese, in *Fonetica e fonologia degli stili dell'italiano parlato. Atti delle 7e giornate di studio del Gruppo di Fonetica Sperimentale (A.I.A.), Napoli, 14-15 Novembre 1996* (F. Cutugno, editor), Roma: Tip. Esagrafica (Collana degli Atti dell'Associazione Italiana di Acustica), 157-175.

VARIAZIONE, UNIVERSALI LINGUISTICI E CONTRIBUTO DELL'INTONAZIONE

Lunella Mereu
Università Roma Tre
mereu@uniroma3.it

In questa sede, il cui scopo primario è discutere metodologie di analisi e questioni di confine tra fonologia e altri aspetti della linguistica, vorrei affrontare un problema linguistico non fonologico per mostrare come l'interfaccia tra la fonologia (dell'intonazione) e gli altri livelli di analisi sia rilevante per stabilire principi universali nella costruzione degli enunciati linguistici. In altre parole, scopo della mia presentazione sarà quello di mostrare come ciascun livello di analisi, non escluso quello fonologico, possa arrivare a delle giuste generalizzazioni in una prospettiva che tenga conto delle interfacce o della 'multidimensionalità' dell'analisi linguistica, piuttosto che mantenendosi ancorato a principi di autonomia e a una discussione interna teorica intra-livello tipica della linguistica del XX secolo.

Partirò quindi dal fenomeno della variazione sintattica, cioè dall'ordine delle parole nelle lingue del mondo, ed arriverò a formulare una proposta di principi universali che regolano tale variazione nei termini delle interfacce fonologia – morfo-sintassi – pragmatica. Rilevante per la proposta sugli universali sarà il confronto tra la marcatura morfo-sintattica delle nozioni pragmatiche (di topic e focus) in alcune lingue tipologicamente diverse e lontane dal tipo occidentale, e la prosodia dei fenomeni sintatticamente marcati in lingue come l'italiano. Prenderò in considerazione i contorni intonativi dei casi di emarginazione dell'italiano analizzando in particolare le differenze tra topic a destra e topic a sinistra, ed estendendo alcune delle generalizzazioni sulle posizioni del topic anche alla distinzione tra focus iniziale e focus postverbale. La letteratura non fonologica ha già evidenziato similarità tra topic e focus in alcune posizioni e/o situazioni pragmatiche (Büring 1997; Choi 1997, 1999; Erteschik-Shir 2007; Mereu 2004, 2009; Molnár e Winkler 2006), anche la letteratura fonologica (Crocco 2009, Crocco e Savy 2007; Mereu e Trecci 2004) ha iniziato ad occuparsi di prosodia e prominenza del topic, ma ci sono già diversi modelli teorici che fanno previsioni diverse al riguardo. Su questa variabilità teorica sarà opportuno riflettere.

Dal punto di vista metodologico affermerò l'importanza degli studi che privilegiano gli enunciati in contesto piuttosto che singole frasi 'di laboratorio', e il canale parlato (spontaneo o semispontaneo) piuttosto che il letto (da fonti scritte o anche orali).

Riferimenti bibliografici

- Büring, Daniel. 1997. *The Meaning of Topic and Focus. The 59th Street Bridge Accent*. London: Routledge.
- Choi, Hye-Won. 1997. Information structure, phrase structure, and their interface. *Proceedings of the LFG97 Conference (on line)*, Miriam Butt & Tracy Halloway King (eds.). Stanford: CSLI Publications.
- Choi, Hye-Won. 1999. *Optimizing Structure in Context: Scrambling and Information Structure*. Stanford: Center for the Study of Language and Information Publications.

- Crocco, Claudia. 2009. Topic accent and prosodic structure. In Mereu (ed.), *Information Structure and its Interfaces*. Berlino/New York: Mouton de Gruyter, 15-49.
- Crocco, Claudia e Savy, Renata, 2007. Topic in dialogue: prosodic and syntactic features. In *Proceedings of Interspeech 2007*, Antwerp: Belgium, CD-ROM.
- Erteschik-Shir, Nomi. 2007. *Information Structure. The Syntax-Discourse Interface*. Oxford: Oxford University Press.
- Mereu, Lunella. 2004. *La Sintassi delle Lingue del Mondo*. Roma-Bari: Laterza.
- Mereu, Lunella. 2009. Universals of information structure. In Mereu (ed.), *Information Structure and its Interfaces*. Berlino/New York: Mouton de Gruyter, 75-101.
- Mereu, Lunella e Trecci, Alessandra. 2004. Focus sul topic. In Albano Leoni, Federico et al. (eds.), *Atti del Convegno Il Parlato italiano, Napoli 13-15 febbraio 2003*. Napoli: M. D'Auria Editore, CD-ROM.
- Molnár, Valéria e Susanna Winkler (eds.), 2006. *The Architecture of Focus*. Berlin: Walter de Gruyter.

CACUMINALI CALABRESI: MODI E LUOGHI D'ARTICOLAZIONE ALLA LUCE DI MISURAZIONI ACUSTICHE E INDAGINI PER RISONANZA MAGNETICA (IRM)

Antonio Romano^{1,2,3}, Francesco Gambino^{2,3}

¹LFSAG – Laboratorio di Fonetica Sperimentale “Arturo Genre”

²Facoltà di Lingue e Letterature Straniere - ³Università degli Studi di Torino

antonio.romano@unito.it

L'articolazione dei cosiddetti suoni retroflessi/cacuminali calabresi ha suscitato molto interesse in ambito dialettologico e fonetico (v. Soriano & Mancuso, 1998, Romito & Soriano, 1998, Loporcario, 2001). In Calabria questi suoni, simili a quelli attestati in altri dialetti d'Italia (siciliani o salentini), manifestano un'eterogeneità di esiti fonetici più evidente che nelle altre aree (cfr. Romito & Belluscio 1996). I suoni che subiscono processi di retroflessione sono essenzialmente la laterale di origine latina -LL- (ma anche alcune -ll- secondarie romanze) e i nessi consonantici -tr-, -dr-, -ntr-, -ndr-, -ttr-, -str- e -rr- (per le varietà salentine, cfr. Romano, 1999, e Costagliola & Katiwada, 2008; più in generale si veda Celata, 2006).

Nell'ambito del lavoro di ricerca svolto dall'autore F. Gambino per la sua tesi di laurea triennale, abbiamo condotto alcune verifiche acustiche e articolatorie su tre suoni caratteristici della varietà linguistica di Vazzano, un piccolo paese dell'entroterra calabrese in provincia di Vibo Valentia.

In particolare ci siamo concentrati sull'analisi delle tre affricate [tʃ], [tʃ̠] e [tʃ̠̠] corrispondenti a quelle studiate, con una metodologia simile, per la varietà salentina di Parabita (cfr. Romano, 1999, 2002). Di questi tre suoni abbiamo analizzato le caratteristiche acustiche (in particolare l'andamento della seconda e della terza formante e i tempi di tenuta e di rilascio), nelle transizioni con suoni vocalici contigui presenti nelle sequenze /'aC:#'Ca/, /'iC:#'Ci/ e /'uC:#'Cu/ (con C = [tʃ], [tʃ̠] e [tʃ̠̠]). Le caratteristiche articolatorie di questi suoni, inseriti nelle stesse sequenze, sono state osservate facendo ricorso a tecniche di risonanza magnetica (sulla base di 27 IRM eseguite presso il Servizio Radiodiagnostica dell'ospedale San Giovanni Battista - Molinette di Torino) che hanno permesso di ottenere una visione sagittale e coronale degli organi coinvolti nella produzione dei suoni nei diversi contesti, confermando i risultati di studi precedenti su altre varietà. In particolare si ribadiscono con chiarezza la postalveolarità di [tʃ̠] (non pre-palatale, né palatale, né alveolo-palatale, cfr. Romano, 2007, Romano & Badin, 2009) e la mancanza d'indici di retroflessione per [tʃ̠̠] (cfr. Romano, 2002).

Celata C. (2006). “Analisi dei processi di retroflessione delle liquide in area romanza: con dati sperimentali dal corso e dal siciliano”. *Tesi di Dottorato*, Scuola Normale Superiore di Pisa.

Costagliola A. & Katiwada R. (2008). “Salentinian cacuminals/retroflexes (Apulia, southern Italy): a preliminary articulatory study”. *New Trends in Experimental Phonetics* (IV Congr. Int. de Fonética Experimental, Granada, Spagna, 2008) (A. Pamies & E. Melguizo, eds.), *Language Design*, no. sp. 1, 39-46.

Gambino F. (2008). "Studio acustico e articolatorio di consonanti affricate del dialetto di Vazzano (VV)". *Tesi di Laurea Triennale*, Fac. di Lingue e Lett. Str. - Università di Torino (Rel. Antonio Romano), a.a. 2006-2007.

Loporcaro M. (2001). "Le consonanti retroflesse nei dialetti italiani meridionali: articolazione e trascrizione". *Bollettino del CSLFS*, 19, 207-233.

Romano A. (1999). A phonetic study of a Sallentinian variety (southern Italy). *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS99)*, San Francisco, USA, 1051-1054.

Romano A. (2002). "La fonetica strumentale applicata ai dialetti d'Italia a un secolo dall'"Etude sur la phonétique italienne" di F.M. Josselyn". In A. Regnicoli (a cura di), *La fonetica acustica come strumento di analisi della variazione linguistica in Italia*, Roma, Il Calamo, 7-14.

Romano A. (2007). "La fonetica sperimentale e gli atlanti linguistici: la sintesi romanza di 'pidocchio' e lo studio degli esiti palatali". In J. Dorta (a cura di), *Temas de Dialectología*, Tenerife, Instituto de Estudios Canarios, 178-204.

Romano A. & Badin P. (2009). "An MRI Study on the Articulatory Properties of Italian Consonants". *Estudios de Fonética Experimental*, XVIII (ed. speciale per il Simposio Internacional - 30è aniversari del laboratori de fonètica de la UB, Barcelona, 2008), 327-344.

Romito L. & Belluscio G.M.G. (1996). "Studio elettropalatografico dell'opposizione fonematica [ll], [dd], [ɖɖ] nel dialetto di Catanzaro e [t̪], [ʎ], [d̪], [ð̪] nella parlata albanese di San Basile". *Atti del XXIV Conv. naz. dell'A.I.A.* (Trento, 1996), 141-144.

Romito L. & Sorianello P. (1998). "Ridefinizione delle consonanti retroflesse nei dialetti calabresi". *Com. pres. al conv. int. SILFI* (Catania, 1998), preprint 12 pp.

Sorianello P. & Mancuso A. (1998). "Le consonanti retroflesse nel cosentino: analisi preliminare". In P.M. Bertinetto & L. Cioni (a cura di), *Unità fonetiche e fonologiche: produzione e percezione (Atti delle VIII Giornate di Studio del G.F.S. dell'Ass. Italiana di Acustica, Pisa, 1997)*, Roma, Esagrafica, 142-154.

METODOLOGIE NEGLI STUDI CROSS-LINGUISTICI. RIFLESSIONI SULLA RELAZIONE FRA DATI ACUSTICI E PERCETTIVI.

Bianca Sisinni

Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL) - Università del Salento

bianca.sisinni@unisalento.it

In questo lavoro viene affrontato un argomento già ampiamente discusso in letteratura presenta questioni ancora irrisolte (si vedano fra gli altri, Stevens et alii, 1996; Bohn et alii, 1999; Strange & Shafer 2008). Sulla base di dati empirici diversificati, si cercherà di riflettere sulle scelte metodologiche più idonee per la conduzione degli studi sull'acquisizione delle categorie fonetico-fonologiche della seconda lingua (L2) da parte di parlanti nativi di L1 (lingua materna). Risulta controverso, infatti, il rapporto che intercorre fra dati acustici (di produzione) e dati percettivi le cui analisi conducono spesso a risultati fra loro divergenti.

Le principali teorie sull'acquisizione dei fonemi non nativi (Speech Learning Model - SLM, Flege 1995; Perceptual Assimilation Model - PAM, Best 1995; Native Language Magnet Model - NLM, Khul 1992) affermano che alla base di tale processo cognitivo vi siano vari fattori fra i quali le similarità/differenze fra i fonemi dei due sistemi (nativo e non nativo). Al contrario di quanto affermato in precedenza dalla Contrastive Analysis Hypothesis (Lado 1957), questi modelli teorici sostengono che il processo di acquisizione sarebbe facilitato per quei fonemi non nativi nuovi, non categorizzabili (totalmente o parzialmente), o non prototipici, rispetto ai fonemi della L1. Viceversa, i fonemi non nativi simili, categorizzabili e prototipici rispetto ai fonemi della L1, sarebbero più difficili da acquisire.

In base a ciò, nell'ambito degli studi cross-linguistici, risulta necessario individuare quei fonemi di L2 che possono essere considerati simili o differenti rispetto ai fonemi di L1, per poi poter formulare inferenze circa i relativi processi di acquisizione. Una volta stabilito quali siano i fonemi non nativi più difficili/facili da discriminare percettivamente, si può verificare come questi vengano articolati dai parlanti nativi di L1.

Un primo obiettivo di questo studio è quello di confrontare i sistemi fonologici L1 e L2, per rilevarne differenze e similarità, ricorrendo a due diverse metodologie. In primo luogo è stato effettuato un confronto dei due sistemi all'interno dello spazio acustico delineato dagli assi cartesiani corrispondenti alle prime due frequenze formantiche F1 ed F2. Successivamente, lo stesso confronto è stato condotto attraverso misurazioni comportamentali, facendo eseguire test percettivi il cui scopo è individuare le similarità/dissimilarità che i parlanti nativi di L1 percepiscono fra i fonemi nativi e non nativi. Sebbene l'oggetto dei confronti fosse il medesimo, i risultati ottenuti non sono concordanti.

Un secondo obiettivo prevede la valutazione delle produzioni di L2 da parte di parlanti nativi di L1 attraverso l'analisi acustica delle stesse e i rate (giudizi) espressi da parlanti nativi di L2, in modo tale da evidenziare eventuali discrepanze fra i risultati ricavati ottenuti dai due livelli di analisi.

Hanno preso parte a questo lavoro due gruppi di soggetti, un gruppo sperimentale (GS) ed un gruppo di controllo (GC). Entrambi i gruppi sono costituiti da studentesse universitarie, parlanti native dell'Italiano Salentino (IS), basate sui cinque fonemi /i/, /e/, /a/, /o/, /u/ (Grimaldi 2003, 2009), frequentanti il primo anno della Facoltà di Lingue

dell'Università del Salento. Sia il GS che il GC hanno seguito lezioni di Lingua Inglese (L2) tenute da una parlante madrelingua dell'Inglese Americano (AE), ma solo il GS ha seguito un training fonetico - per 7 settimane, 3 ore/settimana - mirato all'apprendimento dei fonemi vocalici non dittongati dell'AE: /i:/, /ɪ/, /e/, /æ/, /ʌ/, /ɑ:/, /ɔ:/, /ɜ:/, /ʊ/ e /u:/ (Ladefoged 2001). Entrambi i gruppi sono stati sottoposti a dei test di produzione e percezione di L2 prima del training (pretest), dopo il training (posttest) e a distanza di due mesi dalla fine del training (2°posttest), per verificare se le capacità di percepire ed articolare i fonemi non nativi fossero maggiori per il GS rispetto al GC.

Per stabilire quali fonemi di L2 fossero da considerare come simili/differenti rispetto ai fonemi di L1, il GS e il GC hanno prodotto i fonemi dell'IS, mentre i fonemi dell'AE sono stati prodotti da 3 parlanti native dell'AE. Tutte le produzioni sono state analizzate acusticamente (F1 e F2 nel tratto stabile di 0,02 sec) e normalizzate con la formula di Lobanov (1971) per renderle comparabili. I valori Hz sono stati in seguito convertiti in Hz-like attraverso NORM.

La comparazione nello spazio acustico è avvenuta calcolando la distanza euclidea fra fonemi di L1 e fonemi di L2: essa è stata utilizzata come "indice di similarità", in quanto ad una distanza minore corrisponderebbe una maggiore vicinanza acustica ed un maggiore grado di similarità fra fonemi appartenenti a sistemi linguistici differenti. In particolare, le distanze euclidee sono state calcolate fra i fonemi dell'IS e determinati fonemi dell'AE (/ɪ/, /æ/, /ʌ/, /ɜ:/ ed /ʊ/) tralasciando quei fonemi della L2 (/i:/, /u:/, /e/, /ɔ:/ e /ɑ:/) acusticamente affini ai fonemi dell'IS, e verosimilmente meno ambigui da un punto di vista percettivo (cfr. Figura 1).

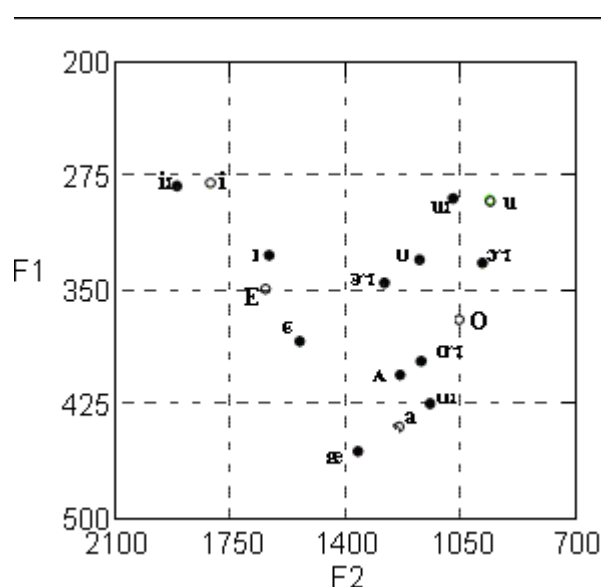


Figura 1: i valori formantici medi (Hz-like) dei fonemi dell'AE (cerchio pieno) e dell'IS (cerchio vuoto)

Le distanze euclidee, quindi, sono state calcolate per le coppie in (1):

/i:/, /ɪ/, /ε/, /æ/, /ʌ/, /ɑ:/, /ɑ:/, /ɜ:/, /ɔ:/, /ʊ/ e /u:/

(1) /ɪ/ - /i/ - /ɪ/ - /E/

/æ/ - /a/ - /æ/ - /E/

/ʌ/ - /a/ - /ʌ/ - /O/

/ɜ:/ - /E/ - /ɜ:/ - /O/

/ʊ/ - /u/ - /ʊ/ - /O/

Inoltre, GS e GC hanno eseguito un test percettivo, l'identification test, il cui scopo è quello di misurare la perceived phonetic distance (Flege et alii, 2004) fra L1 ed L2, attraverso giudizi qualitativi (fit index) espressi dai parlanti nativi di L1 sul grado di similarità (fit index elevato)/dissimilarità (fit index basso) fra i fonemi nativi e i fonemi di L2.

Dalla comparazione delle distanze euclidee delle coppie in (1) e dei fit index degli stessi fonemi di L2 rispetto ai fonemi di L1, è stato evidenziato come il confronto fra due sistemi linguistici non possa limitarsi esclusivamente ad una delle due metodologie, in quanto fra i risultati è emersa in maniera evidente una discrepanza: al dato acustico della distanza euclidea non corrisponde, nella quasi totalità dei casi, il dato ricavato dal giudizio percettivo delle parlanti native di L1 espresso sui fonemi di L2, confermando ancora una volta che la similarità/dissimilarità nello spazio acustico non è necessariamente indice di similarità/dissimilarità.

In secondo luogo, per verificare se la capacità di articolazione di tre fonemi non nativi (/ɪ/, /æ/ e /ɜ:/) percepiti come differenti rispetto al sistema L1, si fosse modificata per il GS e/o il GC nel corso del pretest, posttest, 2°posttest, sono state prese in esame le aree di esistenza dei tre fonemi - e l'analisi statistica dei valori formantici medi F1 ed F2 - e i rate espressi da 6 parlanti nativi dell'AE sugli stessi fonemi. Anche in questo caso sembrano sussistere delle differenze fra i dati acustici ed i dati percettivi: in particolare, nel caso del fonema /æ/, ad un'iperarticolazione nello spazio acustico rispetto allo stesso fonema prodotto da parlanti nativi dell'AE, è corrisposto un rate elevato, indice della "bontà" della produzione. Ciò dimostra che la valutazione delle produzioni non native non può fermarsi alla loro comparazione nello spazio acustico o al solo giudizio espresso da parlanti nativi di L2, in quanto si rischierebbe di perdere dati importanti ai fini di una valutazione completa ed esaustiva delle stesse. Tuttavia, da questo studio emerge che i dati percettivi, in entrambi i casi, sembrano essere fortemente indicativi di quanto avviene in un'interazione naturale, poiché essi riflettono i processi di categorizzazione dei fonemi non nativi rispetto ai fonemi nativi (identification test) e forniscono un feedback veritiero circa il grado di intelligibilità - ovvero la capacità di articolazione - dei parlanti di L2 (rate).

Riferimenti bibliografici

Best, C. T. , (1995), A direct realist view of cross-language speech perception, in Strange W. (Ed.), Speech perception and linguistic experience: Issues in cross-language research , 171-204, Timonium, MD: York Press.

Best, C. T., Tyler, M. D., (2008) Nonnative and second language speech perception: Commonalities and complementarities. In M.J. Munro & O.-S.

Bohn (eds) *Second Language speech learning: the role of language experience in speech perception and production*, 13-34, Amsterdam: John Benjamins.

Bohn, O.-S., & Flege, J. E., (1992), The production of new and similar vowels by adult German learners of English, *Studies in Second Language Acquisition*, 14: 131-158.

Bohn, O.-S., Strange, W. & Trent, S. A. , (1999), On what it takes to predict perceptual difficulty in cross-language vowel perception, *Journal of the Acoustical Society of America*, 105 (2): 2.

Flege, J. E., (1991), Perception and production: the relevance of phonetic input to L2 phonological learning, in Heubner T., Ferguson C. (eds), *Crosscurrents in Second Language acquisition and Linguistic Theory*, 249-290, Philadelphia, John Benjamins Publishing Company.

Flege, J. E., (1995), *Second-language Speech Learning: Theory, Findings and Problems*, in Strange W.(ed), *Speech Perception and Linguistic experience: Issues in Cross-language research*, 229-273, Timonium MD, York Press.

Flege, J. E., MacKay, I. R. A., (2004), Perceiving vowel in a second language, *Studies in Second Language Acquisition*, 26: 1-34.

Grimaldi, M., (2003), *Nuove ricerche sul vocalismo tonico del Salento Meridionale. Analisi acustica a trattamento fonologico dei dati*, Alessandria, editore dell'Orso.

Grimaldi, M., (2009), "Acoustic Correlates of Phonological Microvariation", in D. Tock and W. L. Wetzels (eds.), *Romance Languages and Linguistic Theory 2006*. Amsterdam, John Benjamins, ISBN: 9789027248190.

Kuhl, P. K., (1993), Innate predispositions and the effects of experience in speech perception: The native language magnet theory, in B. de Boysson-Bardies, S. da Chonen, P. Juszcyk, P. McNeilage & J. Mortan (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life*, 259-274, Dordrecht, Holland, Kluwer Academic.

Ladefoged, P., (2001), *Vowels and consonants: An introduction to the sounds of language*, Malden, MA,

Blackwell. Lado, R., (1957), *Linguistics across cultures*, Ann Arbor, MI, University of Michigan Press.

Lobanov, B. M., (1971), Classification of Russian Vowels Spoken by Different Speakers, *The Journal of the Acoustical Society of America*, 49: 606-608, Issue 2B.

Munro, M. J., (2008), Foreign accent and speech intelligibility, in J. G. Hansen Edwards, M. Zampini (Eds.), *Phonology and Second Language Acquisition*, 193-218, Amsterdam/Philadelphia, John Benjamins.

Stevens, K., Liberman, A., Studdert-Kennedy M., & Öhman, S., (1996), Cross language study of vowel perception, *Language Speech*, 12: 1-23.

Strange, W., & Shafer, V. L., (2008), Speech perception in second Language learners: The re-education of selective perception, , in J. G. Hansen Edwards, M. Zampini (Eds.), *Phonology and Second Language Acquisition*, 153-192, Amsterdam/Philadelphia, John Benjamins.

INVARIANZA E VARIABILITÀ, IERI E OGGI. SPUNTI PER UNA DISCUSSIONE A PIÙ VOCI

Arianna Uguzzoni
Università degli Studi di Bologna
arianna.uguzzoni@unibo.it

Scopo di questo intervento è stimolare ad una riflessione su qualche aspetto dell'ampia problematica della variabilità in campo fonico, considerata in relazione al polo opposto della invarianza.

Elenco qui in forma schematica alcuni spunti per una discussione a più voci.

1. Come si configura oggi dal punto di vista teorico il macrotema della invarianza e della variabilità nello speech, oggetto dell'importante simposio del 1983? (Perkell & Klatt (eds.), 1986).

2. E' accettabile l'ipotesi per cui l'invarianza si colloca a livello cognitivo, mentre la variabilità si colloca al livello della implementazione? (Docherty & Foulkes, 2000).

3. Invarianza articolatoria, acustica, percettiva o invece invarianza "negoziata"? (Lindblom, 1989; Schwartz, 1990).

4. La variabilità è un fenomeno casuale e senza limiti o è un fenomeno sistematico governato da regole? (Johnson, Ladefoged, Lindau, 1993).

5. Quali sono le fonti della variabilità in campo fonico?

6. I progressi nello studio della variabilità hanno implicazioni per alcuni problemi centrali della fonologia e della fonetica? (Foulkes, 2006).

Riferimenti bibliografici

Perkell, J.S., Klatt, D.H., 1986 (eds.), *Invariance and variability in speech processes*, Hillsdale, N.J.: Lawrence Erlbaum Associates.

Docherty, G.J., Foulkes, P., 2000, Speaker, speech, and knowledge of sounds. In N. Burton-Roberts, P. Carr, G.J. Docherty (eds.), *Phonological knowledge. Conceptual and empirical issues*, Oxford: Oxford University Press, 105-129.

Lindblom, B., 1989, Phonetic invariance and the adaptive nature of speech. In B.A.G. Elsendoorn, H. Bouma (eds.), *Working models of human perception*, London: Academic Press, 139-173.

Schwartz, J.-L., 1990, Perception de la parole: invariance et variabilité, *Colloque de physique*, 51, 461-470.

Johnson, K., Ladefoged, P., Lindau, M., 1993, Individual differences in vowel production, *Journal of the Acoustical Society of America*, 94, 701-714.

Foulkes, P., 2006, Phonological variation: a global perspective, In B. Aaarts, A. McMahon (eds.), *Handbook of English Linguistics*, Oxford: Blackwell, 625-669.

VARIAZIONE DEI PARAMETRI NEUROLINGUISTICI NELL'UNSPOKEN SPEECH

Pamela Mattana¹, Andrea Paoloni¹, Mirko Grimaldi²

Università La Sapienza, Roma, Fondazione Ugo Bordoni, Roma¹

Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL), Università del Salento²

pamela.mattana@gmail.com, pao@fub.it, mirko.grimaldi@ateneo.unile.it

Il campo delle neuroscienze cognitive del linguaggio, un settore di ricerca che attualmente ha una forte vocazione interdisciplinare, può vantare origini molto antiche. Tuttavia fino all'avvento delle tecniche neurofisiologiche e di neuroimaging l'esplorazione dell'attività corticale e sottocorticale e la sua connessione con la facoltà del linguaggio era basata sull'osservazione a posteriori, spesso post mortem, dei danni che una lesione al cervello poteva aver causato alle abilità linguistiche di un soggetto. Era sulla base di osservazioni delle conseguenze di danni cerebrali che si avanzavano ipotesi su quale fosse il funzionamento e la struttura del nostro encefalo. Oggi, grazie alle tecnologie sopra menzionate, è possibile vedere il cervello "al lavoro" in maniera poco o per nulla invasiva, ed è stato perciò possibile quello sviluppo in termini di conoscenze ed investimenti che hanno visto le neuroscienze cognitive protagoniste negli ultimi 20 anni. Le neuroscienze cognitive, e in particolare quel settore dedicato all'esplorazione delle connessioni neurali che stanno alla base della conoscenza del linguaggio (per usare una nozione chomskiana), richiedono un dialogo costante tra discipline diverse che spesso esulano di molto dalla tradizionale preparazione del linguista. Vi è la necessità infatti di acquisire e coniugare competenze di base in neurofisiologia, in psicologia, e in ingegneria sia per la creazione e la somministrazione degli stimoli utilizzati nei protocolli sperimentali, sia per il successivo trattamento dei dati ottenuti.

In questo campo di studi un settore di ricerca attualmente di frontiera è rappresentato dalla possibilità di catturare i 'segnali' connessi con l'attività vocale senza che venga emesso nessun suono dall'apparato fonatorio. Per esempio, è stato di recente dimostrato che il segnale elettromiografico prodotto dai muscoli orofacciali o da quelli della laringe contiene informazioni sufficienti per discriminare e sintetizzare in modo accurato un set di parole che viene articolato ma non acusticamente realizzato (Chan et al. 2002, Jorgensen et al. 2003, Jou et al. 2006, Walliczek et al. 2006). Ancora più di recente Porbadnigk et al. (2008) suggeriscono che il segnale elettroencefalografico può essere sfruttato per catturare lo speech silente (ovvero solo pensato). Frank Guenther e la sua équipe dell'Università di Boston presso il Department of Cognitive & Neural System, hanno invece sfruttato un modello a reti neurali, il DIVA (Directions Into Velocities of Articulators), per un tentativo ancora più pionieristico. Il DIVA è in grado di apprendere proprio come farebbe un bambino nella fase del babbling, per cui la generazione in output viene sempre confrontata con il pattern standard nell'input: via via che si susseguono le produzioni, l'output rassomiglia sempre più all'input (si veda Guenther & Vladusich, in press). Questo modello ha compiuto un balzo in avanti grazie all'integrazione della mappa dei correlati neurali connessi con lo speech che ne ha permesso l'applicazione in un esperimento davvero avveniristico. Infatti nell'aprile del 2008 ad un ragazzo con lesioni spinali gravi, i cui movimenti si limitano al lento sbattere delle ciglia, finora unico mezzo per comunicare con il mondo esterno, sono stati impiantati nella corteccia cerebrale degli elettrodi basati su un sistema wireless. Questi elettrodi hanno catturato il segnale elettrico proveniente dall'area della corteccia motoria, che veniva generato dal soggetto a cui è stato richiesto di "pensare" di produrre una vocale,

in questo caso la /u/. Il segnale così catturato è stato trasferito al sintetizzatore DIVA che ha riprodotto la vocale.

Con il nostro intervento intendiamo presentare un progetto di ricerca che si colloca in questo settore pionieristico, ed eventualmente discutere i primi dati grezzi ottenuti dagli esperimenti pilota. Nell'ottica dell'interdisciplinarietà prima descritta ci si avvarrà del supporto della Fondazione "Ugo Bordonì" di Roma per quanto concerne l'elaborazione del segnale e del supporto del CRIL di Lecce per l'organizzazione del protocollo, la strumentazione e la rilevazione dei segnali EEG.

Il nostro obiettivo è quello di poter replicare l'esperimento di Guenther e colleghi senza impiantare gli elettrodi direttamente nella corteccia motoria, ma utilizzando esclusivamente una cuffia EEG per Potenziali Evento Correlati (ERPs a 64 canali), sfruttando quindi un segnale più debole e con più rumore (sulla scia del primo esperimento tentato da Porbadnigk et al. 2008), e ricorrendo, per la sintesi, al modello DIVA. Dopo aver individuato le aree corticali più efficaci per la rilevazione del segnale dallo scalpo, procederemo alla discriminazione dei pattern di produzione che saranno infine sintetizzati in segnale vocale.

Un primo esperimento consisterà nel far leggere a 3 soggetti un numero limitato di vocali (/i/, /u/, /a/) e di bisillabi (per es.: casa, porta, rosso, etc.) che verranno via via visualizzati su uno schermo mentre un attivatore farà partire l'acquisizione del segnale EEG al momento più opportuno. La "lettura" avverrà una prima volta sulla base del solo pensiero (il soggetto penserà intensamente e dettagliatamente una sola volta alla parola che compare sullo schermo), mentre il soggetto dovrà muovere la mandibola senza emettere tuttavia alcun suono. Si effettueranno 10 ripetizioni per ogni tipo di stimolo, nel tentativo di ottenere una prima base di dati sperimentale su cui tentare l'identificazione e la sintesi degli stimoli "pensati".

Il corpus così acquisito conterrà un primo file che corrisponderà all'attività del cervello in quiete (prima che le vocali e le parole vengano pensate), un secondo file contenente il tratto di segnale corrispondente alla presentazione degli stimoli, un terzo file contenente il tratto di segnale corrispondente al movimento mandibolare senza che venga emesso nessuna vocalizzazione.

Nella comunicazione verranno criticamente illustrati i presupposti teorici di partenza, verranno quindi descritte e discusse in dettaglio tutte le metodiche utilizzate per l'identificazione, e verranno presentati i primi dati grezzi.

Riferimenti bibliografici

Chan, A., K. Englehart, Hudgins, B., and Lovely, D. (2002), Hidden markov model classification of myoelectric signals in speech, *Engineering in Medicine and Biology Magazine*, IEEE, 21:143–146.

Jou, Szu-Chen, Schultz, Tanja, Walliczek, Matthias, Kraft, Florian, Waibel, Alex (2006), Towards continuous speech recognition using surface electromyography, In *INTERSPEECH-2006*, 17-21 September, Pittsburgh.

Kröger B.J., Kannampuzha J., Neuschaefer-Rube C. (2009), Towards a neurocomputational model of speech production and perception, *Speech Communication*, 51, 9: 793-809.

Guenther, F. (in press), A Neural Theory of Speech Acquisition and Production, in M. Grimaldi,

L. Craighero, "Is a Neural Theory of Language Possible?", Special Issue of *Journal of Neurolinguistics*.

Jorgensen, C., Lee, D., and Agabon, S. (2003), Sub auditory speech recognition based on emg/epg signals. In Proceedings of the International Joint Conference on Neural Networks

Porbadnigk, A. (2008), Eeg-based speech recognition: Impact of experimental design on performance. Institut für Algorithmen und Kognitive Systeme, Universität Karlsruhe (TH), Karlsruhe, Germany.

Walliczek, Matthias, Kraft, Florian, Jou, Szu-Chen, Schultz, Tanja, Waibel, Alex (2006), Subword unit based non-audible speech recognition using surface electromyography, In *INTERSPEECH-2006*, 17-21 September, Pittsburgh.

VARIAZIONE NELLE CATEGORIE INTONATIVE

Amedeo De Dominicis
Università della Tuscia, Viterbo
dedomini@unitus.it

Alla base di molti problemi tipici delle indagini intonative sembra di poter individuare una causa comune. Si tratta di quello che è stato definito il basso grado di categorialità dei diversi tipi *accent* o prominenza intonativa. Qui il termine categorialità è inteso nell'accezione della teoria della cosiddetta *percezione categoriale* (Massaro & Cohen 1983; Kohler 1987, 1990, 1991, 2006; Gartenberg & Panzlaff-Reuter 1991; D'Imperio & House 1997; Ladd & Morton 1997; Massaro 1998; Remijsen & van Heuven 1999; Gussenhoven 1999, 2004, 2006; Niebuhr & Kohler 2004; Falé & Faria 2005; Schneider et al. 2006; Vanrell Bosch 2006; Gili Fivela 2005, 2006a, 2006b, 2008).

Negli esperimenti condotti, la natura categoriale dell'intonazione appare controversa. Nel *paper* che intenderei presentare, riporto i risultati di una serie di esperimenti percettivi, condotti con soggetti di area romana. Ho manipolato il contorno intonativo della frase *mangia la pasta*, pronunciata come semplice informazione (De Dominicis 1992). Le manipolazioni riguardano la forma del contorno (accentazione) relativo al focus *in situ* (collocato a destra della frase, sul costituente *pasta*) e riproducono le categorie intonative che sono ritenute rilevanti o per il tipo di *focus* (contrastivo, informativo) o per la diatopia (PI, FI, RM): L*+H, H*+L, H*, L+H*, H+L*. Le figure 1-5 illustrano i cinque tipi di manipolazioni.

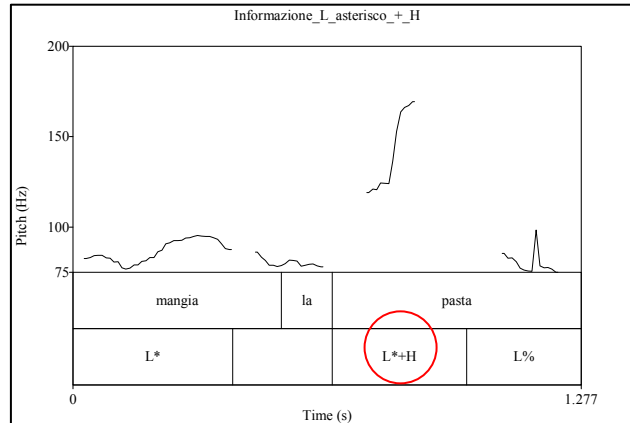


Figura 1. L*+H: accento di *topic* (Frascarelli & Hinterhölzl 2007).

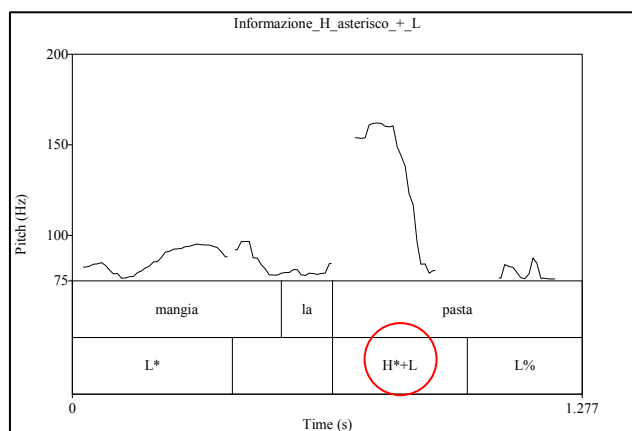


Figura 2. H*+L: focus ristretto nuovo e contrastivo RM (Frascarelli 2004), contrastivo PI (Gili Fivela 2008).

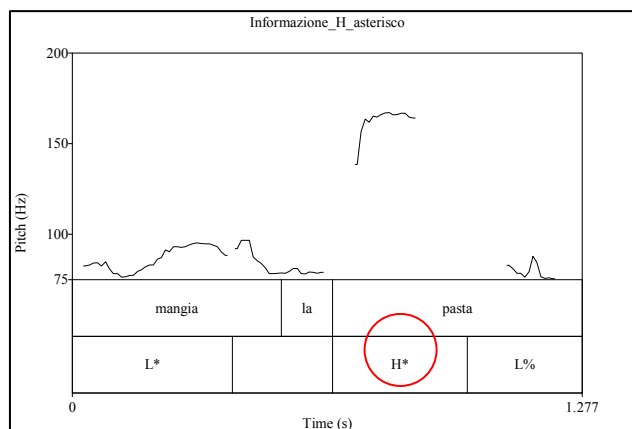


Figura 3. H*: focus contrastivo e informativo FI (Avesani 2003; Avesani & Vayra 2003, 2004), ristretto nuovo e contrastivo RM (Frascarelli 2004).

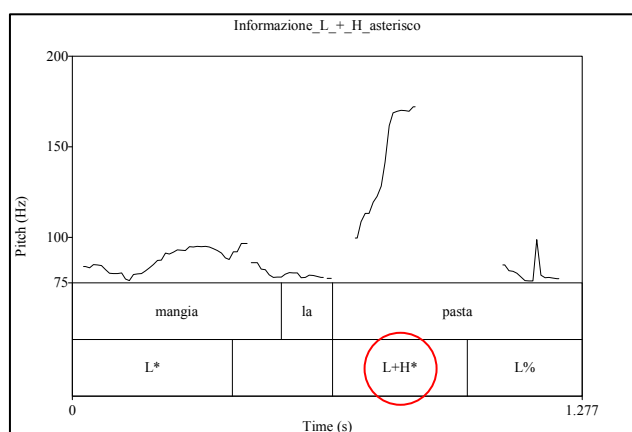


Figura 4. L+H*: focus ristretto contrastivo FI (Avesani 2003; Avesani & Vayra 2003, 2004).

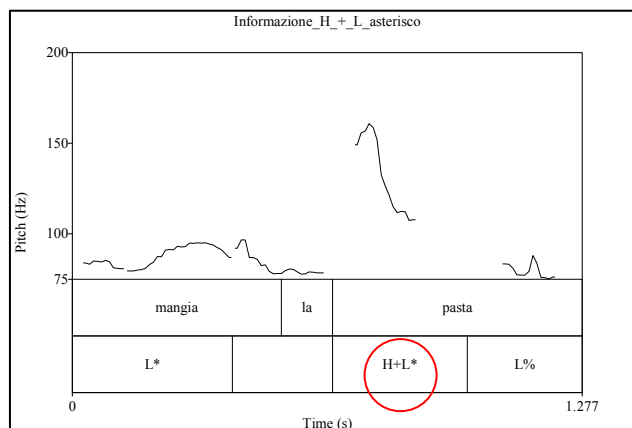


Figura 5. H+L*: focus ristretto informativo FI (Avesani 2003; Avesani & Vayra 2003, 2004).

I soggetti che hanno ascoltato i segnali corrispondenti alle cinque varianti di contorno non solo non sono stati in grado di predirne la funzione pragmatica (tipo di focus, provenienza geografica del parlante), ma non hanno nemmeno saputo distinguere percettivamente un contorno dall'altro.

Il mio esperimento non è ancora terminato. In letteratura si trovano, tuttavia, verifiche rigorose del medesimo fenomeno della cosiddetta 'bassa categorialità' delle opposizioni intonative. I risultati, come si è detto prima, sono controversi, ma sembrano tutti alludere al fatto che le opposizioni intonative sono di natura prototipica (Kuhl 1991; Schneider et al. 2006; Gili Fivela 2008)¹⁰. Alla luce di questa ipotesi, il motivo per cui i tipi intonativi L*+H, H*+L, H*, L+H*, H+L* sono scarsamente (o meno accuratamente) discriminati potrebbe risiedere nel fatto che si tratta di categorie percettive vicine al medesimo prototipo (H).

Riferimenti bibliografici

Avesani, Cinzia (2003) La prosodia del focus contrastivo. Un accento particolare?. In G. Marotta & N. Nocchi (eds.) *La coarticolazione. Atti delle XII Giornate di studio del Gruppo di Fonetica Sperimentale dell'AIA*, 157-167. Pisa.

Avesani, Cinzia & Vayra, Mario (2003) Broad, narrow and contrastive focus in Florentine Italian. In M. J. Solé, D. Recasens & J. Romero (eds), *Proceedings of the 15th International Congress of Phonetic Sciences*, 1803-1806. Barcellona, 3-9 agosto 2003.

Avesani, Cinzia & Vayra, Mario (2004) *Focus ristretto e focus contrastivo in italiano*. In Albano Leoni F., Cutugno F., Pettorino M & Savy R. (eds), *Atti del Convegno*

¹⁰ La natura prototipica dell'intonazione può essere interpretata anche in senso iconico, come risulta, ad esempio, in un passo di Bolinger (1985: 106): *intonation is part of a gestural complex, a relatively autonomous system with attitudinal effects that depend on the metaphorical associations of up and down – an elaborate scheme of iconism. It assists grammar – in some instances may be indispensable to it – but is not ultimately grammatical.*

"Il parlato italiano", CD- ROM, F01, 1-20. Napoli, 13-15 febbraio 2003, M. D'Auria Editore - CIRASS.

Bolinger, Dwight (1985) The inherent iconism of intonation. In John Haiman (ed.), *Iconicity in Syntax*, 97-108. Amsterdam & Philadelphia: Benjamins.

D'Imperio, Mariapaola & House, David (1997) Perception of questions and statements in Neapolitan Italian. In G. Kokkinanis & N. Fakotakis & E. Dermatas (eds.), *Proceedings of Eurospeech '97*, Rodi, vol. 1: 251-254.

De Dominicis, Amedeo (1992) *Intonazione e contesto*. Alessandria: Edizioni dell'Orso.

Falé, Isabel & Faria, Isabel Hub (2005) Intonational contrasts in European Portuguese: a categorical perceptual approach. In *Proceedings of Interspeech 2005*, 1705-1708. Lisbona.

Frascarelli, Mara (2004) L'interpretazione del *focus* e la portata degli operatori sintattici. In Albano Leoni F. & Cutugno F. & Pettorino M. & Savy R. (eds.) *Atti del Convegno "Il Parlatto Italiano"*, (13-15 febbraio 2003), CD- ROM, B06. Napoli, M. D'Auria Editore - CIRASS.

Frascarelli, Mara, & Hinterhölzl, Roland (2007) Types of topics in German and Italian. In Suzanne Winkler, & Kerstin Schwabe (eds.), *On information structure, meaning and form*, 87-116. Amsterdam/ Philadelphia: John Benjamins.

Gartenberg, Robert & Panzlaff-Reuter, Christa (1991) Production and perception of F0 peak patterns in German. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)* 25, 29-115. University of Kiel.

Gili Fivela, Barbara (2005) La percezione degli accenti: il ruolo dell'allineamento e dello 'scaling' dei bersagli tonali. In Piero Cosi (ed.), *Atti del primo Convegno Nazionale dell'Associazione Italiana di Scienze della Voce*, 313-326. Padova: EDK Editore, CD-ROM.

Gili Fivela, Barbara (2006a) The coding of target alignment and scaling in pitch accent transcription. *Italian Journal of Linguistics* 18.1: 189-221.

Gili Fivela, Barbara (2006b) 'Scaling' e allineamento dei bersagli tonali: l'identificazione di due accenti discendenti. In Renata Savy & Claudia Crocco (eds.), *Atti del secondo Convegno Nazionale dell'Associazione Italiana di Scienze della Voce*, 214-232. Torriana (RN): EDK Editore, CD-ROM.

Gili Fivela, Barbara (2008) *Intonation in Production and Perception. The case of Pisa Italian*. Alessandria: Edizioni dell'Orso.

Gussenhoven, Carlos (1999) Discreteness and gradience in intonational contrasts. *Language and Speech* 42: 283-305.

Gussenhoven, Carlos (2004) *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Gussenhoven, Carlos (2006) Experimental approaches to establishing discreteness of intonational contrasts. In Stefan Sudhoff & Denisa Lenertova & Roland Meyer & Sandra Pappert & Petra Augurzy & Ina Mleinek & Nicole Richter & Johannes Schliesser (eds.), *Methods in empirical prosody research*, 321-334. Berlin, New York: de Gruyter.

Kohler, Klaus (1987) Categorical pitch perception. *Proceedings of the 11th International Congress of Phonetic Sciences (ICPhS)*, vol. 5: 331-333. Tallinn.

Kohler, Klaus (1990) Macro and micro F0 in the synthesis of intonation. In John Kingston & Mary E. Beckman (eds.), *Papers in Laboratory Phonology I: between the grammar and physics of speech*, 115-138. Cambridge: CUP.

- Kohler, Klaus (1991) Terminal intonation patterns in single-accent utterances of German: phonetics, phonology and semantics. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)* 25: 115-185. University of Kiel.
- Kohler, Klaus (2006) Paradigms in experimental prosodic analysis: from measurements to function. In Stefan Sudhoff & Denisa Lenertová & Roland Meyer & Sandra Pappert & Petra Augurzky & Ina Mleinek & Nicole Richter & Johannes Schließer (eds.), *Methods in Empirical Prosody Research*, 123-151. Berlin, New York: de Gruyter.
- Kuhl, P. K. (1991) Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* 50: 93-107.
- Ladd, D. Robert & Morton, Rachel (1997) The perception of intonational emphasis: continuous or categorical?. *Journal of Phonetics* 25: 313-342.
- Massaro, Dominic W. (1998) Categorical perception: important phenomenon or lasting myth?. In R. H. Mannell & J. Robert-Ribes (eds.), *Proceeding of the 5th International Congress of Spoken Language Processing*, Vol. 6., 2275-2279. Sydney.
- Massaro, Dominic W. & Cohen, Michael M. (1983) Phonological context in speech perception. *Perception and Psychophysics* 34: 338-348.
- Niebuhr, Oliver & Kohler, Klaus (2004) Perception and cognitive processing of tonal alignment in German. In B. Bel & I. Marlien (eds.), *Proceedings of the International Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages (TAL)*: 155-158. Beijing.
- Remijsen, Bert & van Heuven, Vincent J. (1999) Gradient and categorical pitch dimensions in Dutch: diagnostic test. In *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)*, 3: 1865-1868. San Francisco, CA.
- Schneider, Katrin & Lintfert, Britta & Dogil, Grzegorz & Möbius, Bernd (2006) Phonetic grounding of prosodic categories. In Stefan Sudhoff & Denisa Lenertová & Roland Meyer & Sandra Pappert & Petra Augurzky & Ina Mleinek & Nicole Richter & Johannes Schließer (eds.), *Methods in empirical prosody research*, 335-361. Berlin: de Gruyter.
- Vanrell Bosch, Maria del Mar (2006) A scaling contrast in Majorcan Catalan interrogatives. In Rüdiger Hoffmann & Hansjörg Mixdorff (eds.), *Proceedings of Speech Prosody 2006*, 807-810. Dresden: TUDpress.