



5° Convegno Nazionale AISV

4-6 febbraio 2009, Università di Zurigo

La dimensione temporale del parlato

The temporal dimension of speech

Indice

Programma	3 - 5
Elenco dei poster	6
Indice degli abstract	7
Abstract (in ordine alfabetico)	8 - 85
Elenco dei partecipanti	86
Comitato scientifico	87
Comitato organizzatore	88
Sponsor	89

Mercoledì, 4 febbraio 2009

08:30	Registrazione
	Apertura del convegno
09:00-09:30	Piero Cosi, Presidente AISV Andreas Fischer, Rettore UZH Consegna Premio "F. Ferrero" 2009
	Relazione invitata
09:30-10:30	The neural mechanisms of temporal acoustic information processing during speech perception (Martin Meyer)
10:30-11:00	Pausa caffè
	Sessione orale: Percezione e riconoscimento (presiede Renata Savy)
11:00-11:30	The anticipatory perception based on events (APE) hypothesis (Béatrice Vaxelaire, Rudolph Sock, Fabrice Hirsch, Véronique Ferbach-Hecker, Johanna-Pascale Roy, Fayssal Bouarourou)
11:30-12:00	Stabilità dei parametri nello <i>speaker recognition</i> : F0, durata e <i>articulation rate</i> (Luciano Romito, Rosito Lio, Pier Francesco Ferri, Sabrina Giordano)
12:00-12:30	Multigranularità temporale e processi paralleli nel riconoscimento del parlato nell'uomo e nelle macchine (Francesco Cutugno)
12:30-14:00	Pausa pranzo
	Tavola rotonda: Ritmo
14:00-16:00	<i>Different ways of analyzing speech rhythm</i> William Barry (chair), Volker Dellwo, Antonio Romano, Christopher Lee
16:00-16:30	Pausa caffè
	Sessione orale: Ritmo e melodia (presiede Carlo Schirru)
16:30-17:30	Un confronto tra diverse metriche ritmiche usando Correlatore 1.0 (Paolo Mairano, Antonio Romano)
17:00-17:30	L'ipotesi delle classi ritmiche e i dialetti italo-romanzi (Stephan Schmid)
17:30-18:00	Assetti intervallari nel canto <i>a mutetus</i> della Sardegna meridionale (Paolo Bravi)
18:15-19:15	Concerto nell'Aula Magna
19:30	Ricevimento al <i>Romanisches Seminar</i>

Giovedì, 5 febbraio 2009	
	Sessione orale: Fonetica, fonologia e linguistica (presiede Michele Loporcaro)
09:00-09:30	Studio elettropalatografico dell'assimilazione nei nessi di nasale + velare in italiano (Irene Ricci, Silvia Calamai, Chiara Bertini, Chiara Celata)
09:30-10:00	Interrogative e assertive in un corpus dialettale recuperato (Amedeo De Dominicis, Pamela Mattana)
10:00-10:30	Quanti processi di elisione categorica ed opzionale nel fiorentino parlato? (Luigia Garrapa)
10:30-11:00	Pausa caffè
11:00-12:30	Sessione Poster
12:30-14:00	Pausa pranzo
	Sessione orale: Dialettologia (presiede Pietro Maturi)
14:00-14:30	Sul fenomeno paragogico nella varietà dialettale di Bitti: aspetti temporali e intensità sonora (Barbara Balloi, Carlo Schirru)
14:30-15:00	Diagnostica fonologica e diagnosi fonetica. Ossitoni lunghi in sillaba libera a Sambuca Pistoiese (PT) (Lorenzo Filipponio, Nadia Nocchi)
15:00-15:30	Lo spazio di -u nel dialetto di Matelica (Tania Paciaroni)
15:30-16:00	Confini prosodici e variazione segmentale. Analisi acustica dell'alternanza monottongo/dittongo in alcuni dialetti dell'Italia meridionale (Giovanni Abete, Adrian P. Simpson)
16:00-16:30	Pausa caffè
16:30-18:00	Assemblea soci AISV
19:30	Cena sociale

Venerdì, 6 febbraio 2009

	Sessione orale: Acquisizione ed emozioni (presiede Cinzia Avesani)
09:00-09:30	Percezione e produzione dei fonemi dell'Inglese Americano in parlanti con un sistema pentavocalico (Bianca Sisinni, Mirko Grimaldi)
09:30-10:30	Presentazione ed esemplificazione di "PHON", un programma per la codifica e l'analisi automatica degli aspetti segmentali del parlato (Claudio Zmarich, Maria Pia Bardozzetti, Caterina Pisciotta, Serena Bonifacio)
10:00-10:30	Alcune considerazioni sull'importanza degli aspetti dinamici nella percezione, produzione ed elaborazione del parlato (Piero Cosi)
10:30-11:00	Enfasi e confini prosodici in due stili di eloquio emozionale (Pier Luigi Salza, Enrico Zovato, Morena Danieli)
11:00-11:30	Pausa caffè
	Relazione invitata
11:30-12:30	From sound to rhythm expectancy (Eric Keller)
12:30-14:00	Pausa pranzo
	Sessione orale: Percezione e lingue seconde (presiede Luciano Romito)
14:00-14:30	Does a talker's own rate of speech affect his/her perception of others' speech rate? (Sandra Schwab)
14:30-15:00	L'accento lessicale in spagnolo LS: esperimenti percettivi con italofoeni e francofoeni a confronto (Iolanda Alfano, Sandra Schwab, Renata Savy, Joaquim Llisterri)
15:00-15:30	Fonologia e percezione. Riflessioni su un modello di percezione dei suoni non-nativi (Mario Vayra, Cinzia Avesani)
15:30-16:00	Persistenza dell'accento straniero. Uno studio percettivo sull'italiano L2 (Giovanna Marotta)
16:00	Chiusura del convegno

Sessione Poster, Giovedì 5 febbraio 2009, 11:00 – 12:30	
	in ordine alfabetico secondo il primo autore
Francesco Cangemi	Il dettaglio fonetico nella dinamica dei contorni intonativi
Roberto Cirillo Gianpaolo Coro Alessandro Bruni	Un Interprete di Dialoghi VoiceXML basato su Automi Finiti
Piero Così	Recenti sviluppi di SONIC per l'italiano: riconoscimento automatico del parlato infantile
Mauro Falcone Antonino Barone Alessandro Bonomi Anna Grazia Santoro Maria Dell'Osso	Loudness e "livello del dialogo" nelle trasmissioni radiotelevisive
Dalia Gamal	La dimensione temporale in tre tipi di parlato: un confronto tra arabo e italiano
Vincenzo Galatà Luciano Romito	Un corpus sperimentale per lo studio cross-linguistico europeo delle emozioni vocali
Nuria Kaufmann Martin Meyer Stephan Schmid	Phonological Contrasts in Foreign Language Learning: A Neuropsychological Study on Palatal Affricates
Bogdan Ludusan Serena Soldo	Sonority based syllable segmentation
Edoardo Mastantuoni	Gradatum diatopico, continuum diafasico e dinamiche diagenazionali: primi risultati di un'indagine in Terra di Lavoro
Renzo Miotti	Fenomeni d'armonica vocalica in area friulana e iberica
Nadia Nocchi Silvia Calamai	Durata e strutture formantiche nel parlato toscano: una indagine preliminare su campioni di parlato semispontaneo
Caterina Pisciotta Massimiliano Marchiori Claudio Zmarich	Balbuzione e coarticolazione
Antonio Romano Paolo Mairano Barbara Pollifrone	Variabilità ritmica di varietà dialettali del Piemonte
Serena Soldo Bogdan Ludusan	Statico vs dinamico, un possibile ruolo della sillaba nel riconoscimento automatico del parlato
Arianna Uguzzoni	Alcune note sulle opposizioni di quantità vocalica
Alessandro Vietti Lorenzo Spreafico Antonio Romano	Tempi e modi di conservazione delle <i>r</i> italiane nei <i>frigoriferi CLIPS</i>

Indice degli abstract (primo autore)

Abete, Giovanni	p. 8
Alfano, Iolanda	p. 10
Balloi, Barbara	p. 12
Bravi, Paolo	p. 14
Cangemi, Francesco	p. 16
Cirillo, Roberto	p. 18
Cosi, Piero	p. 20, 22
Cutugno, Francesco	p. 24
De Dominicis, Amedeo	p. 26
Falcone, Mauro	p. 28
Filipponio, Lorenzo	p. 30
Galatà, Vincenzo	p. 32
Gamal, Dalia	p. 34
Garrapa, Luigia	p. 36
Kaufmann, Nuria	p. 38
Keller, Eric	p. 40
Ludusan, Bogdan	p. 42
Mairano, Paolo	p. 44
Marotta, Giovanna	p. 46
Mastantuoni, Eduardo	p. 48
Meyer, Martin	p. 50
Miotti, Renzo	p. 53
Nocchi, Nadia	p. 54
Paciaroni, Tania	p. 56
Pisciotta, Caterina	p. 58
Ricci, Irene	p. 60
Romano, Antonio	p. 62
Romito, Luciano	p. 64
Salza, Pierluigi	p. 66
Schmid, Stephan	p. 68
Schwab, Sandra	p. 70
Sisinni, Bianca	p. 72
Soldo, Serena	p. 74
Uguzzoni, Arianna	p. 76
Vaxelaire, Béatrice	p. 78
Vayra, Mario	p. 80
Vietti, Alessandro	p. 82
Zmarich, Claudio	p. 84

CONFINI PROSODICI E VARIAZIONE SEGMENTALE. ANALISI ACUSTICA DELL'ALTERNANZA MONOTTONGO/DITTONGO IN ALCUNI DIALETTI DELL'ITALIA MERIDIONALE

Giovanni Abete, Adrian Paul Simpson
Friedrich-Schiller-Universität, Jena
giovanni.abete@libero.it

SOMMARIO

Il presente contributo espone i primi risultati di una ricerca triennale che ha riguardato la variabilità del vocalismo tonico in quattro dialetti dell'Italia meridionale (Pozzuoli e Torre Annunziata in Campania, Belvedere Marittimo in Calabria, Trani in Puglia). In particolare, ci si è concentrati su un fenomeno di alternanza sincronica tra esiti monottongali e esiti dittongali, che è tra le caratteristiche più interessanti dei dialetti su menzionati. Come l'analisi evidenzierà, questa alternanza è particolarmente sensibile alla presenza di determinati confini prosodici, con i dittonghi che emergono in corrispondenza dei confini di ordine gerarchicamente superiore.

Questa ricerca si inserisce in maniera originale nel filone degli studi che negli ultimi anni hanno indagato gli effetti della struttura prosodica non solo a livello soprasegmentale ma anche a livello segmentale (ad es. Fougeron/Keating 1997; Keating et al. 2003; Cho 2004; Cho et al. 2007). Rispetto agli studi citati, generalmente condotti su parlato di laboratorio, la presente ricerca si contraddistingue per l'uso di parlato spontaneo e di varietà substandard quali i dialetti italiani. Questa scelta ha imposto una riflessione approfondita su diversi problemi metodologici, dalla modalità di elicitazione del parlato, alle tecniche di analisi acustica, ai metodi statistici per un'adeguata interpretazione dei dati. Questi aspetti verranno discussi nel presente contributo, con particolare attenzione alla metodologia di analisi dei fenomeni di variazione. L'approccio che qui si adotta parte da un'analisi rigorosa dei dati che emergono dall'uso, per risalire induttivamente ai *patterns* di variazione (cfr. Simpson 1992 e 2006). Le varianti vengono qui definite in rapporto reciproco e non come trasformazioni di unità di livello più astratto imposte preliminarmente all'analisi (è questo il caso delle analisi fonologiche dei fenomeni di parlato connesso basate su forme di citazione). In questo contesto grande attenzione è posta ai dettagli fonetici, che non sono il risultato di trasformazioni automatiche di livello post-lessicale, ma fanno parte della competenza implicita del parlante, e provvedono a costruire significati lessicali, grammaticali e interazionali (cfr. Local 2003 e 2007).

L'analisi è stata condotta su un corpus di 24 parlanti e circa 19 ore di registrazioni, raccolte sul campo da uno degli autori con una versione modificata dell'intervista libera (cfr. Como 2006). I parlanti, tutti maschi adulti, sono per la maggioranza pescatori con basso livello di istruzione, ad eccezione dei parlanti di Belvedere Marittimo, che hanno istruzione mediamente più alta e sono impiegati nel terziario. Il corpus è stato segmentato ed etichettato in maniera parziale, limitando questo lavoro a una lista di parole precedentemente selezionate. L'etichettatura ha previsto diversi livelli: segmenti, parole, sintagmi intonativi, enunciati. Altre informazioni prosodiche sono state inserite nella trascrizione al livello segmentale. Questa impalcatura ha consentito di distinguere tra *tokens* vocalici in diverse posizioni prosodiche: interna al sintagma fonologico, finale di sintagma fonologico ma interna al sintagma intonativo, finale di sintagma intonativo. Per la posizione finale di sintagma intonativo, che riveste una particolare importanza per il fenomeno in esame, sono state operate ulteriori distinzioni in base al tipo di andamento melodico di confine e alla funzione pragmatica da esso assolta (ad es. interrogativa, di "continuazione", tipo "lista" etc.).

Sulla base dei *tokens* vocalici etichettati sono state effettuate analisi acustiche della durata e della struttura formantica. Le durate sono state ottenute direttamente dai *files* di etichettatura, mentre per l'analisi della traiettoria dittongale sono state realizzate delle *scripts* in *Snack* e *tcl/tk* per la stima automatica dei valori formantici, valori che sono poi stati controllati manualmente e, quando necessario, corretti. Un coefficiente di dittongazione per ciascun *token* è stato calcolato sulla base dei valori in *Bark* delle prime due formanti. Il metodo di descrizione della traiettoria dittongale utilizzato nella presente ricerca costituisce una evoluzione del metodo di Holbrook e Fairbanks (1962), e segue essenzialmente Simpson (1998), con qualche differenza rispetto all'algoritmo per la stima delle traiettorie formantiche.

In questo contributo si presenteranno alcuni dati relativi alle variazioni di durata e ai coefficienti di dittongazione nelle diverse posizioni prosodiche, quindi si fornirà una caratterizzazione acustica delle traiettorie dittongali e una descrizione dei *patterns* di variazione che regolano l'alternanza monottongo/dittongo, evidenziando similarità e differenze nelle quattro varietà indagate.

RIFERIMENTI BIBLIOGRAFICI

- Cho, T., McQueen, J. M. Cox, E. A. (2007), "Prosodically driven phonetic detail in speech processing: the case of domain-initial strengthening in English". *Journal of Phonetics*, 35, 210-243.
- Cho, T. (2004), "Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English". *Journal of Phonetics*, 32, 141-176.
- Como, P. (2006). Elicitation techniques for spoken discourse. In: Brown, K. (a cura di) *Encyclopedia of language and linguistics*, seconda edizione, volume 4. Amsterdam, Elsevier: 105-109.
- Fougeron, C., Keating, P. A. (1997). "Articulatory strengthening at edges of prosodic domains". *Journal of the Acoustical Society of America*, 101 (6), 3728-3740.
- Holbrook, A., Fairbanks, G. (1962). "Diphthong formants and their movements". *Journal of Speech and Hearing Research*, 5, 33-58.
- Keating, P., Cho, T., Fougeron, C., Hsu, C. (2003). Domain-initial articulatory strengthening in four languages. In Local, J., Ogden, R., Temple, R. (a cura di) *Phonetic interpretation*, Papers in Laboratory Phonology 6. Cambridge, Cambridge University Press: 143-161.
- 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". In: *Proceedings of the International Congress of Phonetic Sciences XVI*, 1-10.
- Simpson, A. P. (1992). "Casual speech rules and what the phonology of connected speech might really be like". *Linguistics* 30: 535-548.
- Simpson, A. P. (1998). Characterizing the formant movements of German diphthongs in spontaneous speech. In Schröder, B., Lenders, W., Hess, W., Portele, T. (a cura di), *Computer Linguistik und Phonetik zwischen Sprache und Sprechen*, Tagungsband der 4. Konferenz zur Verarbeitung natürlicher Sprache – KONVENS – 98. Frankfurt, Lang: 192-200.
- Simpson, A. P. (2006). Phonetic processes in discourse. In: Brown, K. (a cura di), *Encyclopedia of language and linguistics*, seconda edizione, volume 9. Amsterdam, Elsevier: 379-385.

L'ACCENTO LESSICALE IN SPAGNOLO LS: ESPERIMENTI PERCETTIVI CON ITALOFONI E FRANCOFONI A CONFRONTO

Iolanda Alfano¹, Sandra Schwab², Renata Savy³, Joaquim Llisterri¹,

¹Universitat Autònoma de Barcelona; Université de Genève²; Università di Salerno³

Iolanda.Alfano@campus.uab.cat, sandra.schwab@unige.ch, rsavy@unisa.it, Joaquim.Llisterri@uab.cat

SOMMARIO

Lo studio dei meccanismi che governano il processo percettivo in lingua straniera (LS) ha da sempre destato grande interesse; vari sono stati i tentativi di sistematizzarne le fasi e di comprendere la relazione con le competenze in produzione dei soggetti non nativi. I modelli teorici sviluppati si sono concentrati prevalentemente sulla percezione di tratti segmentali; se pur con le debite differenze tra i vari approcci, si riscontra in letteratura l'attribuzione di un peso rilevante alla lingua materna. Anche studi sperimentali, condotti su tratti soprasegmentali sembrano indicare una forte incidenza del fattore lingua madre sulla percezione (Cutler *et al.*, 1986; Otake *et al.*, 1993; Wang *et al.*, 1999).

Il lavoro che presentiamo verte sull'analisi del comportamento percettivo di soggetti italofoeni e francofoeni, in risposta a stimoli in lingua spagnola, alla luce delle diverse caratteristiche accentuali, tanto in produzione come in percezione, delle tre lingue.

Il disegno sperimentale impiegato è stato già seguito in lavori analoghi con soggetti nativi (Llisterri *et al.*, 2005; Alfano, 2006) e non nativi (Alfano *et al.*, 2007; Alfano *et al.*, in stampa; Schwab e Llisterri, in stampa). I corpora impiegati consistono in parole e non parole trisillabiche pronunciate in isolamento, con struttura sillabica CV e con tre diversi profili accentuali: proparossitono (PP), parossitono (P) e ossitono (O). Ciascun item è stato ripetuto 10 volte; di ogni vocale è stata analizzata la durata e la frequenza fondamentale ad inizio, centro e fine vocale. Gli stimoli proposti nei test sono stati così manipolati: per ogni item è stato creato uno stimolo "di base" con i valori medi di durata e frequenza delle dieci ripetizioni; alle parole proparossitone sono stati sostituiti i valori di durata e di f_0 delle corrispettive parossitone (che indichiamo con PP>P) e alle parossitone i valori delle ossitone (P>O). Al fine poi di valutare il peso dell'interazione dei due correlati, è stata effettuata una manipolazione dei valori congiunti di durata ed f_0 . Analisi e sintesi sono state realizzate mediante Praat (Boersma e Weenink, 2003). Gli stimoli sono stati proposti in ordine *random* a tre gruppi di soggetti nativi di italiano e due gruppi di francofoeni (divisi a seconda del livello di conoscenza della LS) per una prova di identificazione della sillaba tonica.

Gli italofoeni percepiscono correttamente la sede accentuale in percentuali che sfiorano il 100% negli stimoli di base con profilo PP e P, ma presentano globalmente oltre un 15% di errori nella percezione delle O. La sola manipolazione di un parametro, f_0 o durata, non risulta sufficiente ad indurre un cambio nel profilo accentuale percepito. L'alterazione congiunta di ambedue i parametri, invece, determina un cambio in percezione in percentuali discrete (tra il 56,7% e il 90%, considerando parole e non parole) per la categoria PP>P, ma in percentuali, in ogni caso, inferiori al 40% per P>O.

Sebbene, dunque, in italiano come in spagnolo l'accento sia libero e quindi dotato di valore fonologico, la percezione in LS risulta complessa. Si evince, in primo luogo, una difficoltà nella percezione dei profili O, sia negli stimoli di base che in quelli manipolati di parole e non parole, a differenza di quanto avviene in L1 (Llisterri *et al.*, 2005; Alfano, 2006). Tale risultato è imputabile, con buona probabilità, alle caratteristiche acustiche degli stimoli: le vocali toniche in parole O hanno durata mediamente diversa in spagnolo e in italiano (extra-lunghe nel primo, tendenzialmente brevi nel secondo, cfr Alfano *et al.*, in stampa). In secondo luogo, emergono differenze notevoli nelle strategie percettive impiegate in L1 vs. LS: i parametri dominanti in L1 vengono accantonati, ma non si verifica una reale 'sostituzione' con i parametri percettivi della L2; il risultato è specularmente a quello ottenuto in un altro esperimento su soggetti ispanofoni con stimoli italiani (Alfano *et al.*, 2007) e mette in luce un forte grado di incertezza e confusione nella percezione.

I francofoeni identificano correttamente la sede accentuale negli stimoli di base in percentuali globalmente superiori al 70% ottenendo la migliore performance sulle PP e la peggiore sulle O. Così come per gli italofoeni, l'alterazione di un solo parametro non determina in maniera chiara un cambio in percezione, mentre la manipolazione congiunta di durata ed f_0 provoca un cambio, ancora una volta, in percentuali globalmente più elevate per PP>P (fino al 77% dei casi) che per P>O (in percentuali che oscillano tra il 32 e il 44% dei casi). La situazione è dunque simile a quella verificata con gli italofoeni. Date le caratteristiche dell'accento in francese, inoltre, ci si aspetterebbe che il profilo ossitono sia quello percepito più facilmente; si riscontra nei dati ottenuti, invece, la tendenza opposta. È possibile che la consapevolezza dei soggetti della libertà accentuale dello spagnolo abbia favorito una propensione verso i profili non presenti in francese.

I risultati ottenuti indicano che l'interpretazione della lingua madre che agisce come filtro nella percezione in LS risulta, quanto meno, non esauriente, suggerendo che le strategie percettive messe in atto non sono solo "linguistiche", ma dipendono da un insieme di fattori psicolinguistici esterni e anche puramente "acustici".

RIFERIMENTI BIBLIOGRAFICI

- Alfano, I. (2006). La percezione dell'accento lessicale: un test sull'italiano a confronto con lo spagnolo, in Savy R., Crocco C. (a c. di), *Teorie, modelli e sistemi di annotazione, Atti del II Convegno Nazionale AISV*, Salerno 2005, Padova: EDK Editore, pp. 632-656.
- Alfano, I., Llisterri, J., Savy, R. (2007). The perception of Italian and Spanish lexical stress: A first cross-linguistic study, in *Proc. of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany. pp. 1793-1796.
- Alfano, I., Savy, R., Llisterri, J. (in stampa). Sulla realtà acustica dell'accento lessicale in italiano ed in spagnolo: la durata vocalica in produzione e percezione, in *Atti del IV Convegno Nazionale AISV*, Arcavacata di Rende (CS), 3-5 dicembre 2007.
- Boersma P., Weenink, D. (2003). Praat: doing phonetics by computer (V. 4.0.4). <http://www.praat.org/>
- Cutler, A. y Mehler, J., Norris, D., Segui, J. (1986). The syllable's differing role in the segmentation of French and English, in *Journal of Memory and Language*, 25, pp. 385-400.
- Llisterri, J., Machuca, M., de la Mota, C., Riera, M., Ríos, A. (2005). "La percepción del acento léxico en español", in *Filología y lingüística*. Madrid: CSIC-UNED-U. de Valladolid, 1, pp. 271-297.
- Otake, T., Hatano, G., Cutler, A., Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese, in *Journal of Memory and Language*, 32, pp. 258-278.
- Schwab, S. y Llisterri, J. (in stampa). La perception de l'accent lexical espagnol par des apprenants francophones, manoscritto non pubblicato.
- Wang, Y., Spence, M., Jongman, A., Sereno, J (1999). Training American listeners to perceive Mandarin tones, in *JASA*, 106, pp. 3469- 3658.

SUL FENOMENO PARAGOGICO NELLA VARIETÀ DIALETTALE DI BITTI: ASPETTI TEMPORALI E INTENSITÀ SONORA.

Barbara Balloi, Carlo Schirru
Università degli Studi di Sassari
ska@uniss.it, balloibarbara@yahoo.it

SOMMARIO

La paragoge, particolare forma di epentesi, si manifesta come noto con la comparsa di una vocale aggiuntiva in parole ossitone o in parole terminanti per consonante.

Secondo quanto attestato da Wagner¹, le varietà dialettali sarde (come d'altra parte anche l'italiano regionale di Sardegna²) si caratterizzano per una forte propensione alla paragoge dovuta, in primo luogo, ad una refrattarietà di fondo per l'ossitonia in monosillabi e bisillabi, nel cui contesto i parlanti reagiscono con l'aggiunta della vocale finale *e*, come nei seguenti casi: *so* (io sono) > *soe*, *tu* (tu) > *tue*, *chi?* (*chi?*) > *chie?*³.

L'elemento paragogico vi si presenta, d'altra parte, in veste di vocale d'appoggio in finale di parola terminante per consonante, quando questa sia pronunciata in isolamento, davanti a pausa o in condizioni d'eloquio a velocità lenta. Il fenomeno è riscontrabile, in particolare, dopo le consonanti finali *s*, *t*, *n*, *nt*, *r* dei plurali maschili e femminili di sostantivi e aggettivi, dei pronomi e degli avverbi e delle desinenze della coniugazione verbale del sardo⁴: *féminas* (femmina) > *féminasa*, *bònos* (buoni) > *bònosos*, *nèm(m)os* (nessuno) > *nèm(m)ososo*, *fin(ts)as* (perfino) > *fin(ts)asa*, *sèmper* (sempre) > *sèmpere*, *fachet* (egli-ella fa) > *fachete*, *aian* (loro avevano, s. nuorese) > *aiana*, *aiant* (loro avevano, s. logudorese) > *aianta*, *cértas* (certe) > *cértasa*; ciò, alla stessa stregua di quanto avviene nel corrispondente italiano regionale il quale, nel caso di un prestito quale *ticket* (*ticket*), da luogo all'esito *tickete*. Contesti, questi, al cui interno la vocale paragogica presenta da un lato la medesima tipologia timbrica della corrispondente che precede la consonante finale; mentre, da un altro, come affermato da Pittau oltreché ben evidenziato da Contini - il quale ha condotto sul fenomeno una specifica indagine acustica in alcune parlate dell'area meridionale dell'isola -, lo stesso segmento si caratterizza anche per una pronuncia relativamente più lieve, secondo Pittau, o meno intensa, secondo Contini⁵.

Partendo da tali premesse, il presente lavoro, che è parte integrante di un più ampio progetto di ricerca sperimentale rivolto tra l'altro allo studio acustico del vocalismo dell'italiano regionale di Sardegna prodotto in area bittese, prende in esame, al momento, gli aspetti temporali e di intensità dell'insieme delle vocali paragogiche presenti in un corpus della stessa varietà, comparativamente alle rispettive vocali di riferimento.

Nello specifico, ai fini dell'indagine vengono utilizzate delle produzioni indotte, costituite da enunciati relativi alle generalità - quali il nome e il cognome delle persone, l'età e una serie di informazioni inerenti la composizione delle rispettive famiglie, unitamente alle attività legate al tempo libero e alle materie scolastiche preferite - di un campione di 12 locutori bittesi, 6 maschi e 6 femmine, di età compresa tra i 9 e gli 11 anni. Campione scelto comunque all'interno di un gruppo più ampio di 46 soggetti ai quali è stato precedentemente somministrato un questionario linguistico teso a rilevare, in maniera concorrenziale all'italiano, la frequenza d'uso del sardo e i relativi argomenti di riferimento.

Le produzioni sonore sono state registrate nei locali dell'istituto dell'Infanzia Primaria di Bitti. Allo scopo, ci si è avvalsi di un microfono NGS MS4pro (risposta in frequenza 20-20.000 Hz, 6x5 – 58 dB \pm 3 dB; impedenza: 32 / Hz; sensibilità: 105 dB/m W) e del programma Audacity, versione 1.2.5, utilizzato su piattaforma Macintosh e impostato ad una frequenza di campionamento di 44.1 KHz e rispettiva quantizzazione a 16 bit. Operando sempre su piattaforma Macintosh, il segnale è stato in seguito normalizzato a -3 dB con l'ausilio dello stesso software Audacity; esso è stato quindi ricampionato a 22.05 KHz, con l'uso, questa volta del programma PRAAT (versione 4.6.27).

¹ Cfr. M.L. Wagner, "La lingua sarda. Storia, spirito e forma", Giulio Paulis (a cura di), Ilisso, 2001, pp. 286-87.

² Cfr. M. Lorinzi, "Sociolinguistica della ricerca linguistica. Punti di vista divergenti sulle consonanti scempie e geminate nell'italiano di Sardegna", Actas do XIX Congreso Internacional de Lingüística e Filoloxía Románicas (Universidade de Santiago de Compostela, 1989), a cura di Ramón Lorenzo, A Coruña, 1996, pp. 311 – 334.

³ Cfr. M. Pittau: "Grammatica del sardo nuorese", Patron 1972, p.18; "Grammatica della lingua sarda: varietà logudorese", Delfino, 1991, p.39.

⁴ Cfr. M. Contini, "Etude de Géographie Phonétique et de Phonétique Instrumentale du Sarde", Edizioni dell'Orso, 1987, pp. 462-468. Per l'area campidanese cfr. in particolare C. Schirru, "Mandas: testo poetico orale in trascrizione fonetica", Quaderni Patavini di Linguistica, Monografie, 6, UP, 1989, pp. 319-355.

⁵ Cfr. Pittau, cit. 1972, p.17; Contini, cit. 1987, pp.462-468.

Quest'ultimo software è stato inoltre utilizzato quale standard di riferimento per l'effettuazione della totalità delle misurazioni di durata e intensità dei segmenti del corpus in analisi. Si tenga comunque presente che in alcuni casi di particolare complessità tecnica si è anche fatto riferimento al software Multi-Speech (Copyright (c) 1995-1999 Speech Technology Research, Ltd), su piattaforma Windows. Il segnale è stato di seguito segmentato secondo una procedura mista, ampiamente testata nel tempo ⁶, di natura uditivo-impressiva oltreché acustico-visiva, basata sull'osservazione dei parametri acustici di frequenza e intensità sonora all'interno in particolare dei rispettivi sonagrammi.

I dati ottenuti sono stati inseriti in una specifica matrice e codificati ai fini della successiva analisi statistica (quantitativa e qualitativa insieme), effettuata - ancora su piattaforma Macintosh - con l'uso del pacchetto Systat.

I risultati permettono di disporre di indicazioni oggettive originali sulla specifica area linguistica in analisi, utili anche in termini di confronto con quanto emerso in particolare dal già detto studio di Contini.

⁶ Cfr. fra altri: C. Schirru, "Aspetti vocalico-temporali dell'italiano in Sardegna. Primi dati sperimentali", Atti delle 4e Giornate di Studio del Gruppo di Fonetica Sperimentale (A.I.A.), Torino, 11-12 novembre 1993, XXI, 1994, pp. 131-140; "Peculiarità temporali nel vocalismo dell'italiano in Piemonte", Atti delle 6e Giornate di Studio del Gruppo di Fonetica Sperimentale (A.I.A.), Roma, 23-24 Novembre 1995, XXIII, 1996, pp. 61-69; "Ulteriori considerazioni prosodiche sul vocalismo italiano del Piemonte", Bollettino dell'Atlante Linguistico Italiano, III Serie, 20, 1996, pp. 79-100; J. E. Flege, C. Schirru, I.R.A. MacKay, "Interaction between the native and second language phonetic subsystems", Speech Communication, 2003, 40, pp. 467-491.

ASSETTI INTERVALLARI NEL CANTO *A MUTETUS* DELLA SARDEGNA MERIDIONALE

Paolo Bravi

pa.bravi@tiscali.it

SOMMARIO

La definizione e l'analisi delle scale, oltre ad essere stata uno degli aspetti più rilevanti dell'analisi in ambito etnomusicologico fin dalle origini della disciplina (Ellis 1885, Hornböstel 1913, Sachs 1982 ed. or. 1962), è un elemento apparentemente ineludibile (cfr Nattiez 1981) nella definizione dei caratteri di un "sistema musicale" (Giannattasio 1998).

È stato frequentemente osservato che le strutture intervallari relative a stili di canto non assimilabili ai generi della musica classica (o di sua derivazione) diffusi in ambito extraeuropeo e folklorico hanno caratteristiche diverse rispetto al sistema euroculto (Giuriati 1991). Tale diversità da un lato riguarda la definizione dei gradi su cui si basa il sistema musicale (ad esempio, le strutture pentatonali, sostanzialmente estranee al sistema occidentale classico, hanno una diffusione ampia in stili di canto presenti in società e culture geograficamente assai lontane), dall'altro lato riguarda l'ampiezza degli intervalli⁷, raramente assimilabile al sistema temperato che da circa tre secoli costituisce il sistema di riferimento in ambito occidentale colto.

Lo stile di canto *a mutetus* riguarda la forma più importante di poesia improvvisata diffusa nella Sardegna meridionale. Viene praticato in maniera prevalente da poeti improvvisatori (*cantadoris*) semi-professionisti, principalmente in occasione di gare poetiche organizzate in concomitanza con feste patronali. Le esecuzioni si svolgono in buona parte in forma solistica, mentre il passaggio dall'una all'altra sezione della forma metrico-musicale è segnato dagli interventi in polifonia di un coro bivocale composto dalla voce gutturale del *basciu* (con emissione simile a quella del *bassu* dei cori a tenore diffusi nell'area centrale della Sardegna) e dalla voce – anch'essa talvolta in parte gutturale – della *contra*, cui in genere si sovrappone la voce del poeta improvvisatore.

Nel caso del canto *a mutetus*, pur essendoci una terminologia specifica relativamente articolata (Bravi 2007/08, Zedda 2008), non esiste una "etnoteoria" (Cardona 1985, Baily 2005) sistematica che permetta di inquadrare i dati rilevabili all'ascolto relativi agli assetti intervallari di questo stile di canto nel quadro di una teoria *emic*⁸. Per questo motivo la ricerca si è svolta unicamente in forma sperimentale, adottando due metodologie diverse. La prima [metodo A] è quella inaugurata, agli inizi degli anni Settanta, da Tjernlund, Sundberg e Fransson (Tjernlund et al. 1972). Il presupposto di base di tale metodo è che le frequenze prevalenti in una esecuzione siano il correlato acustico della scala in uso. La realizzazione di istogrammi di frequenza che mostrano la distribuzione delle frequenze fondamentali offre indicazioni essenziali sugli assetti intervallari del canto. Il metodo da un lato presenta il vantaggio di offrire una valutazione di sintesi oggettiva sui comportamenti intonativi dei cantori, dall'altro presenta limiti legati al fatto che la distribuzione delle frequenze non permette di distinguere fenomeni particolari ma assai comuni nell'intonazione cantata quali, ad es. il vibrato o il portamento. Ovvio a questo tipo di inconvenienti il secondo tipo di metodologia di ricerca adottato [metodo B], che prevede l'esame puntuale dei contorni intonativi realizzati dai cantori, specie nei punti in cui all'ascolto si avverte la presenza di movimenti melodici fondati su intervalli non comuni.

Gli assetti intervallari del canto *a mutetus* evidenziano caratteristiche significativamente diverse rispetto agli intervalli definiti dal sistema temperato. L'esame svolto attraverso il metodo [A] mostra alcune differenze evidenti nel comportamento intonativo della maggioranza degli esecutori rispetto alla struttura intervallare temperata. In modo particolare, il II, il III e il IV grado mostrano assetti ricorrenti e spesso condivisi da diversi cantori, seppur non costanti. L'analisi puntuale dei percorsi melodici [metodo B] mostra da un lato i limiti del concetto di 'nota' (intesa come 'punto stabile' che discretizza lo spazio diastematico, vd. Sorce Keller 1990), dall'altro i contesti melodici in cui appaiono alcuni fra i tratti intonativi più caratteristici rilevabili nel canto *a mutetus*.

⁷ Per *intervallo* si intende nel lessico musicologico la "[m]isura delle differenze d'altezza fra due gradi di una scala" (Baroni 2004: 245).

⁸ Il concetto di *emic* / *etic*, di derivazione linguistica, è stato da tempo applicato anche all'ambito delle scienze umane, e in particolare all'antropologia. Il riferimento d'origine è in questo senso Pike (1954). Per quanto riguarda l'applicazione della distinzione nell'ambito dell'analisi etnomusicologica, v. Macchiarella (2000).

Il lavoro che si intende presentare si incentra sui punti seguenti:
 Sintetica presentazione del genere di canto (con dimostrazione video) e descrizione del corpus analizzato
 Descrizione del metodo di analisi [A] e di alcuni risultati in evidenza
 Descrizione del metodo di analisi [B] e di alcuni risultati in evidenza, con ascolti delle versioni originali comparate a versioni in cui l'intonazione è modificata artificialmente e ricondotta agli intervalli del sistema temperato
 Conclusioni e ipotesi di *future work*

RIFERIMENTI BIBLIOGRAFICI

- BAILY, JOHN
 2005 *La teoria musicale nelle tradizioni orali. In Enciclopedia della musica. Vol. V: L'unità della musica.*, Einaudi, Torino: 537-554.
- BARONI, MARIO
 2004 *L'orecchio intelligente. Guida all'ascolto di musiche non familiari.* Libreria Musicale Italiana, Lucca.
- BRAVI, PAOLO
 2007/08 *A sa moda campidanese. Pratiche, poetiche e voci degli improvvisatori nella Sardegna meridionale.* Tesi di Dottorato, Dottorato di Ricerca in "Metodologie della ricerca etnoantropologica", Università di Siena, tutors: prof. Giulio Angioni, prof. Ignazio Macchiarella).
- CARDONA, GIORGIO RAIMONDO
 1985 *La foresta di piume. Manuale di etnoscienza.* Laterza, Roma-Bari.
- ELLIS, ALEXANDER
 1885 *On the musical scales of various nations.* In "Journal of the Society of Arts", 33: 485-527.
- GIANNATTASIO, FRANCESCO
 1998 *Il concetto di musica.* Bulzoni, Roma (ed. or. La Nuova Italia Scientifica, 1992).
- GIURIATI, GIOVANNI
 1991 *Scala.* In AGAMENNONE, MAURIZIO / FACCI, SERENA / GIANNATTASIO, FRANCESCO / GIURIATI, GIOVANNI, *Grammatica della musica etnica.* Bulzoni, Roma: 83-112.
- HORNBÖSTEL, ERICH M. VAN
 1913 *Melodie und Skala.* In "Jahrbuch der Musikbibliothek Peters", 19: 11-23.
- MACCHIARELLA, IGNAZIO
 2000 *Analisi ed etnomusicologia. Una introduzione.* In "Bollettino di analisi e teoria musicale", VII, 1: 9-71.
- NATTIEZ, JEAN JACQUES
 1981 *Scala.* In "Enciclopedia", vol. XII, Einaudi, Torino: 454-470.
- PIKE, KENNETH
 1954 *Language in relation to a unified theory of the structure of human behaviour.* Summer Institute of Linguistics, Glendale.
- SACHS, CURT
 1982 ed. or 1962 *Le sorgenti della musica.* Boringhieri, Torino (ed. or. *The wellspring of music.* Martinus Nijhoff, The Hague, 1962).
- SORCE KELLER, MARCELLO
 1990 *Alcune considerazioni per uno studio analitico della melodia nelle trascrizioni popolari e in quelle extraeuropee.* In appendice a DE NATALE, MARCO, *Analisi della struttura melodica.* Guerini e ass. Milano: 209-229.
- TJERNLUND, PER et al. [TJERNLUND, PER / SUNDBERG, JOHAN / FRANSSON, FRANS]
 1972 *Grundfrequenzmessungen an schwedischen Kernspaltflöten.* In "Studia Instrumentorum Musicae Popularis", 2: 77-96.
- ZEDDA, PAOLO
 2008 (a cura di) *L'arte di is mutetus. Il canto e l'improvvisazione nei poeti sardi del Campidano.* Edizioni Gorée, Iesa.

IL DETTAGLIO FONETICO NELLA DINAMICA DEI CONTORNI INTONATIVI

Francesco Cangemi
Laboratoire Parole et Langage, Université de Provence
francesco.cangemi@lpl-aix.fr

SOMMARIO

Recenti studi sulla fonologia dell'intonazione hanno sottolineato l'importanza di fattori dinamici (forma del picco accentuale, interazione con la curva di intensità, etc.) nell'analisi dei contorni intonativi dell'italiano di Napoli [1, 2], dell'inglese [3] e del tedesco [4]. Nel quadro standard della fonologia metrico-autosegmentale i fattori dinamici della curva di F0 sono considerati come dettaglio fonetico non rilevante nella determinazione della forma fonologica di un contorno intonativo. Ad esempio, se tra due bersagli tonali alti si può distinguere una transizione di tipo lineare da una a catenaria [5], si postula che i due toni di un accento bitonale possano essere prodotti esclusivamente attraverso un'interpolazione lineare tra i due target, e che fattori quali la forma dell'interpolazione e/o la pendenza del movimento siano conseguenza dell'ancoraggio segmentale dei toni [6]. Alcuni dati di italiano napoletano mostrano invece che, in contesti pragmatici differenti, il contorno intonativo relativo a una stessa stringa segmentale può differire sistematicamente anche per il tipo di movimento (concavo o convesso) tra due bersagli tonali [7] (Fig. 1).

In questo studio si fornisce anzitutto una documentazione di diversi tipi di interpolazione in due contorni tonali prenucleari, sottolineando l'importanza del percorso che la curva di frequenza fondamentale esegue nel tempo. In secondo luogo si mostra che indici usuali come l'allineamento e l'altezza dei bersagli iniziali e finali non sembrano differenziarsi in maniera sufficiente da motivare l'esistenza di un contrasto percettivo robusto. Si propone dunque una modellizzazione di queste differenze attraverso un indice di curvatura, capace di rappresentare in maniera più efficace i due tipi di interpolazione.

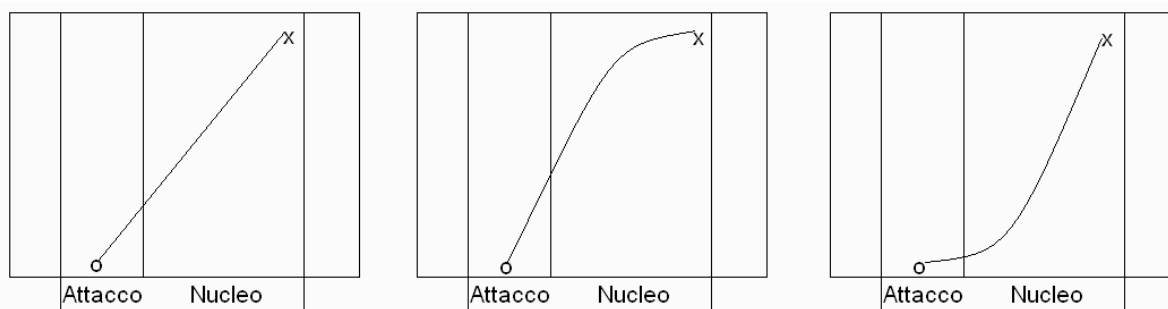


Figura 1. Stilizzazioni dell'interpolazione tra due bersagli o ed x: lineare (a sinistra); non lineare, convessa (al centro); non lineare, concava (a destra)

L'analisi verrà condotta su un corpus di italiano napoletano [8]. Tre parlanti hanno letto cinque ripetizioni di tre frasi bersaglio, disegnate senza occlusive sorde al fine di garantire un'estrazione meno problematica dei valori di frequenza fondamentale (es. « Valeria viene alle nove »). Prima di pronunciare ognuna delle frasi, i parlanti hanno letto silenziosamente un paragrafo di contestualizzazione; in questo modo si è potuto manipolare il contesto pragmatico degli enunciati. Ogni frase è stata registrata nel contesto di domanda polare a focalizzazione stretta (es. «È proprio Valeria che viene alle nove?») e in quello di affermazione con topic contrastivo parziale (es. «Valeria viene alle nove, ma degli altri invitati non so nulla») [9].

Per ognuna delle novanta frasi a disposizione si sono misurate alcune proprietà relative all'ascesa della F0 sulla prima sillaba tonica (es. Valeria). Anzitutto, il bersaglio basso e quello alto sono stati rispettivamente individuati nell'attacco e nel nucleo della sillaba. Per i due bersagli sono state misurate l'altezza in hertz e l'allineamento in millisecondi rispetto a diversi segnapiosti segmentali (inizio della sillaba, inizio della vocale, fine della sillaba). Per ogni movimento ascendente, inoltre, si è rilevata l'altezza del punto intermedio sull'asse del tempo tra i due bersagli tonali (Fig. 2).

Le altezze dei tre punti sono state utilizzate per calcolare un indice di curvatura (*range proportion* [10]). Infine, i valori di ogni punto del contorno della F0 nella sillaba protonica e in quella tonica sono stati utilizzati per calcolare un punto di flessione nel contorno intonativo (*two-line regression* [1]). I risultati mostrano che la differenza tra i contorni intonativi dei due contesti pragmatici è rappresentata in maniera più robusta da questi due indici che da quelli, tradizionali, di allineamento dei bersagli tonali.

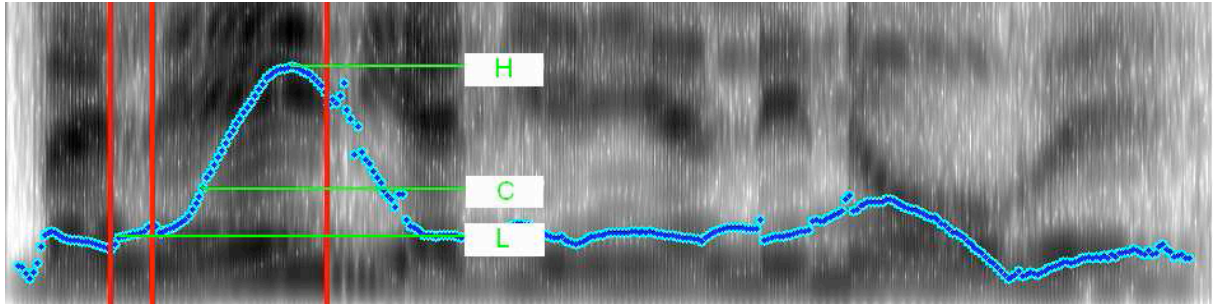


Figura 2. Spettrogramma e curva F0 per l'enunciato "Valeria viene alle nove" nel contesto di domanda polare a focalizzazione stretta. Le linee verticali rosse delimitano l'attacco ed il nucleo della prima sillaba tonica. Le linee orizzontali verdi indicano l'altezza del bersaglio basso (L), di quello alto (H) e del punto intermedio nel tempo tra i due (C).

RIFERIMENTI BIBLIOGRAFICI

- [1] D'Imperio, Mariapaola (2000a): *The role of perception in defining tonal targets and their alignment*, PhD Dissertation, The Ohio State University
- [2] D'Imperio, Mariapaola - Terken, Jacques - Pitermann, Michel (2000b): "Perceived tone «targets» and pitch accent identification in Italian", in *Proceedings of Australian International Conference on Speech Science and Technology (SST)*, 201-211
- [3] Knight, Rachael-Anne - Nolan, Francis (2006): "The effect of pitch span on intonational plateaux", *Journal of the International Phonetic Association*, 36 (1): 21-38
- [4] Niebuhr, Oliver (2007): "The signalling of German rising-falling intonation categories - the interplay of synchronization, shape, and height", *Phonetica*, 64 (2-3): 174-193
- [5] Pierrehumbert, Janet (1980): *The phonology and phonetics of English intonation*. PhD Dissertation, MIT
- [6] Ladd, D. Robert - Faulkner, D. - Faulkner, H., - Schepman, Astrid (1999): "Constant 'segmental' anchoring of f0 movements under changes in speech rate", *Journal of the Acoustical Society of America*, 106: 1543-1554
- [7] Petrone, Caterina - D'Imperio, Mariapaola (2008): "Tonal structure and constituency in Neapolitan Italian Evidence for the Accentual Phrase in statements and questions", *Speech Prosody*
- [8] D'Imperio, Mariapaola - Cangemi, Francesco - Brunetti, Lisa (2008): "The phonetics and phonology of contrastive topic constructions in Italian", in *Third Conference on Tone and Intonation (TIE3)*
- [9] Büring, Daniel (1997): *The Meaning of Topic and Focus – the 59th Street Bridge Accent*, London, Routledge
- [10] Dombrowski, Ernst – Niebuhr, Oliver (2005): "Acoustic patterns and communicative functions of phrase-final F0 rises in German: activating and restricting contours", *Phonetica*, 62: 176-195
- [11] Dilley, Laura - Ladd, Robert D. - Schepman, Astrid (2005): "Alignment of L and H in bitonal pitch accents: testing two hypotheses", *Journal of Phonetics*, 33 (1): 115-119

UN INTERPRETE DI DIALOGHI VOICEXML BASATO SU AUTOMI FINITI

Roberto Cirillo, Gianpaolo Coro, Alessandro Bruni

roberto.cirillo@abla.it, gianpaolo.coro@abla.it, alessandro.bruni@abla.it

SOMMARIO

L'articolo che presentiamo riguarda l'interpretazione di dialoghi telefonici sviluppati utilizzando il linguaggio VoiceXML.

I recenti sviluppi nell'ambito della telefonia hanno riguardato principalmente l'espansione del concetto di "risponditore automatico" (IVR). Se prima questo era inteso come una semplice aggiunta ad un centralino telefonico, per fornire capacità minimali di smistamento intelligente delle richieste, in arrivo ad un call center, oggi questo tipo di soluzione si orienta alla gestione di un dialogo complesso con un utente, per mezzo di riconoscimento e sintesi vocale, permettendo di svolgere operazioni anche complicate. In quest'ottica, negli ultimi anni, il mercato ha mostrato esigenze di rinnovo tecnologico sia nell'ambito del riconoscimento e della sintesi vocale, che in quello dei dialoghi. Il Voice XML è nato proprio allo scopo di descrivere e gestire tali interazioni, nell'ottica dello sviluppo di sistemi integrativi per i call center sempre più efficaci.

Il panorama commerciale, nel settore IVR, è molto variegato, e le diverse implementazioni di interpreti VoiceXML hanno portato questo linguaggio ad allontanarsi dagli standard suggeriti dal consorzio W3C. Di fatto, le implementazioni di applicazioni IVR in uso si basano su variazioni dagli standard a volte profonde.

In questo articolo descriveremo un'implementazione di un interprete per il VoiceXML, che cerca di superare i limiti interpretativi imposti dagli standard ed è abbastanza elastico da riuscire ad interpretare i "dialetti" utilizzati dagli applicativi commerciali più in uso.

La motivazione per un tale interesse è che le norme redatte dal W3C, per la costruzione di un interprete per applicazioni *VXML-compliant*, suggeriscono l'uso di un descrittore DTD, ma la norma a nostro avviso non è sufficientemente robusta e contestualizzata per l'ambito specifico delle applicazioni vocali, e ciò ne scoraggia l'uso da parte dei produttori. Nell'esame degli interpreti presenti sul mercato abbiamo sperimentato un nuovo approccio che consiste, dato un documento VXML, nella estrapolazione di una struttura sufficientemente astratta, riconducibile ad un Automa a Stati Finiti, che fa uso di informazione proveniente dall'esterno per proseguire nella computazione.

Le difficoltà sono nella deduzione di una struttura a stati finiti a partire da un documento, e nell'estensione del concetto di automa finito a quello di macchina a stati finiti la cui stessa topologia dipende da un agente esterno. L'automa infatti prende delle decisioni in base alle risposte del suo interlocutore, ma al contempo, essendo dinamica la generazione delle pagine VoiceXML, la struttura stessa dell'automa viene modificata in base a tale risposta. Per tale motivo abbiamo deciso di indicare questi modelli con il nome di "Automati ad Oracolo", in analogia alle "Macchine di Turing ad Oracolo", delle quali sono sottoinsiemi. Nel nostro caso l'Oracolo è l'interlocutore umano, che fornisce il risultato ad una richiesta di interazione.

L'articolo dimostra come un tale approccio sia più elastico nell'interpretazione dei documenti VoiceXML, rispetto ad una soluzione basata sull'interpretazione della DTD standard. Anche i dialetti, entro certi limiti, possono essere interpretati e trattati correttamente.

Altro punto a favore dell'approccio ad automi è la loro espandibilità, per inglobare nuovi tag o strutture XML, introdotte nelle versioni successive del VoiceXML.

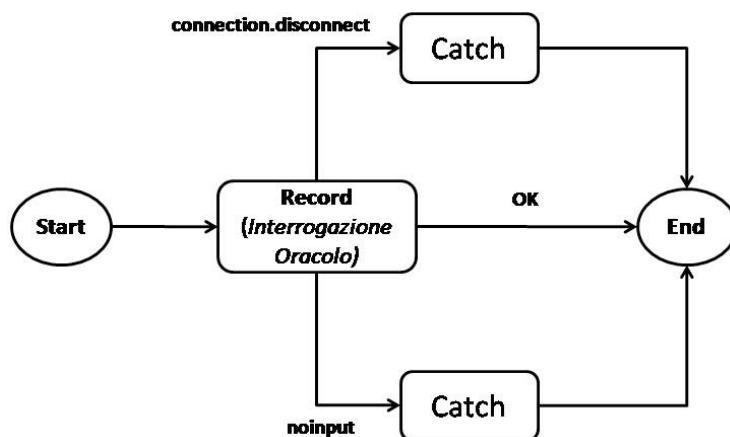


Figura 1. Automa Finito, costruito da un dialogo Voice XML

In figura 1 è rappresentato l'automa generato dal nostro sistema dopo una successione di pagine VoiceXML, per un'applicazione telefonica nella quale si chiede all'utente di registrare un messaggio audio. Il blocco *Record* rappresenta l'interazione con la persona, alla quale questa può rispondere senza parlare (evento *noinput*), oppure riagganciando la cornetta del telefono (evento *connection.disconnect*), oppure registrando la propria voce (in tal caso si procede al termine del dialogo).

Un esempio di differenza in efficacia interpretativa è visibile nell'interpretazione di alcuni tag proprietari dei produttori di IVR, come ad esempio `<genesys:field>`, dove l'etichetta *genesys* è definita nel namespace esterno <http://www.genesyslab.com/vxml/2.0/ext/20020430>.

Un interprete VoiceXML strettamente legato alla DTD 2.0 del W3C, considererebbe tale tag non valido, ignorando le richieste contenute nel field. Un automa come il nostro, invece, bada alla struttura XML più che al nome del tag, e questo consente di contemplare namespaces esterni, a differenza di quanto richiesto dal VoiceXML, e di trattare il tag `<genesys:field>` come un tag *field* senza etichetta.

Queste capacità consentono al nostro interprete di analizzare molti più documenti VoiceXML, adattandosi ad applicazioni vocali sviluppate da produttori diversi, senza perdite importanti di informazione.

I modelli qui presentati sono stati inseriti in un IVR ed hanno mostrato delle prestazioni computazionali del tutto soddisfacenti per applicazioni commerciali.

Il VoiceXML è un linguaggio che ha bisogno di una grossa flessibilità dal punto di vista dell'espressività, se si tiene conto anche del fatto che il grado di complessità delle applicazioni vocali è in crescente aumento. Attualmente tale flessibilità non è concepita nella specifica della DTD ufficiale, almeno per quanto riguarda la versione 2.0. L'XML permette di aggirare questi problemi in quanto è un linguaggio di *Markup Informativo*, a differenza del VoiceXML che potremmo definire di *Markup Elaborativo*, perché ha spesso anche il compito di elaborare le risposte. Per questo motivo è difficile pensare che possa esistere una DTD robusta, che includa le esigenze sentite da tutti gli sviluppatori IVR e che li metta d'accordo. In quest'ottica, un approccio adattativo come il nostro, presenta delle ottime prospettive di utilizzo futuro.

RECENTI SVILUPPI DI SONIC PER L'ITALIANO: RICONOSCIMENTO AUTOMATICO DEL PARLATO INFANTILE

Piero Cosi

ISTC-spdf CNR, Istituto di Scienze e Tecnologie della Cognizione, Sede di Padova "Fonetica e Dialettologia",
Consiglio Nazionale delle Ricerche, via Martiri della Libertà, 2 - 35127 Padova, Italia
piero.cosi@pd.istc.cnr.it

SOMMARIO

In questo lavoro vengono descritti i risultati dei più recenti esperimenti di riconoscimento automatico di parlato infantile effettuati, mediante l'utilizzazione del sistema denominato SONIC, su un corpus di parlato letto da bambini di età compresa fra i 7 e i 13 anni.

Il corpus utilizzato è stato raccolto presso alcune scuole del Trentino da parte dell'ITC-IRST (Giuliani & Gerosa, 2003) ora FBK (Fondazione Bruno Kessler), nell'ambito di un progetto europeo denominato PF-STAR.

Il tasso di errore di riconoscimento iniziale di 15.1% per un insieme di 33 unità fonetiche (21,8% considerando un insieme di 40 unità fonetiche) è stato successivamente ridotto al 12.2% (18,6% considerando 40 unità) utilizzando una combinazione delle più aggiornate tecniche di adattamento comprendenti la normalizzazione di lunghezza del tratto vocale (Vocal Tract Length Normalization VTLN), la normalizzazione della varianza dei coefficienti Cepstrali (Cepstralcoefficients Variance Normalization, CVN) e l'utilizzazione di modelli fonetici addestrati in modalità indipendente dal parlante utilizzando le più recenti strategie iterative denominate Structural MAP Linear Regression (SMAPLR) e Speaker Adaptive Training (SAT).

Mentre il tasso di errore del sistema allenato su voci di bambini è paragonabile e addirittura migliore di quello ottenuto da sistemi simili sullo stesso corpus (ad esempio paragonabile al 22.7% ottenuto da un sistema analogo con 28 unità fonetiche come quello utilizzato in Giuliani & Gerosa, 2003)), esiste ancora un significativo margine di miglioramento per un sistema che utilizzi modelli acustici allenati su parlato adulto e utilizzati per decodificare parlato infantile. Infatti quando sono state applicate entrambe le tecniche VTLN e SMAPLR in una condizione di disallineamento adulti/bambini il sistema finale ha ottenuto un tasso di errore fonetico del 19.3% dimostrando di ridurre l'errore fonetico iniziale del 28%. Ciò nonostante, persiste ancora un notevole 30% di differenza relativa fra l'utilizzazione di modelli acustici allenati su parlato adulto e modelli acustici allenati su parlato infantile per la decodifica di quest'ultimo.

CONCLUSIONI E SVILUPPI FUTURI

Lo sviluppo di sistemi di riconoscimento di parlato infantile spesso si presenta come un compito di ardua soluzione a causa della spesso totale mancanza di risorse acustiche utilizzabili per l'allenamento dei modelli acustici. In questo lavoro, il sistema di riconoscimento denominato SONIC e sviluppato per l'inglese è stato adattato all'italiano ed in particolare è stato considerato il caso del parlato infantile di bambini compresi nella fascia di età compresa fra i 7 e i 13 anni.

Questi nuovi modelli acustici per il parlato infantile italiano sono stati incorporati nel *CLT (Colorado Literacy Tutor)*, sviluppato al *CSLR (Centre for Speech and Language Research)* della *University of Colorado di Boulder*, per la lingua inglese, quale primo passo verso lo sviluppo della sua corrispondente versione italiana *l'Italian Literacy Tutor*, un sistema interattivo e personalizzato per l'apprendimento della lingua italiana.

RIFERIMENTI BIBLIOGRAFICI

- [1] The Colorado Literacy Tutor: <http://www.colit.org/>
- [2] Cole R., van Vuuren S., Pellom B., et al. 2003. "Perceptive Animated Interfaces: First Steps Toward a New Paradigm for Human Computer Interaction", in Proc. of the IEEE, vol. 91, no. 9, pp. 1391-1405, Sept., 2003
- [3] Pellom B. 2001. "SONIC: The University of Colorado Continuous Speech Recognizer", Technical Report TR-CSLR-2001-01, University of Colorado, USA, 2001.
- [4] Pellom B. and Hacioglu K. 2003. "Recent Improvements in the CU SONIC ASR System for Noisy Speech: The SPINE Task", Proc. ICASSP, Hong Kong, 2003.
- [5] Hagen A., Pellom B., and Cole R. 2003. "Children's Speech Recognition with Application to Interactive Books and Tutors", Proc. ASRU, St. Thomas, USA, 2003.
- [6] Hagen A., Pellom B., Van Vuuren S., and Cole R. 2004. "Advances in Children's Speech Recognition within an Interactive Literacy Tutor", Proc. HLT-NAACL, Boston Massachusetts, USA, 2004.
- [7] Yapanel U.H., Hansen J.H.L. 2003. "A New Perspective on Feature Extraction for Robust In-Vehicle Speech Recognition", in Proceedings EUROSPEECH 2003, Geneva, Switzerland, September 1-4, 2003, 1281-1284.
- [8] Lee K., Hagen A., Romanyshyn N., Martin S., and Pellom B. 2004. "Analysis and Detection of Reading Miscues for Interactive Literacy Tutors", Proc. 20th Int. Conf. on Computational Linguistics (Coling), Geneva, CH, 2004.
- [9] Hagen A., Pellom B. 2005. "A Multi-Layered Lexical-Tree Based Token Passing Architecture for Efficient Recognition of Subword Speech Units", in 2nd Language & Technology Conference, Poznan, Poland, April, 2005
- [10] Cosi P., Pellom B. 2005. "Italian Children's Speech Recognition For Advanced Interactive Literacy Tutors", in CD-Rom Proceedings INTERSPEECH 2005, Lisbon, Portugal, 2005, pp. 2201-2204.
- [11] Cosi P., Delmonte R., Biscetti S., Cole R., Pellom B. and van Vuuren S. 2004. "Italian Literacy Tutor: tools and technologies for individuals with cognitive disabilities", Proc. InSTIL/ICALL Symposium, Venice, Italy.
- [15] Gerosa M., Giuliani D. and Brugnara F. 2007. "Acoustic Variability and automatic recognition of children's speech", Speech Communication, Vol. 49, 2007, Proc. ICASSP, Hong Kong, 2003.
- [16] Siohan O., Myrvoll T., and Lee C.H. 2002. "Structural Maximum a Posteriori Linear Regression for Fast HMM Adaptation", Computer, Speech and Language, 16, 5-24, Jan, 2002.
- [17] Welling L., Kanthak S., Ney H. 1999. "Improved Methods for Vocal Tract Length Normalization", Proc. ICASSP, Phoenix Arizona, 1999.
- [18] Giuliani D. and Gerosa M. 2003. "Investigating Recognition of Children's Speech", Proc. ICASSP, Hong Kong, 2003.

ALCUNE CONSIDERAZIONI SULL'IMPORTANZA DEGLI ASPETTI DINAMICI NELLA PERCEZIONE, PRODUZIONE ED ELABORAZIONE DEL PARLATO

Piero Cosi

ISTC-spdf CNR, Istituto di Scienze e Tecnologie della Cognizione, Sede di Padova "Fonetica e Dialettologia",
Consiglio Nazionale delle Ricerche, via Martiri della Libertà, 2 - 35127 Padova, Italia
piero.cosi@pd.istc.cnr.it

SOMMARIO

La dimensione temporale è un elemento costitutivo non solo dei meccanismi di produzione del parlato, intervenendo, a livello segmentale, nella determinazione delle durate e nella pianificazione e nel controllo di tutti i gesti articolatori e, a livello soprasegmentale, nell'allineamento dei contorni intonativi con le parti dell'enunciato, ma anche, nella percezione del segnale verbale e, più in generale, nell'interpretazione di un qualsiasi atto comunicativo.

Inoltre, sia la configurazione delle caratteristiche facciali che la sincronizzazione delle azioni facciali sono importanti nell'espressione e nel riconoscimento delle emozioni (Cohn, 2007). La configurazione delle azioni facciali (espressioni relative sia a specifiche emozioni sia ad unità di azione individuali) rispetto alle emozioni ed all'intenzione comunicativa è stata un importante tema di ricerca. Meno invece si conosce circa la sincronizzazione delle azioni facciali, anche perché la misurazione manuale della sincronizzazione è assai complicata e laboriosa. Tuttavia, sappiamo che (Cohn, 2007) siamo altamente sensibili alla sincronizzazione delle azioni facciali nelle interazioni sociali (Edwards, 1998). Le azioni facciali più lente, ad esempio, sembrano essere più genuine e naturali (Krumhuber & Kappas, 2005), come pure lo sembrano essere quelle più sincrone nei loro movimenti (Frank & Ekman, 1997). In particolare, le espressioni facciali più sottili diventano visibili soltanto quando le informazioni di movimento sono a disposizione di chi le percepisce (Ambadar, Schooler, & Cohn, 2005).

La dinamica è cioè particolarmente importante per inferire l'intenzione comunicativa. Alcuni studi condotti dal gruppo di ricerca CMU/Pitt utilizzando tecniche automatiche di analisi di immagini facciali per misurare la sincronizzazione delle azioni facciali, hanno provato che le caratteristiche dinamiche riescono a discriminare fra i sorrisi intenzionali e quelli spontanei con un livello di precisione dell' 89% (Cohn & Schmidt, 2004). Usando caratteristiche simili, il divertimento, l'imbarazzo ed il sorriso "gentile" sono stati discriminati con una precisione dell' 83% (Kanade, Hu, & Cohn, 2005), che è paragonabile a quella umana. Lavori più recenti suggeriscono inoltre che la coordinazione multimodale dell'espressione facciale, del movimento della testa e dei gesti sono caratteristiche specifiche dell'imbarazzo (Keltner, 1995).

In questo lavoro vengono illustrati alcuni dei più significativi apporti tecnologici che nel corso degli ultimi anni sono stati influenzati dalla dimensione temporale del parlato nel campo dell'analisi del segnale vocale, della sintesi della voce da testo scritto, del riconoscimento automatico del segnale verbale (Furui, 2005). Per quanto riguarda la realizzazione di facce parlanti animate, saranno discussi poi alcuni esempi dell'influenza degli aspetti dinamici nella percezione e nella interpretazione delle espressioni facciali e più in generale degli intenti comunicativi, nella trasmissione di emozioni, stati d'animo e atteggiamenti, nell'interazione faccia a faccia (Cohn, 2007).

RIFERIMENTI BIBLIOGRAFICI

- [1] Cohn, J.F. (2007). Foundations of human-centered computing: Facial expression and emotion. *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'07)*, Hyderabad, India, 5-12.
- [2] Edwards, K. (1998). The face of time: Temporal cues in facial expressions of emotion. *Psychological Science*, 9(4), 270-276.
- [3] Krumhuber, E., & Kappas, A. (2005). Moving smiles: The role of dynamic components for the perception of the genuineness of smiles. *Journal of Nonverbal Behavior*, 29, 3-24.
- [4] Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stakes lies. *Journal of Personality and Social Psychology*, 72(6), 1429-1439.
- [5] Ambadar, Z., Schooler, J., & Cohn, J. F. (2005). Deciphering the enigmatic face: The importance of facial dynamics to interpreting subtle facial expressions. *Psychological Science*, 16, 403-410.
- [6] Cohn, J. F., & Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing*, 2, 1-12.
- [7] Kanade, T., Hu, C., & Cohn, J. F. (2005). Facial expression analysis. Paper presented at the IEEE International Workshop on Modeling and Analysis of Faces and Gestures, Beijing, China.
- [8] Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement and shame. *Journal of Personality and Social Psychology*, 68(3), 441-454.
- [9] Furui S. (2005). 50 years of progress in speech and speaker recognition. *Proceedings SPECOM 2005*, Patras, Greece, 1-9.

MULTIGRANULARITÀ TEMPORALE E PROCESSI PARALLELI NEL RICONOSCIMENTO DEL PARLATO NELL'UOMO E NELLE MACCHINE

Francesco Cutugno
Dipartimento di Scienze Fisiche Università di Napoli Federico II – NLP Group
cutugno@na.infn.it

SOMMARIO

Gli studi sul riconoscimento del parlato negli uomini e nelle macchine sono spesso basati su modelli che derivano dall'osservazione del comportamento umano. Molte teorie si confrontano con l'idea di definire una unità di base di tipo atomico che possa essere considerata come la base di partenza per l'avvio dei processi di riconoscimento e comprensione del messaggio verbale.

I sistemi di *Automatic Speech Recognition* (ASR) finalizzati alle applicazioni tecnologiche integrano poi le (poche) conoscenze che si hanno dei meccanismi cognitivi con approcci stocastici basati su apprendimento statistico. Tali processi, quale che sia l'algoritmo stocastico impiegato, di fatto associano un simbolo astratto, in qualche modo riconducibile ad un repertorio finito di oggetti appartenenti ad una specifica categoria linguistica, ad una porzione di segnale acustico. Un secondo processo, normalmente basato su una da regole di concatenazione deterministiche (tecnicamente dette grammatiche) o statistiche (n-grammi) fra questi simboli, prova ad aggiungere conoscenza di tipo *top-down* al processo *bottom-up* legato all'individuazione ed al riconoscimento dell'esistenza di informazione acustica collegata alla sequenza dei simboli.

Se, sul piano delle applicazioni tecnologiche consideriamo questi simboli le nostre unità di base, ci rendiamo conto di quanto sia affascinante il problema di quale unità scegliere per ottimizzare il processo e di quali siano i collegamenti teorici fra queste unità e i livelli normalmente considerati fondamentali nell'analisi linguistica. Più forti saranno questi legami, più semplice sarà il processo di formalizzazione delle loro caratteristiche sia in ambito linguistico che informatico e sarà più facile modellare la conoscenza *top-down* a cui si faceva in precedenza riferimento.

Allo stesso tempo molti lavori recenti [1], provenienti da svariati campi scientifici, sembrano concordare sull'idea che in un segnale parlato l'informazione sia distribuita su diverse scale temporali e che l'uomo sia in grado di riconoscere il parlato proprio grazie ad un'elaborazione parallela utilizzando processi separati che catturano differenti proprietà del segnale collegate a costanti temporali che variano fra i pochi millisecondi (componenti segmentali), alle centinaia (componenti ritmiche) alle frazioni di secondo (componenti frasali). Tale concetto è noto come multigranularità.

Detto in altre parole (per un'ampia discussione su questo tema cfr [2]), si inizia da più parti a ritenere che un modello sufficientemente esplicativo dei meccanismi di riconoscimento e comprensione del parlato negli uomini e nelle macchine, potrebbe essere fortemente basato sull'idea che diversi processi possano contemporaneamente fare uso di unità di base differenti, dando in questo modo luogo ad un processo multiplo parallelizzato e multigranulare.

La ricerca nel settore dell'elaborazione del segnale vocale presenta sempre più spesso risultati interessanti in relazione alla possibilità di effettuare analisi di natura prosodica automaticamente e in molti casi senza conoscere il contenuto informativo-testuale del segnale che si elabora. Si va dall'estrazione di proprietà di tipo ritmico basati su misure di sonorità (per una rassegna si veda [3]), alla segmentazione sillabica automatica (per una rassegna si veda [4]), all'analisi del profilo melodico di un enunciato del quale si forniscono parametri che sono in relazione con accenti sia di parola che di frase [5]. Altri tipi di analisi consentono di ottenere vettori di cosiddette *features* collegate a proprietà spettroacustiche delle componenti a variazione lenta del segnale vocale [6], in opposizione alle normali tecniche spettroacustiche che solitamente descrivono le variazioni fini.

Nel settore del riconoscimento automatico del parlato, numerosi studi effettuati su metodi per la gestione e l'utilizzo di eventi prosodici, affermano che l'introduzione di questi fenomeni risulta essere il passo più importante, allo stato dell'arte, per migliorare le prestazioni dei sistemi automatici del parlato ([7],[8], [9], [10]).

Il presente lavoro passa in rassegna la bibliografia sopra citata, discutendone i punti critici e individuando le implicazioni per una teoria della descrizione dei fenomeni linguistici basati su costanti temporali non segmentali nell'ottica di contribuire alla modellazione dei meccanismi del riconoscimento del parlato nell'uomo e nelle macchine.

Il passo successivo sarà quello di presentare un progetto per un riconoscitore automatico di parlato multilivello in cui alcuni moduli software operano stime su differenti livelli informativi ottenibili direttamente da vari processi di analisi. Ogni processo investiga proprietà del segnale che evolvono con costanti temporali diverse: un primo layer basato su processi markoviani fattoriali costituisce una baseline pseudo-segmentale che avanza nel segnale con un passo di circa 20 ms ma cerca comunque di riconoscere il parlato operando sui processi che evolvono con un fattore temporale di circa 200 ms, un secondo layer, basato su un preprocessing che tenta una segmentazione preliminare del segnale in sillabe e uno strato finale di rescoring che corregge le

stime effettuate dai due strati precedenti sulla base di template ritmici che consentono una parziale revisione degli eventuali errori commessi dai processi diretti.

Le prestazioni fornite dai singoli layers saranno discusse nell'ottica di valutare le possibili strategie di correzione degli errori riscontrati e in quella di individuare la strategia che porti ad un algoritmo che integri delle risposte ricavate in ognuno di essi in maniera ottimale. Verrà infine delineato il problema di un ulteriore livello di integrazione fra i risultati del sistema sopra presentato e le stime fornite dai modelli *top-down* comunque presenti nel sistema.

RIFERIMENTI BIBLIOGRAFICI

- [1] D. Poeppel. The analysis of speech in different temporal integration windows: Cerebral lateralization as asymmetric sampling in time. *Speech Communication*, 41:245-255, 2003.
- [2] G. Coro, A Step Forward in Multi-granular Automatic Speech Recognition, Tesi di Dottorato in Scienze Informatiche e Computazionali, Università di Napoli, Federico II, 2007.
- [3] B. Ludusan, S. Soldo, in questo volume.
- [4] S. Soldo, B. Ludusan, in questo volume.
- [5] The Roles of Physiology, Physics and Mathematics in Modeling Prosodic Features of Speech (Invited Paper), *Proceedings of Speech Prosody 2006*, Dresden, Germany (May 2006).
- [6] G. Coro. Il modulation spectrogram nel riconoscimento automatico del parlato. Atti del I° Convegno Nazionale AISV, 2004, CD-Rom.
- [7] A. Ganapathiraju, J. Hamaker, J. Picone. Syllable-based large vocabulary continuous speech recognition. *Transaction on Speech and Audio Processing*, 9, 358-366, 2001.
- [8] D. Vergyri, A. Stolcke, V. Gadde, L. Ferrer, E. Shriberg. Prosodic knowledge sources for automatic speech recognition. *Proceedings of ICASSP*, 2003:211, 2003.
- [9] E. Shriberg, A. Stolcke. Prosody modeling for automatic speech recognition and understanding. *Proceedings of Workshop on Mathematical Foundations of Natural Language Modeling*, 2002.
- [10] S. King T., Stephenson S., Isard P. Taylor, A. Strachan. Speech recognition via phonetically featured syllables. *Proceedings of ICSLP*, 2:124-145, 1972.

INTERROGATIVE E ASSERTIVE IN UN CORPUS DIALETTALE RECUPERATO

Amedeo De Dominicis, Pamela Mattana
Laboratorio di Fonetica, Università degli Studi della Tuscia
dedomini@unitus.it, pamelamattana@gmail.com

SOMMARIO

Il lavoro prende in esame il *corpus* dialettale recuperato dagli archivi dell'audioteca della Provincia di Viterbo, relativo a registrazioni sul campo effettuate negli anni '70. Tale archivio venne digitalizzato e parzialmente restaurato dal Laboratorio di Fonetica dell'Università della Tuscia (De Dominicis & Mattana in corso di stampa) nell'ambito di un progetto denominato prima Co.Va.I.D. e poi "Calliope" (De Dominicis 2002).

I dati si riferiscono al parlato di Bomarzo, una cittadina vicina a Viterbo.

Abbiamo isolato nel *corpus* a nostra disposizione un gruppo di frasi interrogative e assertive e le abbiamo analizzate con il software *Praat*, estraendone alcuni parametri, riguardanti i valori tonali sulla sillaba nucleare e sulla coda (annotati in To.B.I.), le durate delle sillabe, della parola fonologica, del piede e del sintagma fonologico finali di frase.

Al momento, abbiamo portato a termine l'indagine solo su un piccolo numero di interrogative chiuse (sì/no) e assertive affermative. Il *corpus* completo verrà presentato nel *paper* definitivo.

L'indagine parziale mostra, tuttavia, alcuni risultati. Sul piano tonale non sembrano manifestarsi opposizioni nette tra interrogative e assertive: il *pattern* H*+L ricorre in maniera prevalente sulle sillabe nucleari delle interrogative, ma si trova anche su alcune assertive; altrettanto accade per il *pattern* L% in coda. La sola differenza sembra riguardare il livello di caduta del tono L della sillaba nucleare e il valore del L% sulla coda: nelle interrogative il valore di F0 su L e su L% è più alto.

D'altra parte, si osserva un sistematico incremento di durata delle unità metriche (soprattutto la parola fonologica o il gruppo clitico e il sintagma fonologico, meno il piede) in fine di frase interrogativa. Le assertive, invece, presentano durate inferiori, tranne nel caso di espressioni marcate da *focus* largo: in quest'ultimo caso, entrambe le tipologie frasali presentano unità metriche di lunghezza comparabile.

Tale differenziazione può trovare una spiegazione di tipo iconico. Le interrogative rappresentano forme grammaticali marcate rispetto alle assertive, quindi devono funzionalmente distinguersi. Di solito, tale funzione si esprime per le caratteristiche tonali della sillaba nucleare o comunque della parte finale della frase. Perciò, questa porzione finale viene temporalmente dilatata mediante risorse metriche. Nel caso delle interrogative di Bomarzo, le caratteristiche tonali distintive sono, almeno parzialmente, neutralizzate e, quindi, l'unico elemento di differenziazione con le assertive è rappresentato dall'allungamento delle unità metriche terminali. Tuttavia, si tratta di una caratteristica che qualifica solo un supporto di marcatezza e, quindi, si ritrova anche nelle assertive focalizzate.

Naturalmente, tali risultati devono essere considerati con prudenza, per diversi ordini di ragioni. Innanzitutto, la consistenza del *corpus* analizzato è ancora scarsa. Perciò contiamo di ampliare il numero di casi analizzati nella versione finale del *paper*.

In secondo luogo, appare comunque difficile isolare nel *corpus* delle frasi con modalità grammaticale sicuramente assertiva/interrogativa e ancor più arduo classificare il tipo interrogativo (aperto, chiuso, *check*, ecc.) a causa della natura delle registrazioni utilizzate: si tratta infatti di reperti di una campagna effettuata negli anni '70, con scopi etnografici, da personale non addestrato e in ambiente spesso rumoroso; il materiale registrato è spesso di scarso interesse per i nostri scopi (canzoni popolari, filastrocche); la piccola parte di parlato spontaneo non rumoroso è sovente decontestualizzata (si tratta di singole parole interrogative o di asserzioni fortemente marcate emotivamente e informativamente, ma la conversazione non viene registrata nella sua interezza e quindi non è facile interpretarne la funzione grammaticale e pragmatica).

Inoltre, per esigenze di congruità di comparazione di dati fonologici, abbiamo volutamente escluso le frasi con coda intonativa vuota (es. frasi terminanti con una parola accentata sull'ultima sillaba).

Ciò detto, il fenomeno di allungamento metrico finale di frase delle interrogative non è tuttavia privo di attestazioni nella letteratura scientifica. Ad esempio, Caldwell (1998: 430 sgg.) riporta casi di allungamento vocalico in alcuni pronomi interrogativi tamil; Annie Rialland (2006) presenta casi di interrogative prive di innalzamento di *pitch* finale, marcate invece da allungamento finale in lingue dei gruppi gur, kwa, mande, kru, benue-congo, ijo e bantu; van Heuven e van Zanten (2005) individuano una diversa velocità di eloquio nelle interrogative polari rispetto alle assertive in manado malay (lingua austronesiana) e in due lingue germaniche (la varietà inglese delle isole Orkney e l'olandese).

RIFERIMENTI BIBLIOGRAFICI

- Caldwell R. (1998), *Comparative Grammar of the Dravidian or South-Indian Family of Languages*. Asian Educational Services.
- De Dominicis A. (2002), *Co.Va.I.D. (COnservazione e VAlorizzazione degli archivi vocali dell'Italiano e dei suoi Dialetti)*, *La comunicazione*, numero unico speciale a cura di Giuseppe Rinaldo e Roberto Piraino contenente gli Atti della Conferenza TIPI (Tecnologie Informatiche nella Promozione della lingua Italiana), LI, 97-98.
- De Dominicis A. & Mattana P. (in corso di stampa), *Il Progetto Bomarzo*, Atti del quarto Convegno Nazionale dell'Associazione Italiana di Scienze della Voce 2007, Arcavacata di Rende.
- van Heuven, V.J. & van Zanten, E. (2005), *Speech rate as a secondary prosodic characteristic of polarity questions in three languages*, *Speech Communication* 47: 87–99.
- Rialland, A. (2006), *Question prosody: an African perspective*, in C. Gussenhoven and T. Riad (eds.), *Tones and Tunes: Studies in Word and Sentence Prosody*, Mouton de Gruyter, Berlin.

LOUDNESS E “LIVELLO DEL DIALOGO” NELLE TRASMISSIONI RADIOTELEVISIVE

Mauro Falcone*, Antonino Barone[^], Alessandro Bonomi[^], Anna Grazia Santoro[^], Maria Dell’Osso[^]
Fondazione Ugo Bordoni*, Istituto Superiore delle Comunicazioni e delle Tecnologie dell’Informazione[^]
falcone@fub.it

SOMMARIO

Il segnale audio che riceviamo attraverso i media (radio, tv, internet, ecc.) può essere, e di fatto lo è, pesantemente, affetto da diversi tipi di elaborazioni e alterazioni. È un fatto ben noto che questi segnali, e quindi anche la voce, sono codificati, e quindi compressi perdendo parte della loro originaria informazione, secondo diversi standard (mpeg2, mpeg4, ecc.). Meno noto è il fatto che questi segnali possono essere elaborati in modo da modificarne, con diversi fini, il loro contenuto energetico. In particolare il segnale è sicuramente manipolato (rispetto ai suoi naturali livelli) in fase di mixing, ma ulteriori e più o meno arbitrarie modifiche sono possibili successivamente. Solo negli ultimi anni si è iniziato a studiare il problema del loudness nelle trasmissioni radiotelevisive cercando di risolvere sia il problema del dislivello “channel to channel”, sia del livello “program to program” e infine quello del “program to advertising”. Quest’ultimo caso in particolare è stato oggetto di diversi indagini [1][2] sia in quanto soggetto a normative giuridiche, sia perché aspetto percepito come particolarmente fastidioso dagli ascoltatori. Con la raccomandazione internazionale ITU-R 1770 [3], e le sue successive modifiche, si è dato un primo fondamentale contributo alla soluzione del problema, risolvendo che tipo di misura deve essere effettuata per misurare il livello del segnale audio. Definito pertanto “come” misurare il livello, rimane (almeno) ancora un secondo e fondamentale punto da risolvere, ovvero “quando” effettuare la misurazione. Non è infatti corretto misurare il livello indiscriminatamente su tutto il segnale audio trasmesso per quantificare correttamente il livello del loudness, ovvero del volume percepito dall’ascoltatore, ma devono essere selezionati solamente quelle parti percettivamente rilevanti, trascurando tutto il rimanente. A tal fine sono oggi utilizzati, nella maggior parte dei casi, due diversi approcci: il “dialogue intelligence” ed il “gating”. Nel primo caso il sistema opera una caratterizzazione del segnale in “parlato” e “non parlato” per poi eseguire la misurazione solo sul primo tipo (ipotizzando appunto che in ogni caso la veicolazione maggioritaria delle informazioni avviene attraverso la voce), nel secondo caso invece si definisce una soglia di riferimento tale che tutto ciò che ha valore superiore viene considerato di interesse per la misura, mentre tutto ciò che è inferiore viene tralasciato. Ognuno di questi metodi ha specifici vantaggi, ed ovviamente svantaggi. Se il “dialogue intelligence” può essere operato online senza conoscere il livello medio dell’intensità del segnale, a suo svantaggio c’è il fatto che non può essere utilizzato su segnali musicali, e che risulta computazionalmente complesso, ed è “non deterministico” a meno di utilizzare uno standard aperto che definisca l’algoritmo di voice detection. La tecnica del “gating” al contrario è di facile realizzazione, può essere applicata su qualsiasi tipo di segnale audio, ma è necessario conoscere il livello medio di intensità del segnale perché la soglia deve necessariamente essere relativa. Non ha infatti senso un gating assoluto senza conoscere quale sia il livello medio di intensità sonora dell’audio che stiamo considerando.

In questo lavoro si mostrano i risultati di un’ampia campagna di misura effettuata attraverso strumentazione professionale e l’utilizzo dello strumento LM100 della Dolby. Strumento che è oggi il riferimento internazionale per la misura del loudness attraverso la tecnica di “dialogue intelligence”. Per quanto riguarda il “gating” invece, è stato sviluppato un apposito software dagli autori. Sfortunatamente nell’utilizzo del “dialogue intelligence”, essendo vincolati al funzionamento dello strumento utilizzato, non è possibile variare alcuna funzionalità in quanto il software utilizzato per il “voice detection” è chiuso ed in ogni caso protetto da brevetto. Al contrario il software sviluppato permette una facile configurazione dei parametri e realizza sia la misura di intensità in RMS, sia la normativa ITU-T 1770. La misura di loudness proposta in questa normativa è subito stata accettata con entusiasmo sia dalla comunità scientifica, sia dalle industrie legate alla produzione e diffusione dell’audio nel broadcast, ed è stata implementata nel nostro sistema di misura. Questa misura tiene conto sia dell’effetto di interferenza dell’ascoltatore, o meglio della sua testa che viene approssimata come una sfera di circa 21 cm, sia di una curva di adattamento relativa alla sensibilità dell’apparato uditivo umano per i suoni della stessa classe di quelli comunemente trasmessi dalle televisioni. Queste curve di “adattamento” (A per la telefonia, B per i segnali di media qualità, e via così per C, D ecc.) sono ben note in psicoacustica, e nella raccomandazione in questione, in particolare, si propone una revisione della curva B.

Le due tecniche. “dialogue intelligence” e “gating”, vengono messe a confronto su un’ampia ed eterogenea quantità di programmi, relativi a tutte le principali emittenti nazionali. Da questa prima analisi si evince come sia possibile trovare una equivalenza tra le due metodologie solo parzialmente, e sotto vincoli molto stringenti sia

della tipologia dei contenuti e sia della corretta, o perlomeno omogenea, realizzazione del materiale nella fase di missaggio.

Vengono riportati i risultati di campagne di misura atte a studiare e quantificare tutte e tre le situazioni confronto del loudness (C2C, P2P, P2A).

Il materiale utilizzato è stato acquisito nell'anno corrente principalmente dalle emittenti RAI e Mediaset, e nell'orario di prime time e comunque sempre nell'arco di maggior ascolto.

Infine viene analizzato, in una prima campagna pilota, il loudness in funzione del medesimo programma in diversi contesti di fruizione/trasmissione, ovvero dei diversi media utilizzati nella trasmissione e nella fruizione. Tutto il materiale audio è manualmente etichettato a due livelli: un primo livello ne individua inizio e fine programma secondo i palinsesti relativi, e un secondo livello dove il segnale viene diviso in tre classi ovvero in A (il programma vero e proprio), B (gli spot pubblicitari), e C (che contiene tutto il resto come jingles, prossimamente, ecc.).

I risultati sperimentali riportati costituiscono un punto di riferimento importante, in quanto non ci risulta siano disponibili pubblicamente studi simili a questo. Inoltre lo studio tra le diverse tecniche di selezione è, anch'esso, una novità nel panorama degli studi sul loudness, e vuole essere un primo contributo alla soluzione di questo problema che già vede schierati i diversi produttori di strumentazione di misura su i due fronti opposti. Infine le investigazioni pilota sulle differenze dello stesso contenuto audio attraverso diversi media, e lo studio delle possibili alterazioni dei livelli di parlato rispetto ad una acquisizione lineare di laboratorio costituiscono i punti di partenza per lo sviluppo di nuove attività sperimentali di ricerca.

RIFERIMENTI BIBLIOGRAFICI

[1] Falcone M., Barone A., Bonomi A., *“Abbassa quello “spot”, per favore”*, Atti del 3° Convegno Nazionale AISV, 29 Nov. – 1 Dic., 2006, Trento, Italia

[2] Travaglini A., *“Broadcast loudness: mixing, monitoring and control”*
122nd AES convention, 5-7 Maggio 2007, Vienna, Austria

[3] ITU, *“Algorithms to measure audio programme loudness and true-peak audio level”*, ITU-R BS.1770-1 Recommendation, ITU, 2006, Ginevra, Svizzera

DIAGNOSTICA FONOLOGICA E DIAGNOSI FONETICA. OSSITONI LUNGHII IN SILLABA LIBERA A SAMBUCA PISTOIESE (PT)

Lorenzo Filipponio, Nadia Nocchi
Università di Zurigo
filippon@rom.uzh.ch, nadia_nocchi@yahoo.com

SOMMARIO

~~Lungo la linea La Spezia Rimini (o, meglio, Carrara Fano, cfr. Pellegrini 1992) la distribuzione delle isoglosse che determinano la distinzione tra le varietà gallo-italiche emiliane e quelle toscane è nell'area tra Pistoia e Bologna certamente più compressa rispetto a quella rilevabile nell'Appennino Tosco-Emiliano occidentale, in area reggiano-garfagnina. Cionondimeno, è possibile osservare anche nell'alto Appennino bolognese e pistoiese una ripartizione dei fenomeni la cui distribuzione rende conto della scansione cronologica con cui essi sono avvenuti.~~

~~Un luogo di indagine privilegiato è il territorio di Sambuca Pistoiese (PT), politicamente toscano ma idrograficamente adriatico, costituito da piccole frazioni dislocate lungo quattro valli parallele e isolate tra loro (da ovest verso est: Reno, Limentra Occidentale, Limentrella, Limentra Orientale). Questa situazione geografica ha di fatto permesso, pur nello spazio ridotto di 77 chilometri quadrati, il mantenimento di caratteristiche linguistiche peculiari e differenziate, in un contesto generale di forte conservatività. Per quanto concerne la dimensione temporale del parlato, per esempio, si può osservare che la compensazione ritmica generale nei proparossitoni, caratteristica ascrivibile all'italo-romanzo settentrionale (cfr. Loporearo 2005), interessa tutte le località sambugane, a eccezione di quelle situate nella valle del torrente Limentra Orientale, che mostrano invece una schietta e radicata toscanità. Per esempio, gli esiti per *ST_M_CU e *T_P_DU risultano essere [ʌStom>ego], [ʌtov>edo] a Lagacci (valle del Reno), [ʌStom>go], [ʌtov>do] a Pàvana (valle della Limentra Occidentale), ma [ʌst] [mako], [ʌtj] [epido] a Torri (valle della Limentra Orientale). Nei primi due casi si ha vocale tonica breve e allungamento della consonante postonica, oltre all'indebolimento della vocale atona interna (a Pàvana addirittura cancellazione, similmente ad altre varietà in territorio bolognese, come il lizzanese, per cui cfr. Malagoli 1930), probabilmente in virtù di una trafilata [a], [i] > [ɛ] poi ricostruita come [e] secondo un meccanismo adottato nell'area anche per le [o] atone finali dopo [ʌVn], cfr. a Lagacci [ʌbo] [ne] da [ʌbo] [no] > [ʌbo] [nɛ] (a Pàvana [ʌb_]); nel caso di Torri, invece, si osservano il mantenimento della quantità della vocale tonica scaturita dal protoromanzo allungamento di vocale tonica in sillaba libera e la conservazione del vocalismo atono (cfr. Filipponio in corso di stampa^a).~~

~~Diversa è la situazione se si osserva la distribuzione degli ossitoni lunghi in sillaba libera che — come insegna Martinet (1975, p. 205) — è dirimente per stabilire se le varietà analizzate hanno quantità vocalica distintiva o meno. Infatti, in questa posizione è possibile osservare il comportamento della vocale tonica senza i condizionamenti dovuti al nesso con la consonante postonica: ciò permette di valutare se dietro il percepito allungamento della consonante postonica si nasconde un *close contact* (Martinet 1966) che in queste varietà non può essere considerato fonologicamente pertinente a scapito della quantità vocalica (diversa è la situazione delle varietà emiliane più innovative, in cui a quest'ultima sembra sostituirsi effettivamente il taglio sillabico come arbitro fonologico, cfr. Vennemann 2000 per la discussione teorica e Filipponio in corso di stampa^b per la verifica sui dati del bolognese), oppure una quantità consonantica ancora fonologicamente distintiva, segno di una varietà il cui fondo gallo-italico non si è evoluto ulteriormente nella direzione intrapresa da gran parte dei dialetti eisalpini.~~

~~Le indagini sul campo finora svolte hanno suggerito, rispetto alla compensazione ritmica, una presenza più limitata sul territorio degli ossitoni lunghi in sillaba libera: si tratta di un'ossitonia secondaria, dovuta all'apocope di sillabe finali atone o alla caduta di morfemi di plurale in iato primario o secondario con la vocale tonica ([kan] [ta]) seconda persona plurale del verbo "cantare", di solito in coppia minima con il participio passato [kan] [ta], [lu] da [ʌlu] [i], [ʌfjo] da [ʌfjo] [li] > [ʌfjo] [i] la cui presenza è assodata a Pàvana (cfr. Guccini 1998), la località sicuramente più esposta ai flussi — umani e linguistici — provenienti da settentrione, ma non è mai stata verificata nelle altre frazioni di Sambuca. A questo scopo sono state svolte diverse inchieste sul campo nelle località di Lagacci, Pòsola (valle della Reno), Castello di Sambuca, San Pellegrino al Cassero e Stabiazioni (valle della Limentra Occidentale). Il massiccio spopolamento dell'area ha reso difficoltoso il reperimento di dialettofoni attendibili, ma la quantità dei dati raccolti e la qualità garantita da una parte degli informatori (in particolare per Lagacci, Castello di Sambuca e Stabiazioni) permettono di verificare la presenza — o meno — degli ossitoni lunghi in sillaba libera, e, con ciò, di delimitare l'isoglossa che separa le varietà con~~

quantità vocalica distintiva da quelle, più conservative, che non hanno fatto proprio questo tratto fondante della struttura fonologica.

Per quanto concerne gli aspetti più propriamente sperimentali, le registrazioni sono state effettuate nelle abitazioni degli informatori, in buone condizioni di silenzio, con un registratore digitale Fostex FR-2LE dotato di microfono Sennheiser MK2. Il corpus è costituito da una serie di ossitoni primari romanzi (che anche nelle varietà con quantità vocalica distintiva hanno conservato la brevità della vocale tonica) e di ossitoni secondari, fatti pronunciare in isolamento, in posizione finale di frase e in posizione interna di frase (possibilmente non al confine di sintagma). Ovviamente, l'analisi spettrografica terrà conto primariamente della durata, ma cercherà anche di non tralasciare eventuali aspetti collaterali.

RIFERIMENTI BIBLIOGRAFICI

- Filipponio, L., in corso di stampa^a, *La quantità vocalica nei proparossitoni etimologici al confine tra toscano e gallo-italico*, in *Actes du XXV Congrès de Linguistique et Philologie Romanes*, Innsbruck, 3-8 settembre 2007.
- Filipponio, L., in corso di stampa^b, *La struttura di parola dei dialetti della valle del Reno*, Sala Bolognese, Forni.
- Guccini, F., 1998, *Dizionario del dialetto di Pàvana*, Porretta terme, Gruppo di Studi Alta Valle del Reno Nuèter.
- Loporearo, M., 2005, *La lunghezza vocalica nell'Italia settentrionale alla luce dei dati del lombardo alpino*, in Max Pfister e Gabriele Antonioli (a cura di), *Itinerari linguistici alpini. Atti del convegno di dialettologia in onore del prof. Remo Bracchi*, Bormio, 24-25 settembre 2004, Sondrio, IDEVV.
- Malagoli, G., 1930, *Fonologia del dialetto di Lizzano in Belvedere (Appennino bolognese)*, L'Italia Dialettale, VI, pp. 125-196.
- Martinet, A., 1966, *Close contact*, *Word*, XXII, pp. 1-6.
- Martinet A., 1975, *Evolution des langues et reconstruction*, Paris, Presses Universitaires de France.
- Pellegrini, G.B. 1992, *Il "Cisalpino" e l'italo-romanzo*, *Archivio Glottologico Italiano*, LXXVII, pp. 272-296.
- Vennemann, T., 2000, *From quantity to syllable cuts: on so-called lengthening in the Germanic languages*, *Rivista di Linguistica/Italian Journal of Linguistics*, XII, pp. 251-282.

UN CORPUS SPERIMENTALE PER LO STUDIO CROSS-LINGUISTICO EUROPEO DELLE EMOZIONI VOCALI

Vincenzo Galatà, Luciano Romito
Laboratorio di Fonetica, Università della Calabria
vgalata@libero.it, luciano.romito@unical.it

SOMMARIO

La presente proposta rappresenta il primo stadio di una ricerca attualmente in corso e volta allo studio cross-linguistico europeo delle emozioni vocali in quattro nazioni: Italia, Francia, Inghilterra e Germania.

Lo studio delle emozioni espresse attraverso la voce ha subito negli ultimi anni un rinnovato interesse in diversi ambiti di ricerca e numerosi sono gli studi presenti in letteratura (sia per l'encoding, ricerca dei meccanismi e degli indici acustici interessati nella produzione del parlato emotivo, sia per il decoding, indagine sui processi percettivi e sulla capacità degli esseri umani a decodificare il parlato emotivo). Gli obiettivi perseguiti sono anch'essi assai diversi e spaziano dalla mera indagine tendente a chiarire le regole sottese alla trasmissione e alla caratterizzazione delle emozioni attraverso la voce, all'implementazione di parlato emotivo in sistemi di sintesi vocale ecc.

Il crescente interesse per le emozioni espresse attraverso la voce si desume soprattutto dall'ampio numero di pubblicazioni (tra cui l'istituzione di giornali e riviste dedicate) e dalla nascita di apposite organizzazioni (HUMAINE, ISER, ecc.) su tale ambito di studio.

Se da un lato si rilevano innumerevoli studi sul parlato emotivo nelle singole lingue, dall'altro gli studi di tipo cross-linguistico e cross-culturale risultano essere assai sparuti (per gli studi cross-linguistici volti all'encoding si vedano Braun-Oba, 2007; Anolli et al., 2008; per quelli volti al decoding si vedano invece Kramer, 1964; Beier-Zautra, 1972; Albas et al., 1976; McCluskey-Albas, 1981; Van Bezooijen, 1983; Scherer et al., 2001; Tickle, 1999, 2000; Piôt, 1999; Abelin-Allwood, 2000; Droomey et al., 2005; Shochi et al., 2007; Sawamura et al., 2007). I motivi sono prevalentemente dovuti alla difficoltà che lo studio delle emozioni vocali impone, difficoltà ulteriormente esacerbate nello studio di tipo cross-linguistico-culturale.

I risultati degli studi appena menzionati provano come le emozioni vocali siano, alla pari di quelle facciali, riconosciute cross-linguisticamente e cross-culturalmente, con risultati nettamente al di sopra della semplice casualità. Da una meta-analisi sugli studi presenti in letteratura sulle emozioni vocali condotta da Laukka (2004) per verificare se le emozioni siano cross-culturalmente riconosciute e se vi siano specifici patterns acustici della voce per categorie discrete di emozioni, emerge infatti che: a) l'accuratezza di riconoscimento è superiore a quella della casualità per categorie di emozioni più ampie; b) il decoding cross-culturale è inferiore al decoding intra-culturale del 7%; c) esistono specifici patterns acustici nella voce delle emozioni che vengono utilizzati per comunicare emozioni discrete. Tuttavia, se da una parte Scherer-Banse-Wallbott (2001: 78) affermano "it seems reasonable to assume that the recognition of vocal emotion expressions might work across language and culture boundaries", più avanti gli stessi autori, sottolineano come "[...] encoders and decoders from several different countries would need to be studied, allowing the construction of an encoder-decoder-emotion matrix and to test whether decoders from the countries involved would recognize emotion portrayals by encoders from their own countries most accurately" (idem, pag. 88).

La presente proposta si inserisce proprio nel quadro scientifico appena delineato. Dopo una prima ricognizione di 64 database di parlato emotivo (Ververidis-Kotropoulos, 2006) alla ricerca di raccolte disponibili per le lingue in esame (ovvero italiano, francese, inglese e tedesco), e verificata l'assoluta assenza di corpora di parlato emotivo mistilingue che potessero essere utilizzati ed il cui materiale registrato potesse essere confrontato all'interno di uno studio di tipo cross-linguistico-culturale delle emozioni, si è resa necessaria la raccolta di materiale per l'attuazione della ricerca precedentemente menzionata con riferimento alle emozioni definite da Ekman (1992) come basic.

Obiettivi della presente proposta sono pertanto quelli di:

motivare e illustrare le caratteristiche del corpus raccolto, con particolare riferimento al protocollo di elicitazione adottato;

fornire i risultati della procedura di validazione, operazione assolutamente necessaria per l'attuazione delle successive fasi della ricerca volta, da un lato, all'analisi acustica dei campioni raccolti per la caratterizzazione delle emozioni nelle lingue in esame, dall'altro, alla somministrazione di esperimenti percettivi nella direzione proposta da Scherer-Banse-Wallbott (2001: 88) per appurare la capacità di soggetti di lingua diversa a decodificare espressioni emotive in una lingua differente dalla propria e per verificare quanto la conoscenza della lingua possa influire sul riconoscimento.

Con riferimento al punto 1, verrà illustrato nei dettagli il protocollo di elicitazione adottato che si compone di tre fasi distinte e conseguenti tra loro e per le quali ci si è ispirati al "paradigma degli scenari" ed al "contenuto

standard” di Anolli et al. (2008), Anolli-Ciceri (1992), Scherer et al. (1991) con la raccolta di materiale sonoro da parte sia di soggetti naif che di attori.

Con riferimento al punto 2, invece, verranno esaminati e commentati i risultati della fase di validazione del corpus verificando a) eventuali differenze di riconoscimento da parte dei soggetti coinvolti nelle registrazioni con riferimento a ciascuna delle tre fasi del protocollo di elicitazione, e b) il contributo, in termini di materiale utile e di riconoscibilità delle produzioni, da parte di soggetti naif da un lato e di attori dall’altro.

BIBLIOGRAFIA

- Abelin, Å., Allwood, J. (2000), Cross linguistic interpretation of emotional prosody, *Speech and Emotion*, pp. 110-113.
- 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.
- Anolli, L., Ciceri, R. (1992), *La voce delle emozioni. Verso una semiosi della comunicazione vocale non-verbale delle emozioni*, Milano: Angeli.
- Anolli, L., Wang, L., Mantovani, F., De Toni, A. (2008), The Voice of Emotion in Chinese and Italian Young Adults, *Journal of Cross-Cultural Psychology* 39, pp. 565-598.
- Beier, E. G., & Zautra, A. J. (1972), Identification of vocal communication of emotions across cultures, *Journal of Consulting and Clinical Psychology*, 39, 166.
- Braun, A., Oba, R., (2007), Speaking Tempo in Emotional Speech: a Cross-Cultural Study Using Dubbed Speech, *Proceedings of the International workshop on Paralinguistic Speech - between models and data, ParaLing'07*, 3 August 2007, Saarbrücken, Germany, available on-line: <http://www2.dfki.de/paraling07/papers/16.pdf>.
- 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), pp. 351-359.
- Ekman, P. (1992), An argument for basic emotions, *Cognition and Emotion*, 6, pp. 169-200.
- 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.
- 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.
- Piot, O. (1999), Experimental study of the expression of emotions and attitudes in four languages, *Proceedings of the 14th International Conference of Phonetic Sciences*, San Francisco, USA, pp. 369-370.
- 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.
- Scherer, K.R., Banse, R., Wallbott, H.G., Goldbeck, T. (1991), Vocal cues in emotion encoding and decoding, *Motivation and Emotion*, 15, pp. 123-148.
- 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.
- 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.
- Ververidis, D., Kotropoulos, C. (2006), Emotional speech recognition: Resources, features, and methods, *Speech Communication*, 48, pp. 1162-1181.

LA DIMENSIONE TEMPORALE IN TRE TIPI DI PARLATO: UN CONFRONTO TRA ARABO E ITALIANO

Dalia Gamal
Università di Ain Shams (Il Cairo)
daliagamal60@hotmail.com

SOMMARIO

Il presente lavoro ha lo scopo di confrontare l'arabo e l'italiano L2 con un *corpus* di italiano lingua prima, tramite l'analisi del ritmo in tre stili diversi: parlato letto di giornali e telegiornali, parlato meno controllate di interviste televisive e parlato semispontaneo di dialoghi *Map Task*.

In un lavoro precedente (Gamal, 2006) è stato analizzato il ritmo dell'italiano L2 in confronto all'arabo, la lingua prima degli informatori, tramite l'esame della distinzione tipologica tra le due lingue, la quale classifica l'italiano come una lingua ad isocronia sillabica e definisce l'arabo una lingua ad isocronia accentuale. Il materiale analizzato constava di turni dialogici di *Map Task*. Non è stata rilevata una chiara influenza della lingua prima sulle produzioni in italiano degli apprendenti, che hanno dimostrato di rispettare la struttura sillabica italiana; inoltre, è stata riscontrata nelle due lingue una notevole differenza nelle strategie di riduzione in funzione della variazione della velocità d'eloquio. Per esempio, in merito della riduzione delle sillabe le pretoniche hanno manifestato una maggiore riduzione in arabo, mentre sono state le postoniche e le toniche a ridursi di più in italiano; e, contrariamente alle aspettative, le sillabe collocate tra due toniche si sono mostrate le meno ridotte. All'interno della sillaba le consonanti sono state manipolate diversamente dalle vocali: la conservazione dell'estensione temporale delle sillabe nelle velocità di eloquio maggiori si realizzava in arabo grazie alle vocali, che resistevano maggiormente agli accorciamenti, e in italiano, invece, per via delle consonanti che a volte aumentavano di durata a scapito delle vocali, generalmente più ridotte tranne nelle toniche e le postoniche.

In questa sede, dunque, per un ulteriore approfondimento di questo inatteso distacco dalla L1 si ricorre al confronto delle produzioni degli apprendenti con produzioni di italiani nativi in diversi tipi di parlato. Qui la durata è sempre il parametro cardine nell'indagine. Come indici dell'organizzazione ritmica vengono considerate le durate delle vocali e delle sillabe, delle catene articolate, delle pause piene e vuote, così come vengono osservate le strategie di raggruppamento delle parole in gruppi di respiro senza controllare a priori la struttura sintattica degli enunciati.

Malgrado si sia consapevoli della portata del materiale fonico spontaneo nell'esplorazione di una data lingua, l'osservazione delle variabili che influenzano il ritmo risulta assai difficile in corpora semispontanei e spontanei, che sono per natura meno controllate. Dunque, questa indagine ha lo scopo di abbinare il controllato e non, oltre alla ricerca di ottenere una maggiore omogeneità tra il materiale italiano L1 e L2. In effetti, parte del materiale è rappresentata dalla lettura dello stesso paragrafo di un editoriale da parte di nativi e di apprendenti; lo stesso vale per il brano del TG italiano. Inoltre, il parlato semispontaneo è elicitato con le stesse mappe italiane, utilizzate nella raccolta del corpus AVIP-API. I brani tratti dal TG e dal giornale arabo vengono ulteriormente registrati con le voci degli apprendenti egiziani che sono i partner delle conversazioni in italiano L2.

Nel corpus arabo si aggiungono al parlato letto alcune stringhe di recitazioni del Corano, variate in funzione della velocità di articolazione, perché il ritmo del Corano è il modello base della ritmicità, che, anche se non si riscontra in pieno nella lettura moderna, fa sempre parte del sottofondo prosodico degli abitanti dei paesi arabi e islamici. Nella recitazione coranica le durate vocaliche seguono un'organizzazione particolare, in quanto sono stabilite in modo sistematico a seconda del contesto segmentale successivo alla vocale. Oltre alla presenza di vocali brevi e lunghe, queste ultime vengono raddoppiate prima della oclusiva glottidale, che in arabo rappresenta un fonema, e triplicate prima delle consonanti doppie. In posizione prepausale le vocali lunghe devono mantenere la stessa durata, singola, doppia o tripla, in tutte le stringhe. Tali regole di 'allungamento' ci offrono dunque tre classi temporali da aggiungere alla quarta delle brevi. Chi recita è tenuto a mantenere la regolarità ritmica conservando questa simmetria tra le varie classi vocaliche comunque sia più alta la velocità e, di conseguenza, più grande la necessità di accorciare le vocali. In arabo, va notato, è stato sempre riconosciuto il valore delle vocali e delle loro durate nella creazione della regolarità ritmica del parlato e nella distinzione tra le parlate delle varie tribù, nonché nella differenziazione tra varie situazioni comunicative, come si osserva, tra gli altri, nello studio di Ibn Jinni che risale all'undicesimo secolo d.C.

Il parlato dei mass media arabi, inoltre, si presenta all'ascolto più pacato rispetto a quello italiano; e le analisi in questo lavoro ci permetteranno di sviluppare una discussione per scoprire gli indici coinvolti nella produzione di tale effetto. Nel dialetto cairota, rispetto all'arabo standard, si rilevano variazioni delle lunghezze vocaliche e omissioni che cambiano la struttura sillabica delle parole e ciò offre un punto di confronto con la produzione degli apprendenti nei vari stili della L2.

Infatti, il passaggio nell'analisi dall'arabo recitato – del Corano e quello moderno del giornale e del TG – alla varietà parlata al Cairo ci può fornire un'immagine più chiara dello 'sfondo ritmico' dei parlanti. E attraverso il confronto tra le loro produzioni in arabo e in italiano e il materiale prodotto da italiani nativi si potrebbero spiegare meglio le strategie ritmiche degli informatori, soprattutto gli apprendenti che nella loro produzione in lingua straniera non danno finora conferma all'ipotesi del transfer.

RIFERIMENTI BIBLIOGRAFICI

- Gamal, D. (2006). Sul ritmo in italiano L2: l'eventuale isocronia, in Atti del 3o convegno Nazionale dell'AISV, pp. 101-118.
- Giannini, A., Pettorino, M. (1999). I cambiamenti dell'italiano radiofonico negli ultimi 50 anni: aspetti ritmo-prosodici e segmentali, in Atti delle 9e Giornate del G.F.S., pp. 65-81.
- Hil_1, A. H. (1996), _____ [Fonologia della lingua araba], Cairo: Wahba.
- Ibn-Jinni, A.-F. (1994). _____ [Le caratteristiche], Cairo: Organizzazione egiziana del libro, 3 voll.
- Pettorino, M., Giannini, A. (2005). analisi delle disfluenze e del ritmo di un dialogo romano, in Albano Leoni, F, Giordano, R. (a cura di), Italiano parlato. Analisi di un dialogo, Napoli: Liguori, pp. 89-104.

QUANTI PROCESSI DI ELISIONE CATEGORICA ED OPZIONALE NEL FIORENTINO PARLATO?

Luigia Garrapa
Universität Konstanz & CRIL, Università del Salento
luigiagarrapa@yahoo.it

SOMMARIO

Nell'italiano parlato le sequenze di due vocali al confine di parola possono essere risolte ricorrendo alla dittingazione o all'elisione della prima vocale, ma possono anche rimanere invariate. L'elisione vocalica cancella prevalentemente le vocali atone finali delle parole funzionali (determinanti e pronomi clitici) seguite da parole lessicali (sostantivi e verbi) che iniziano per vocale, e risulta particolarmente produttiva nella varietà di italiano parlata a Firenze (cf. Agostiniani 1989).

Sebbene l'elisione sia stata oggetto di alcuni (fra l'altro, non recentissimi) studi (cf. Agostiniani 1989, Vogel et al. 1982, Nespor 1990), questi studi presentano l'elisione come obbligatoria per i determinanti maschili singolari *un(o)/l(o)/quell(o)*, ma come opzionale e sostanzialmente imprevedibile (seppure in parte condizionata morfologicamente) per le restanti parole funzionali. Ne consegue che finora l'elisione non è stata analizzata in modo del tutto adeguato: in effetti, nessuno degli studi citati fa riferimento a dati quantitativi, né analizza in modo sistematico l'elisione nei determinanti e nei pronomi clitici dell'italiano.

Il presente studio intende chiarire il funzionamento dell'elisione nel fiorentino parlato e mettere in luce i differenti processi di cancellazione delle vocali finali delle parole funzionali presenti in questa varietà di italiano. I dati analizzati sono rappresentativi per il fiorentino contemporaneo e provengono in parte (3306 contesti di elisione) dal vasto corpus di parlato spontaneo *C-Oral-Rom* (Cresti & Moneglia 2005), ed in parte sono stati elicitati sul campo tramite un esperimento cui hanno preso parte 9 parlanti fiorentini (2394 contesti di possibile elisione). Questi dati sono stati sottoposti ad un'attenta analisi uditiva e ad analisi statistica inferenziale per determinarne la rappresentatività. Le parole funzionali studiate sono i determinanti *l(o)/la, le, un(o)/una, quell(o)/quella, questo/questa, questi, e quelle/queste*, i clitici accusativi *lo/la/li/le*, i clitici di persona *mi/ci* ed il clitico dativo *le*.

L'analisi congiunta dei dati empirici estrapolati dal corpus ed elicitati sul campo mette in luce la presenza di *cinque* processi di elisione vocalica (più o meno produttivi) nel fiorentino. Un *primo* processo elide obbligatoriamente le vocali dei determinanti maschili singolari *un(o)/l(o)/quell(o)* davanti a sostantivi che iniziano per vocale sia atona (ad es. *quell'amico*) che tonica (ad es. *quell'attimo*). Questo processo categorico sembra riconducibile ad allomorfia sia nel fiorentino contemporaneo che nell'italiano parlato nel XX secolo e probabilmente anche in precedenti stadi dell'italiano (cf. Battaglia & Pernicone 1954; Brunet 1979-1986; Regula & Jerney, 1975).

Un *secondo* processo cancella quasi obbligatoriamente (approssimativamente nel 98% dei casi) le vocali dei determinanti femminili singolari *una/la/quella* sia davanti a sostantivi che iniziano per vocale atona (ad es. *un'una idea*) che per vocale tonica (ad es. *un'una elica*). Un *terzo* processo elide opzionalmente ma abbastanza frequentemente (approssimativamente nel 70% dei casi) le vocali degli aggettivi dimostrativi *questo/questa* sia davanti a sostantivi che iniziano per vocale atona (ad es. *quest'questo invito*) che per vocale tonica (ad es. *quest'questo attimo*). Un *quarto* processo cancella facoltativamente (approssimativamente nel 35 ~ 60% dei casi) e con estrema variabilità *interspeaker* ed *intraspeaker* le vocali dei clitici accusativi singolari *lo/la* e del clitico di persona *mi*, prevalentemente quando questi sono seguiti da un verbo che inizia con una vocale atona (ad es. *l'/la invita*) e meno frequentemente se il verbo inizia con una vocale tonica (ad es. *m'/mi evita*). Questi tre processi opzionali e variamente produttivi non sono spiegabili in termini di allomorfia, ma sembrano derivare diacronicamente da un unico processo facoltativo di 'effettiva' elisione, che avveniva prevalentemente davanti a vocale atona e meno frequentemente davanti a vocale tonica nell'italiano parlato nel XX secolo (cf. Battaglia & Pernicone 1954; Brunet 1979-1986; Regula & Jerney, 1975), e che per questo può essere considerato postlessicale. Col tempo l'elisione nei determinanti femminili singolari e negli aggettivi dimostrativi *questo/questa* sembra aver subito morfologizzazione e aver acquistato produttività, diventando quasi obbligatoria per i primi e molto frequente per i secondi, e contestualmente diventando insensibile all'influsso dell'accento eventualmente presente sulla vocale iniziale del sostantivo. Al contrario, l'elisione nei pronomi singolari sembra aver subito una morfologizzazione solo parziale ed aver in parte perso produttività.

Un *quinto* processo di elisione nel fiorentino contemporaneo cancella piuttosto raramente le vocali dei determinanti plurali *questi, le/quelle/queste* e dei clitici plurali *li/le, ci*. Questo processo sembra essere non particolarmente produttivo nell'italiano parlato nel XX secolo e spesso associato ad uno stile piuttosto trascurato (cf. Battaglia & Pernicone 1954; Brunet 1979-1986; Regula & Jerney, 1975). In ultimo, il pronome dativo femminile *le* rifiuta categoricamente l'elisione sia nel fiorentino contemporaneo che nell'italiano parlato nel XX secolo (cf. Battaglia & Pernicone 1954; Brunet 1979-1986; Regula & Jerney, 1975).

I dati empirici considerati smentiscono la presunta imprevedibilità dell'elisione con tutte le parole funzionali diverse dai determinanti *un(o)/l(o)/quell(o)* e mettono in evidenza alcune tendenze interessanti. Per prima cosa, le vocali dei determinanti vengono cancellate molto più frequentemente rispetto a quelle dei pronomi, per le quali l'elisione risulta estremamente variabile. In secondo luogo, l'elisione nei determinati è (quasi) interamente morfologizzata e non risulta inibita dalla presenza dell'accento primario sulla vocale iniziale del sostantivo. Al contrario, l'elisione nei pronomi è morfologizzata solo in parte e risulta parzialmente sfavorita dalla presenza dell'accento primario di parola sulla vocale iniziale del verbo. Oltre a ciò, i dati del corpus mostrano che, se l'elisione (quasi) categorica nei determinanti singolari non risulta minimamente influenzata dallo stile discorsivo, la stessa cosa non è valida per i processi di elisione opzionali ed eccezionali: questi ultimi vengono applicati più frequentemente nel parlato informale (generalmente caratterizzato da una velocità di elocuzione maggiore e da una pronuncia meno accurata) e meno frequentemente in quello formale (generalmente caratterizzato da una velocità di elocuzione minore e da una pronuncia accurata).

I dati empirici analizzati lasciano pensare che l'elisione categorica che cancella le vocali dei determinanti maschili singolari *un(o)/l(o)/quell(o)* sia una regola precompilata nel senso di Hayes (1990) e derivi dalla lessicalizzazione delle forme elise dei determinanti in questione piuttosto che dall' 'effettiva' applicazione di un processo di elisione. Le restanti parole funzionali, invece, sembrano essere salvate nel lessico mentale esclusivamente nella forma terminante per vocale, e l'applicazione opzionale o eccezionale dell'elisione sulle stesse genera le rispettive varianti elise. Dunque l'elisione che cancella quasi obbligatoriamente le vocali dei determinanti femminili singolari *una/la/quella* sembra essere un processo postlessicale altamente produttivo che non risente dello stile discorsivo (formale vs. informale) e risulta molto simile alle *rules of external sandhi* di Kaisse (1985) ed alle *type 1 rules* di Pak (2005). I processi di elisione che cancellano opzionalmente ma non molto frequentemente le vocali dei clitici singolari *lo/la* e *mi* e piuttosto raramente le vocali dei clitici plurali *li/le* e *ci* e di tutti i determinanti plurali si presentano, invece, come processi postlessicali del tutto simili alle *fast speech rules* di Kaisse (1985) ed alle *type 2 rules* di Pak (2005). In ultimo, il processo che elide molto frequentemente, ma non obbligatoriamente, le vocali degli aggettivi dimostrativi singolari *questo/questa* sembra essere un processo postlessicale in transizione: in effetti esso sta gradualmente perdendo le caratteristiche dei processi postlessicali opzionali e sta gradualmente acquisendo le caratteristiche dei processi postlessicali altamente produttivi.

RIFERIMENTI BIBLIOGRAFICI

- Agostiniani, L. (1989): 'Fenomenologia dell'elisione nel parlato in Toscana'. *Rivista Italiana di Dialettologia* XIII: 3-46.
- Battaglia, S. & V. Pernicone (1954): *Grammatica Italiana*. Torino: Loescher.
- Brunet, J. (1979-1986): *Grammaire critique de l'Italien*. Voll 2, 4, 8. Vincennes: Université ParisVIII.
- Cresti, E. & M. Moneglia (2005): *C-Oral-Rom. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: Benjamins.
- Hayes, B. (1990): 'Precompiled Phrasal Phonology'. In Inkelas, S. & D. Zec (eds.), *The Phonology-Syntax Connection*. Chicago, CSLI: 85-108.
- Kaisse, E. (1985): *Connected Speech. The Interaction of Syntax and Phonology*. Orlando: AP.
- Nespor, M. (1990): 'Vowel Deletion in Italian: the Organization of the Phonological Component'. *The Linguistic Review* 7: 375-390.
- Pak, M. (2005): 'Explaining Branchingness Effects in Phrasal Phonology'. In Alderete, J., H. Chung Hye & A. Kochetov (eds.), *Proceedings of the 24th WECOL*. Somerville, MA, Cascadia: 308-316.
- Regula, M. & J. Jerney (1975): *Grammatica Italiana descrittiva su basi storiche e psicologiche*. Bern: Francke.
- Vogel, I., M. Drigo, A. Moser & I. Zannier (1982): 'La cancellazione di vocale in italiano'. *Studi di Grammatica Italiana* 12: 189-230.

PHONOLOGICAL CONTRASTS IN FOREIGN LANGUAGE LEARNING: A NEUROPSYCHOLOGICAL STUDY ON PALATAL AFFRICATES

Nuria Kaufmann
Neuropsychologisches Institut der Universität Zürich
nuria_kaufmann@access.uzh.ch

ABSTRACT

This study addresses the question whether a foreign language can be learned more easily by a speaker of a native language that shares similar phonological categories.

Kuhl (2004) promotes a Critical Theory Hypothesis stating that at the age of 9 months infants lose the ability to distinguish non-native phonological contrasts. Instead they specialise their neuronal networks for their native-language, which Kuhl calls Native Language Neural Commitment. This theory supposes that learning a foreign language after a certain age cannot be accomplished without a recognisable foreign accent. However, Friederici et al. (2002) challenge this hypothesis with a study using an artificial language. From the study it follows that a non-native learner even after the age of twenty can achieve native-like processing patterns in an EEG-analysis.

Furthermore, there are a number of studies showing an improved ability to discriminate foreign language phonemes with higher language proficiency (Winkler et al., 1999). Various models propose different ways how newly acquired phonological categories are mapped onto similar or distinct native categories (Lipski, 2006). Also, a connection between orthographic knowledge and the awareness of sound segments which in turn influence speech perception is strongly assumed. Learning, therefore, can improve the ability to perceive and possibly produce non-native phonological contrasts. Despite the fact that the processing of foreign speech sounds appears to recruit significant greater brain resources over a longer period of activation (Zhang et al., 2005), this pattern seems to diminish with linguistic experience. Mismatch negativity paradigms have shown that fluent non-native speakers develop cortical memory for the foreign language phonemes (Näätänen et al., 1997 and Winkler et al., 1999). Such recognition patterns presumably develop gradually with the exposure to the new language.

In this study the focus lies on two post-alveolar and two palatal affricates of Serbian:

– tʃ] (post-alveolar, voiceless), [tʃ] (alveolo-palatal, voiceless), – [dʒ] (post-alveolar, voiced), – [dʒ] (alveolo-palatal, voiced).

These affricates are going to be presented randomly, as isolated syllables, followed by the vowel /a/. In an Oddball-EEG-experiment the volunteers will wear headphones and will be asked to press a button every time they perceive a different syllable. Two groups of participants are going to be tested: 15 Swiss-German speaking adults and 15 Rhaeto-Romansh (Sursilvan speakers only) speaking adults between the ages of 20 to 30 years. Of the four affricates under investigation, Swiss-German speakers just use the post-alveolar voiceless contrast. Therefore the hypothesis is that Swiss-German speakers would not be able to reliably distinguish the four Serbian affricates; if at all, they would press the button randomly. Rhaeto-Romansh speakers on the other hand are expected to be able to distinguish all four affricate categories as they use very similar phonetic categories.

REFERENCES

- Friederici, A.D., Steinhauer, K. & Pfeifer E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Pnas*, 99(1), 529-534
- Gordon, M. et al. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, 32, 2, 141-174.
- Handbook of the International Phonetic Association (1999). A Guide to the Use of the International Phonetic Alphabet. *Cambridge University Press*
- Hayes-Harb, R. & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, 24, 1, 5-33.
- Kuhl, P. (2004). Early Language Acquisition: Cracking the Speech Code. *Nature Reviews*, 5, 831-843
- Ladefoged, P. & Maddieson, I. (1996). *The Sound of the World's Languages*. Blackwell Publishers
- Lipski, S. C. (2006). *Neural correlates of fricative contrasts across language boundaries*. Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart
- Maddieson, I. (1984). *Patterns of Sounds*. Cambridge University Press, Cambridge
- Morén, B. (2006). Consonant-vowel interactions in Serbian: Features, representations and constraint interactions. *Lingua*, 116, 1198-1244.
- Näätänen, R. et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385, 432-434.
- Recasens, D. & Espinosa, A. (2007). An electropalatographic and acoustic study of affricates and fricatives in two Catalan dialects. *Journal of the International Phonetic Association*, 37, 2, 143-172.
- Schmid, S. (to appear). Les occlusives palatales du vallader. *Actes du XXIV Congrès International de Linguistique et de Philologie Romanes*, Niemeyer.
- Tsao, F.-M. et al. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *Journal of the Acoustical Society of America*, 120, 4, 2285-2294.
- Winkler, I. et al. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, 36, 638-642
- Zhang, Y. et al. (2005). Effects of language experience: Neural commitment to language-specific auditory patterns. *NeuroImage*, 26, 703-720.

FROM SOUND TO RHYTHM EXPECTANCY

Eric Keller, professor emeritus
University of Lausanne, Switzerland
eric.keller@unil.ch

The temporal structure of speech reflects as much the massive intricacy of linguistic structures, as it demonstrates the considerable complexity of speech communication. One only needs to consider the difficulties of a native Rome speaker trying to follow the everyday chit-chat of a Sardinian farmer, or that of a Zurich native trying to understand rapid instructions given in the dialect of the upper Wallis. Comprehension in such challenging speech conditions supposes not only a good understanding of shared language components and potential dialectal variations, but also a rapid and reliable orientation within the temporal speech material. One way to do this is to orient oneself towards prominent prosodic features within the speech chain. Indeed, the decoding process appears to be guided by the *rhythmic qualities* of speech.

But how are these “rhythmic qualities” defined? In an influential 1977 review paper of the issue of isochrony (evidence for rhythmicity in speech), Ilse Lehiste argued that there were no simple acoustic parameters that stood in regular temporal relationship, and that “rhythmicity” was a perceptual construct produced on the basis of a combination and complex interaction of a great number of temporal, intonational and motor parameters. This view, supported by a number of further studies conducted since then, has to the present day formed the general consensus for spontaneous speech.

The issue has thus migrated from the purely mechanical identification of single temporal features – familiar to most prosodic researchers – to the question of the psycholinguistic construction of reference points from multidimensional material. This in turn has raised much less familiar concepts that reach into chaos-based “coordinative structures”, via more or less mysterious “oscillator systems” operating at the neuronal level, to research on evoked potentials about the anticipation of perceptual events. This presentation will explain and tie together many of these less familiar strands of temporal structuring. We will explain Bob Port’s experiments on “dig for a duck”-type sentences within a chaos framework, we shall touch on Plinio Barbosa’s two-oscillator hypothesis, and we shall explain the evoked potentials research that permits us to postulate rhythmic expectancy in the human brain. We shall also demonstrate the particular relevance of the P-centre / vowel onsets in the phonetic chain.

Researchers wishing to pursue new work in this area may wish to take either of two exciting directions. Very promising work is likely to come out of further evoked potentials work, where fairly direct correlates of rhythmic perception are likely to be hiding in EEG responses. The other promising direction lies in the mathematical prediction of rhythmic effects as divergences from “a-rhythmical” statistical predictions of speech events. For that, one needs to develop a classic multidimensional statistical model of temporal structure which permits to measure divergences between actual, rhythmic speech patterns and patterns predicted by the model. Much basic information about research in the second orientation can be found on my web site www.mindsight.cc/cv/Kellerdoc.html. Some key articles are noted below.

REFERENCES

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5, 253-263.

Beats and Bob Port's experiments: Keller, E. (2007). Beats for individual timing variation. In A. Esposito, E. Keller, M. Marinaro, M. Bratanic (eds.), *The Fundamentals of Verbal and Non-Verbal Communication and the Biometrical Issue* (pp. 115-128), IOS Press. (http://www.mindsight.cc/cv/pdf.files/Keller_07_Beats.pdf)

Beats, Bob Port and expectancy:

Keller, E., & Port, R. (2007). Speech timing: Approaches to speech rhythm. *Introduction to Special Session on Timing. Proceedings of the International Congress of Phonetic Sciences*. Saarbrücken, Germany. (http://www.mindsight.cc/cv/pdf.files/Keller_Port_07_IntroTiming_ISPhS.pdf)

Keller, E. (2007). Waves, beats and expectancy. *Special Session on Timing. Proceedings of the International Congress of Phonetic Sciences*. Saarbrücken, Germany. (http://www.mindsight.cc/cv/pdf.files/Keller_07_WavesBeatsExpectancy_ICPhS.pdf)

The statistical modeling of temporal structures: Keller, E., & Zellner Keller, B. (2003). How Much Prosody Can You Learn from Twenty Utterances? *Linguistik online*. 17, 5/03, 57-78. <http://www.linguistik-online.de/>, ISSN 1615-3014. (<http://www.mindsight.cc/cv/pdf.files/Keller-Zellner-Keller-03-ProsodyFrom20Utt.pdf>)

SONORITY BASED SYLLABLE SEGMENTATION

Bogdan Ludusan, Serena Soldo
 Department of Physical Sciences, University "Federico II" of Naples – NLP Group
ludusan@na.infn.it, soldo@na.infn.it

ABSTRACT

Automatic speech segmentation is a topic of great interest in nowadays speech related literature due to its multiple use. One of its most important application areas is Automatic Speech Recognition (ASR), in which speech segmentation techniques are applied for obtaining the units used for recognition. In the recent ASR literature, the syllable is a frequent choice for such a unit because it offers a good representation of the variability present in the speech signal while retaining a also good trainability. This is the reason behind our proposal for an algorithm for automatic syllable segmentation.

The most widely used syllable definition in phonology is based on the sonority scale. A recent definition of the sonority links it to the loudness of a sound, which is related to its acoustic energy relative to other sounds having the same length, stress, and pitch [1]. Based on this measure, a relative ranking of the sonority of sounds was done: low vowels > mid vowels > high vowels > liquids > nasals > obstruents. The Sonority Sequencing Principle (SSP) [2] is used as the principle for syllabification and it states that the sounds inside a syllable increase in sonority from the onset to the nucleus, with a maximum value corresponding to the nucleus and decrease in sonority from the nucleus to the coda.

The sonority was previously used as a feature for segmentation in speech processing, but it was either used to detect only syllable nuclei [3] or to detect syllable boundaries, but combined with other features and in conjunction with statistics from previous segmentations [4].

In [3], Kawai and van Santen use multiple linear regression in order to obtain, what they call, the instantaneous sonority. As predictor variables for the regression they use bandpass-filtered acoustic energy from the central part of each phone. The authors argue that the five frequency bands chosen can efficiently locate boundaries between different phone classes. They report accuracies of over 60% for syllable nuclei detection and over 80% for speech rate recognition for a corpus of read news.

The authors of [4] obtain their segmentation by using time-domain signal processing followed by a refinement of the results based on a fuzzy-logic approach. As time domain feature they use the zero-crossing rate in the intervals of sonority decrease, which, they state, it is related to the to the attenuation of the acoustic intensity of speech that occurs between the transition of adjacent syllables. The second step represents a refinement of these results and it is implemented using statistics from previous segmentation tests together with fuzzy logic rules. The accuracies reported on a corpus of isolated Italian digits are of 87% after the first phase and 95% after the refinement of the results.

Recent work [5] has proved the usefulness of the sonority in other areas, like rhythmic class discrimination. In this paper, the authors propose a formulation for the sonority function, defined on the interval $[0,1]$. The proposed function has values close to 1 for sounds displaying regular patterns, characteristic of sonorant portions of the signal and close to 0 for regions characterized by obstruency.

In [6] the authors refine the previously proposed function using an exponential function. Subsequently, the sonority is defined as a decreasing function of the values of the relative entropies between neighbors columns of the spectrogram of the speech signal:

$$S(t) = \exp\left(-\sum_{i=1}^3 h(p_i, p_{t-i})\right) \quad (1)$$

Based on the previous formulation of the sonority function, we propose an algorithm for the detection of syllable boundaries. The algorithm uses exclusively speech processing techniques (both frequency and time domain), having no knowledge about the phonetic content of the signal, in order to obtain the syllable boundaries from the continuous speech signal.

In a first step, for each of the utterances, their sonority function is computed as described in [6]. It implies the following stages:

- computing the spectrogram of the signal for the frequency band bellow 800 Hz, using a 25 ms window
- normalizing the power spectrum in order to be able to consider the value of each column of the spectrogram as a probability value
- computing the relative entropy of four consecutive columns; the formula for the relative entropy for the column P_t with respect to column P_s is listed in (2); the relative entropy between two values indicate the difference between those values ($h(p_i, p_s) = 0$ only if the values are equal); so, the relative entropy will give

low values for vowels, nasals and voiced stops (because of the regularity introduced by voicing) and high values for voiceless stops, fricatives and flaps [7].

$$h(p_i/p_s)^{\alpha} = \sum_i p_i \log \frac{p_i}{p_s} \quad (2)$$

applying relation (1) will give us the value of the sonority function; by multiplying in the exponential function the relative entropy with -1, we will obtain in the sonority function high values for vowels, nasals and voiced stops and low values for all the other sounds; the value of α was empirically set at 1.5

The second step consists in computing the envelope of the sonority function. This is done by filtering the sonority function with a low order low-pass filter having a very low cutoff frequency in order to obtain only the long-term variations of the signal. Because the sonority function computes the relative entropy between two normalized power spectra and because the pauses in the signal are quasi-periodic we obtain for them a higher than expected sonority. In order to solve this problem we multiply the previously obtained sonority envelope with the energy of the signal. In this way, the non-sonorous area will be deemphasized with respect to the sonorous areas, while preserving the peaks of the sonority envelope.

The syllable boundaries will be placed by taking into account the SSP. As it states that the peaks in the sonority function correspond to the syllable nuclei, by finding the maxima of the function, we will be finding the syllable nuclei. This is done first by imposing a minimum threshold for the sonority maxima and then searching for all local maxima in the signal. Having found out the syllable nuclei and knowing that the syllable boundaries correspond to the minima in the sonority function, the next step consists in finding the minima between each two consecutive maxima.

After obtaining this first segmentation, some of the errors that might appear in the segmentation process must be corrected. One of the most important errors is the existence of spurious maxima close to the syllable nuclei, due to the semi-vowels, nasals or liquids that are in the vicinity of the vowel. These types of errors can be corrected by comparing every two neighbouring maxima and eliminating one of them in case of an insertion.

The algorithm described here represents an ongoing work. While we haven't yet obtain significant results, we still have to investigate the impact of several aspects on the segmentation accuracy. Among these factors, we can mention: the frequency band for which the spectrogram is computed, the distance function used for the spectrogram columns and the possibility of combining the sonority function with different cues, as in [7].

REFERENCES

- [1] Ladefoged, Peter. 1993. A Course in Phonetics, 3rd edition (International Edition). Orlando: Harcourt Brace & Company
- [2] Clements, George N.: CV phonology: a generative theory of the syllable. Cambridge, MA, MIT Press, 1990
- [3] Goh Kawai and Jan van Santen: Automatic detection of syllabic nuclei using acoustic measures, 2002 IEEE Workshop on Speech Synthesis, Santa Monica, California
- [4] Mayora-Ibarra, O. and Curatelli, F.: Time-Domain Segmentation and Labelling of Speech with Fuzzy-Logic Post-Correction Rules. In Proceedings of the Second Mexican International Conference on Artificial intelligence: Advances in Artificial intelligence, 2002
- [5] Galves, A., Garcia, J., Duarte, D., Galves, C.: Sonority as a basis for rhythmic class discrimination, Proceedings of the Speech Prosody 2002, Aix en Provence, France
- [6] M. Cassandro, P. Collet, D Duarte, A. Galves, and J Garcia: An universal linear relation among acoustic correlates of rhythm
- [7] E. Garcia, U. B. Gut, and A. Galves.: Vocale - a semi-automatic annotation tool for prosodic research, Proceedings of the Speech Prosody 2002, Aix en Provence, France

UN CONFRONTO TRA DIVERSE METRICHE RITMICHE USANDO CORRELATORE 1.0

Paolo Mairano, Antonio Romano
Laboratorio di Fonetica Sperimentale "Arturo Genre", Università degli Studi di Torino
paolomairano@gmail.com, antonio.romano@unito.it

SOMMARIO

Basandosi sull'osservazione di Dauer (1983) e Bertinetto (1989) secondo cui le lingue cosiddette isoaccentuali (IA) e isosillabiche (IS) presentano differenti proprietà fonologiche (come la presenza/assenza di riduzione vocalica e una struttura sillabica complessa/semplice), Ramus et al. (1999) hanno proposto tre correlati ritmici ($_C$, $_V$ and $\%V$) che permetterebbero di distinguere i due gruppi di lingue. Questo nuovo approccio alla tipologia ritmica ha dato un nuovo impulso alla ricerca in questo campo cosicché, come è noto, sono stati proposti nuovi correlati: i PVI di Grabe & Low (2002), i Varco di Dellwo & Wagner (2003 e seguenti) e, recentemente, i CCI di Bertinetto & Bertini (2008). I PVI e i Varco rappresentano un tentativo di normalizzare gli effetti della velocità di eloquio sulle misure ritmiche, mentre i valori di CCI (*controlling and compensating index*) permettono di descrivere il grado di compensazione tollerata o richiesta in una lingua e si ispirano a riflessioni maturate in lavori degli anni '70 (v. Lindblom & Rapp, 1973 e Fowler, 1977; v. numerosi altri lavori in bibliografia).

Un primo obiettivo che ci siamo posti in questo lavoro è stato quello di testare il mutamento dei risultati al variare di alcuni fattori: sono stati calcolati $\%V$, $_V$, $_C$, nPVI(V), rPVI(C), Varco(V), Varco(C), CCI(V) e CCI(C) per 27 campioni de *Il vento di tramontana e il sole* (2 inglesi, 2 francesi, 2 tedeschi, 2 italiani, 2 finlandesi – segmentati autonomamente da entrambi gli autori – 10 islandesi, 3 portoghesi, 3 romeni e 1 spagnolo – segmentati da uno solo degli autori). Questa scelta rispecchia la nostra convinzione che sia opportuno studiare campioni di parlato controllato prima di affrontare il parlato spontaneo, dove molti fattori hanno (almeno potenzialmente) una forte influenza sul ritmo e sulle misure a esso associate. Riteniamo inoltre che non sia necessario utilizzare campioni estremamente lunghi poiché alcuni test di discriminazione hanno dimostrato che il cervello umano distingue lingue isosillabiche e isoaccentuali anche con campioni di pochi secondi (v. Ramus et al., 1999). Proponiamo anche di evitare il ricorso alla segmentazione automatica in quanto il guadagno in termini di tempo è controilanciato da una perdita di precisione; al contrario, abbiamo pensato di automatizzare il processo di calcolo dei correlati: a questo scopo l'autore PM ha realizzato **Correlatore**, un programma multiplatforma sviluppato in TCL/TK presto disponibile sul nascente sito del Laboratorio di Fonetica Sperimentale "Arturo Genre" di Torino. Esso calcola i valori di $\%V$, $_V$, $_C$, nPVI(V), rPVI(C), Varco(V), Varco(C), CCI(V) e CCI(C) e costruisce i grafici a partire dai file di annotazione di Praat (TextGrid) etichettati semplicemente come CV (consonante-vocale) o in SAMPA (seguendo alcune convenzioni riportate nella guida).

I dati ottenuti ci permettono di analizzare il variare dei risultati a seconda: a) dei diversi correlati utilizzati; b) dei diversi parlanti di una stessa lingua; c) di chi segmenta. Non sono stati ancora debitamente valutati gli eventuali effetti della velocità d'eloquio (argomento discusso in numerosi studi, v. per es. Dellwo & Wagner, 2003) perché, comunque, i nostri campioni sono piuttosto omogenei a questo riguardo (5-6,5 syll/s). In tutti i casi, si nota un certo grado di sovrapposizione (comunque limitato nella maggioranza dei casi) tra gruppi di lingue che graviterebbero tradizionalmente attorno ai due poli IS e IA, ma i diversi correlati hanno mostrato sensibilità a fenomeni differenti, tutti in relazione al timing. Tuttavia, è difficile stabilire quali correlati rispecchino meglio la tradizionale dicotomia di lingue IS/IA in quanto non sembra possibile avere un riscontro oggettivo del punto esatto in cui un determinato campione debba situarsi all'interno del continuum.

Discutendo della possibilità di una distinta valutazione in termini di produzione e di percezione, allo stato attuale del lavoro stiamo allestendo una batteria di test percettivi (non dissimili da quelli presentati da Ramus et al., 1999) sugli stessi campioni citati sopra, mascherati in modo che la lingua non sia riconoscibile tramite l'uso delle routine sviluppate dall'autore AR e già adottate per i test percettivi finora svolti in seno al progetto AMPER.

Un'altra direzione di ricerca che abbiamo intrapreso recentemente e di cui ci proponiamo di discutere i risultati, si basa sull'osservazione che i rapporti tra le sillabe toniche e quelle atone tendano a essere maggiormente marcati nelle lingue IA che nelle lingue IS (in cui solitamente non sono presenti fenomeni macroscopici di riduzione vocalica). Partendo da queste premesse abbiamo iniziato ad applicare il calcolo dei delta e dei PVI alle misure di intensità (in dB) e altezza (in quarti di tono) delle vocali di alcuni dei nostri campioni, iniziando a ragionare in termini di valori efficaci di queste variabili così come pare necessario a partire dall'osservazione dei risultati di test informali condotti su segnali sintetici in cui si neutralizzano alcune di esse a rotazione. Parallelamente, badando a non confondere i modelli ritmici di una lingua con le tendenze idiosincratice forse presenti nei materiali a nostra disposizione, stiamo allargando il campione di dati su cui

condurre i test. Alcuni dei risultati ottenuti in questa direzione, attualmente ancora in corso di interpretazione, potrebbero essere già disponibili in quest'occasione.

RIFERIMENTI BIBLIOGRAFICI

- Bertinetto, P.M. (1989). "Reflections on the dichotomy 'stress' vs. 'syllable-timing'". *Revue de Phonétique Appliquée*, Mons, 99-130.
- Bertinetto, P.M. & Bertini, C. (in c. di p.). "Prospezioni sulla struttura ritmica dell'italiano basate sul corpus semispontaneo AVIP/API". *Atti delle 4e Giornate AISV (Associazione Italiana di Scienze della Voce, Cosenza, 2007)*, in c. di p.
- Bertinetto, P.M. & Bertini, C. (2008). "On modeling the rhythm of natural languages". *Proc. of the 4th International Conference on Speech Prosody (Campinas, Brazil, 2008)*, 427-430.
- Dauer, R.M. (1983). "Stress-timing and syllable-timing reanalyzed". *Journal of Phonetics*, 11, 51-62.
- Dellwo, V. & Wagner, P. (2003). "Relations between language rhythm and speech rate". *Proc. of the 15th International Congress of Phonetics Sciences (Barcelona, Spain, 2003)*, 471-474.
- Fowler, C. (1977). "Timing control in speech production", Bloomington, Indiana University Linguistic Club.
- Grabe, E. & Low, E.L. (2002). "Durational variability in speech and the rhythm class hypothesis". In: Gussenhoven, C., Warner, N. (eds), *Papers in Laboratory Phonology 7*, Berlin: Mouton de Gruyter, 515-546.
- Lindblom, Bj. & Rapp K. (1973). "Some temporal regularities of spoken Swedish". *Papers in Linguistics from the University of Stockholm*, 21: 1-59.
- Mairano, P. & Romano, A. (2007). "Inter-Subject Agreement in Rhythm Evaluation for Four Languages (English, French, German, Italian)". *Proc. of the 16th International Congress of Phonetic Sciences (Saarbrücken, Germany, 2007)*, 1149-1152.
- Mairano, P. & Romano, A. (2008). A comparison of four rhythm metrics for six languages. Poster presented at the conf. on Empirical Approaches to Speech Rhythm (Univ. College London, 2008).
- Marotta, G. (1985). *Modelli e misure ritmiche: la durata vocalica in italiano*. Bologna: Zanichelli.
- Mendicino, A. & Romito, L. (1991). "«Isocronia» e «base di articolazione»: uno studio su alcune varietà meridionali". *Quaderni del Dip. di Ling. dell'Univ. della Calabria*, S. L. 3, 49-67.
- Pike, K.L. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press.
- Ramus, F., Nespors, M. & Mehler, J. (1999). "Correlates of linguistic rhythm in the speech signal". *Cognition*, 73/3, 265-292.
- Romito, L. & Trumper, J. (1993). "Problemi teorici e sperimentali posti dall'isocronia". *Quaderni del Dip. di Ling. dell'Univ. della Calabria*, S. L. 4, 10, 89-118.
- Vayra, M., Avesani, C. & Fowler, C. (1984). "Patterns of temporal compression in spoken Italian". *Proceedings of the X International Congress of Phonetic Sciences (Utrecht, The Netherlands, 1983)*, vol. 2, 540-546.

PERSISTENZA DELL'ACCENTO STRANIERO. UNO STUDIO PERCETTIVO SULL'ITALIANO L2

Giovanna Marotta
Università di Pisa
gmarotta@ling.unipi.it

SOMMARIO

La ricerca fonologica sull'acquisizione di L2 si è finora concentrata sul versante della produzione, trascurando quello della percezione, nonostante sia da tempo nota la rilevanza dei processi percettivi anche nella resa fonetica dei segmenti (cfr. Best 1995, Flege 1997; 2003, Major 2001). In particolare, risulta ancora poco indagata la tematica relativa alla percezione del cosiddetto 'accento straniero'.

All'interno di questa area di ricerca, un problema specifico concerne il peso dei tratti fonetici sulla percezione del *foreign accent*. Le domande che ci poniamo sono le seguenti:

se i tratti responsabili di "forestierismo" permangono anche nella produzione di parlanti con ottima competenza di L2, in che misura questi tratti dipendono dalle caratteristiche di L1?
qual è il ruolo dei fattori prosodici nella percezione dell'accento straniero?

Per cercare di rispondere a queste domande, abbiamo in programma una serie di test percettivi, diversamente ideati e realizzati, che dovrebbero consentirci di valutare il peso relativo degli elementi segmentali e prosodici nel riconoscimento dell'accento straniero (cfr. Marotta 2008).

In questa sede presenteremo i risultati di un test percettivo appena concluso in cui alcuni frammenti di parlato italiano prodotto da parlanti con diversa L1 (francese, spagnolo, tedesco, inglese) e ottima competenza dell'italiano sono stati valutati da parlanti nativi italiani.

I soggetti sono stati chiamati ad ascoltare gli stimoli acustici naturali, uno per volta, e a giudicare se il parlante era italiano o straniero; se valutato come straniero, i soggetti dovevano indicare la lingua madre del parlante tra le quattro lingue sopra elencate, valutando anche il grado di accento straniero su una scala a tre gradini, che va da 0 (poco accento) a 2 (accento forte). Ogni ascoltatore italiano è stato preliminarmente invitato ad auto-valutare sia la sua competenza nelle quattro lingue straniere, che il suo grado di familiarità con l'accento delle stesse lingue. Ottanta studenti hanno partecipato all'esperimento; più della metà erano di origine toscana, i restanti provenivano da varie regioni italiane.

I risultati mostrano che nella maggioranza dei casi gli ascoltatori sono in grado di percepire la differenza tra parlanti nativi e parlanti non nativi, nonostante i parlanti stranieri avessero un'ottima competenza dell'italiano. Più complesso si è invece rilevato il compito relativo all'identificazione della lingua materna dei parlanti. Soltanto gli stimoli prodotti da parlanti inglesi sono stati identificati con una percentuale di riconoscimento soddisfacente, mentre quelli relativi a parlanti spagnoli presentano i valori di corretto riconoscimento più bassi. Inoltre, gli stimoli prodotti da parlanti tedeschi sono stati spesso confusi con quelli relativi ai parlanti inglesi.

Il grado di successo nel riconoscimento della L1 è dunque inversamente proporzionale alla vicinanza strutturale e fonologica tra L1 e L2: italiano e spagnolo sono discriminati con difficoltà, mentre il parlato dei tedeschi tende ad essere confuso con quello degli inglesi più che con quello degli spagnoli.

Tuttavia, dai nostri dati non risulta una generale corrispondenza tra l'auto-valutazione dell'ascoltatore e la sua *performance* nel test percettivo. In maniera abbastanza prevedibile, soltanto nel caso dell'inglese si osservano valori comparabili tra auto-valutazione e percezione, mentre per le altre lingue straniere si rileva una discrasia più o meno marcata tra il supposto livello di familiarità con un accento straniero e la corretta identificazione della lingua straniera nel test sperimentale. In altri termini, la percezione dell'accento straniero può esser indipendente dal corretto riconoscimento della lingua parlata da colui che è stato identificato come straniero.

RIFERIMENTI BIBLIOGRAFICI

- Best, C.T. (1995), A direct realistic view of cross-language speech perception, in W. Strange (ed.), *Speech perception and linguistic experience*, Baltimore (MD), York Press, 171-206.
- Flege, J.E. (1997), The Role of Phonetic Category Formation in Second-Language Speech Learning, in *New Sounds 97, Proceedings of the Third International Symposium on the Acquisition of Second Language Speech*, University of Klagenfurt, 8-11 September 1997, 79-88.
- (2003), Assessing constraints on second-language segmental production and perception, in A. Meyer e N. Schiller (eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, Berlin, Mouton de Gruyter, 319-355.
- Major, R.C. (2001), *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*, Mahwah-London, Erlbaum Ass.
- Marotta, G. (2008), Sulla percezione dell'accento straniero, in R. Lazzeroni et alii (a cura di), *Diachronica et sincronica. Studi in onore di Anna Giacalone Ramat*, Pisa, ETS.

GRADATUM DIATOPICO, CONTINUUM DIAFASICO E DINAMICHE DIAGENERAZIONALI: PRIMI RISULTATI DI UN'INDAGINE IN TERRA DI LAVORO

Edoardo Mastantuoni

emastantuoni@hotmail.com

SOMMARIO

La presente comunicazione vuol presentare alcune tra le principali linee-guida di un'inchiesta dialettologica in corso nell'alto casertano orientale, dando conto dei seguenti aspetti:

1. La scelta dei punti (6) e la loro particolare disposizione lungo un asse sud-nord, che dal capoluogo di provincia sale fino ai monti del Matese. Il materiale linguistico raccolto nei sei punti – caratterizzato dal continuum diafasico che unisce le varietà locali d'italiano (il substandard) ai dialetti – consente un'interessante analisi in diatopia dei dati elicitati nel territorio d'inchiesta e un'analisi contrastiva a paragone con le koinai napoletane.

2. L'inchiesta in atto, pur concorrendo a colmare in parte la mancanza di documentazione sui dialetti di Terra di Lavoro, si prefigge tuttavia come scopo quello di descrivere, con una “fotografia” delle dinamiche linguistiche in atto, gli usi linguistici di due gruppi sociali: gli ultrasessantenni (che nelle interviste hanno usato prevalentemente il dialetto locale conservativo o forme arcaiche di substandard) e la generazione sotto i quaranta (che usa largamente il substandard o, in alternativa, il dialetto con attenuazione di alcuni tratti locali e largo uso di code-switching e code-mixing). I primi risultati fanno pensare a una diacronia apparente che, più che il sistema della varietà dialettale in se stessa, coinvolge le sue relazioni con quella sovrapposta nel processo di costruzione del testo conversazionale.

3. Il piano descrittivo prescelto privilegia il livello fonetico (vocalismo tonico, atono, consonantismo e fonosintassi), senza tralasciare gli aspetti morfologici e sintattici più rilevanti. L'analisi del livello fonetico viene svolta con metodologia uditiva e descrittiva, con affondi di analisi anche strumentale su singoli aspetti di particolare complessità.

4. Una peculiarità del campione consiste nel fatto che esso è composto quasi esclusivamente da donne, che nell'inchiesta dialettologica classica venivano invece messe in secondo piano rispetto all'informatore di sesso maschile o del tutto trascurate. Anche se oggi è dimostrato che spesso la variazione diasessuale ha un'influenza minima rispetto a quella diagenetale e diastratica (oltre che alle reti sociali che ciascun parlante ha stabilito), la scelta di un campione femminile si è rivelata preferibile per motivi di natura pratica: sia perché le parlanti di sesso femminile si sono dimostrate più facili da avvicinare, soprattutto nel caso delle anziane all'uscita dalla Messa, sia perché la loquacità che le caratterizza, almeno nell'area osservata, ha giocato a favore dell'inchiesta.

5. Per poter essere facilmente accettato nelle comunità, e nel tentativo di stabilire un rapporto con i soggetti da intervistare quanto più possibile cordiale, rilassato e amichevole, mi sono avvalso della collaborazione di alcuni insider. Grazie a questi ultimi, ho scelto cinque informatrici e un informatore per ciascuna delle località, individuati tra i loro parenti, amici e conoscenti.

6. Ciò che differenzia i parlanti sotto i quaranta da quelli sopra i sessanta è che i primi sono generalmente più scolarizzati dei secondi, come è ovvio, il che, nel panorama italiano, implica generalmente un maggior grado d'italianizzazione. Nel contesto delle interviste, la fascia anziana, dialettofona e monolingue, si esprime sempre in dialetto, con piccole inserzioni d'italiano substandard, laddove la fascia più giovane degli intervistati presenta una situazione linguisticamente più composita. Sia i parlanti giovani con il grado d'istruzione più alto (i laureati), sia la maggioranza di quelli con un livello d'istruzione più basso (diploma o licenza media), hanno risposto alle domande e conversato informalmente con l'intervistatore usando l'italiano e relegando alle occasionali porzioni di dialetto una funzione meramente espressiva, enfatica o metaforica. Il dialetto compare poi nelle citazioni e nei discorsi riportati.

7. Parte del gruppo dei giovani è inoltre caratterizzato da diglossia, ovvero da un uso semi-esclusivo del dialetto in famiglia o tra amici, utilizzando, in presenza sia dei parenti sia dell'intervistatore, l'italiano con quest'ultimo e il dialetto con i primi. Si aggiunga che come intervistatore ho cercato di assumere una posizione diafasicamente “neutra”, avvalendomi di un italiano molto regionale misto a dialetto, alternandolo con un italiano più vicino allo standard nei discorsi di natura più tecnica. Ciononostante, il comportamento linguistico di alcuni soggetti giovani si è orientato ugualmente verso il polo dell'italiano, dimostrando così un condizionamento diglottico evidentemente più forte di quello adattivo o imitativo.

8. In termini diatopici, i giovani, nell'uso dell'italiano regionale, dimostrano una convergenza di tratti, dimodoché anche i parlanti di Baia e Latina (di provenienza più “periferica” rispetto a San Nicola la Strada e Castel Morrone) non presentano vistose caratteristiche fonetiche locali, ma piuttosto di aver competenza di un

substandard di tipo campano cui è difficile attribuire una più precisa identità locale (a differenza di quanto accade per i parlanti di altre aree di Terra di Lavoro – come l'Agro Aversano – dove molti tratti basilettali emergono nel substandard come indicatori). Tra questi tratti convergenti possiamo ricordare a titolo di esempio l'uso della fricativa /s/ al posto dell'affricata /ts/, che invece non si osserva a nessun livello diafasico nel parlato degli anziani.

9. Lo stesso discorso non vale, naturalmente, per i dialetti dei rispettivi paesi, che, soprattutto per Baia e Latina, si differenziano molto da quelli di San Nicola e di Castel Morrone. Infatti i punti più settentrionali presentano fenomeni analoghi a quelli di dialetti molisani e del Basso Lazio, sia sul piano segmentale sia su quello prosodico.

RIFERIMENTI BIBLIOGRAFICI

- Albano Leoni, Federico / Sornicola, Rosanna / Stenta Krosbakken, Eleonora / Stromboli, Carolina [2001] (a cura di), *Dati empirici e teorie linguistiche* (Atti del XXXIII Congresso Internazionale di Studi della Società di Linguistica Italiana, Napoli, 28-30 ottobre 1999), Roma, Bulzoni.
- Auer, Peter / Hinskens, Frans [1996], *The convergence and divergence of dialects in Europe: New and not so new developments in an old area* in Ulrich Ammon / Peter Hans Nelde / Klaus J. Mattheier (a cura di), *Sociolinguistica*, no. X, Tübingen, Max Niemeyer, pagg. 1-30.
- Berruto, Gaetano [1998], *Sociolinguistica dell'italiano contemporaneo*, Roma, Carocci.
- [2003], *Fondamenti di Sociolinguistica*, Bari, Laterza.
- Chambers, J.K. / Trudgill, Peter [1999], *Dialectology*, Cambridge, CUP.
- D'Agostino, Mari [2002] (a cura di), *Percezione dello spazio, spazio della percezione. La variazione linguistica fra nuovi e vecchi strumenti di analisi* (ALS, vol. X; Atti del Convegno, marzo 2001), Palermo, Centro Studi Filologici e Linguistici Siciliani.
- De Blasi, Nicola [2006], *Profilo linguistico della Campania*, Bari, Laterza.
- Maturi, Pietro [1999], *Aspetti di fonosintassi nei dialetti campani settentrionali* in *Contributi di Filologia dell'Italia Mediana*, no. XIII, Perugia, pagg. 227-258.
- [2002], *Dialetti e substandardizzazione nel Sannio beneventano*, Francoforte, Peter Lang.
- Milroy, Lesley [1989], *Observing and Analysing Natural Language: A Critical Account of Sociolinguistic Method*, Oxford, Basil Blackwell.
- Radtke, Edgar [1997], *I dialetti della Campania*, Roma, Il Cálamo.
- Rohlf, Gerhard [1966-1969], *Grammatica storica della lingua italiana e dei suoi dialetti*, voll. tre: *Fonetica, Morfologia, Sintassi e formazione delle parole*, Torino, Einaudi.

THE NEURAL MECHANISMS OF TEMPORAL ACOUSTIC INFORMATION PROCESSING DURING SPEECH PERCEPTION

Martin Meyer

Department of Neuropsychology, University of Zurich
Binzmühlestrasse 14, Box 25
CH-8050 Switzerland
e-mail: meyer@access.uzh.ch

SUMMARY

The mounting evidence on the role temporal information processing plays during speech perception has considerably changed the understanding of present neuroscientists of the relationship between language mechanisms and the human brain. While the traditional neurological model of language proposes the existence of two “speech centers” in the left hemisphere that accommodate linguistic functions, namely syntax, semantics etc., recent frameworks favor a view that emphasizes temporal acoustic patterns as the most important foundation of speech. In terms of brain research the recent views differ from the traditional model in that they elucidate the significant involvement of the right perisylvian in speech functions. Essentially, differential preferences of the left and right auditory-related cortex for rapidly and slowly changing acoustic cues that constitute (sub)segmental and suprasegmental parameters, e.g. phonemes, prosody, and rhythm form the basis for a better understanding of the brain basis of language.

This presentation introduces a series of neuroimaging studies that explored the neural underpinnings of early stages of speech perception. In agreement with the “AST”-hypothesis (Poeppel, 2001, 2003) the presented findings supply evidence for a “division of labor” between the left and right perisylvian regions that accommodate the early stages of speech perception. In more detail, functional asymmetries related to speech perception may be accounted for by different hemispheric preferences for temporal resolution: the left auditory areas preferentially extract information over short temporal integration windows (~ 40 Hz, gamma band) and the right auditory areas over long integration windows (~ 4 -10 Hz, theta and alpha bands) (Poeppel, 2001, 2003). In other words, temporal integration windows of different length should be considered the computational mechanism responsible for decoding the inflowing stream of auditory signals. Subsequent to the initial acoustic computation speech perception becomes asymmetric in that the left posterior auditory-related cortex preferentially computes (sub-)segmental information (i.e. formant transitions, rapid frequency modulated (FM) sweeps). Complementarily, the contralateral hub is more proficient at processing slowly changing, suprasegmental auditory information, namely aspects of prosody (speech melody and speech rhythm) but also features of music (instrumental timbre, melody).

Our group performed a series of studies using the functional magnetic resonance technique and electrical brain imaging methods to examine the framework’s predictions. These confluent observations demonstrate a functional asymmetry in all posterior perisylvian regions including the auditory core areas.

Due to these accounts that tested temporal gap and voice-onset-time discrimination the left posterior perisylvian region is capable of extracting rapidly changing acoustic pattern from speech and non-speech stimuli (Jäncke et al. 2002, Meyer et al. 2005, Zaehle et al. 2004, Zaehle et al. 2007, Zaehle et al. 2008). Complementarily, we concluded from our results that the right posterior perisylvian cortex is most suited for computing slowly modulating acoustic patterns, namely speech rhythm and speech melody (Geiser et al. 2008, Meyer et al. 2002, Meyer et al. 2004)

Anatomical data on structural asymmetry of cortical volume in these brain regions and density of cortico-cortical fibre connections in the left perisylvian territory corroborate the functional observations (Glasser & Rilling 2008) and should be considered an indispensable support for the new frameworks that stipulate a novel understanding of the relationship between speech and the brain.

Eventually, this presentation underlines the importance of temporal acoustic cues that decode rapid and slow modulations in speech and emphasizes to what extent these modulations could be considered the essential structural elements that help the listener group words and phrases to perform a more efficient integration of syntactic and semantic information and to achieve a proper representation of a spoken utterances.

Finally, in concordance with Poeppel and Embick (2005), I aver the strong need for an extensive revision of the classical 19th and 20th neurological models of language processing that are conceptually ill-defined and anatomically and underspecified (Meyer 2008).

REFERENCES

- Geiser, E., Zaehle, T., Jancke, L., & Meyer, M. (2008). The neural correlate of speech rhythm as evidenced by metrical speech processing: an fMRI study. *Journal of Cognitive Neuroscience*, 20, 541– 552.
- Glasser, M. F. & Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cereb. Cort.*, 18, 2471-2482.
- Jäncke, L., Wustenberg, T., Scheich, H., & Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *Neuroimage*, 15, 733– 746.
- Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science*, 25, 679 – 693.
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech and Communication*, 41, 245– 255.
- Poeppel, D. & Embick, D. (2005). Defining the relation between linguistics and neuroscience. In A. Cutler (Ed.), *Twenty-first century psycholinguistics. Four cornerstones* (pp. 103– 118). Mahwah (NJ): Lawrence Erlbaum.
- Meyer, M. (2008). Functions of the left and right posterior temporal lobes during segmental and suprasegmental speech perception. *Z Neuropsych* 19, 101-115.
- Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum Brain Mapp*, 17, 73- 88.
- Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., & von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain Lang*, 89, 277– 289.
- Meyer, M., Zaehle, T., Gountouna, V. E., Barron, A., Jancke, L., & Turk, A. (2005). Spectro-temporal processing during speech perception involves left posterior auditory cortex. *Neuroreport*, 16, 1985– 1989.
- Zaehle, T., Geiser, E., Alter, K., Jancke, L., & Meyer, M. (2008). Segmental processing in the human auditory dorsal stream. *Brain Res.*, 1220, 179-190.
- Zaehle, T., Jancke, L., & Meyer, M. (2007). Electrical brain imaging evidences left auditory cortex involvement in speech and non-speech discrimination based on temporal features. *Behav Brain Funct*, 3, 63.
- Zaehle, T., Wustenberg, T., Meyer, M., & Jancke, L. (2004). Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. *Eur J Neurosci*, 20, 2447– 2456.

FENOMENI D'ARMONIA VOCALICA IN AREA FRIULANA E IBERICA

Renzo Miotti
Università di Verona
renzo.miotti@univr.it

SOMMARIO

L'armonia vocalica è un fenomeno assimilatorio che consiste nell'estensione di tutti o d'alcuni tratti d'una vocale ad altri segmenti vocalici, normalmente adiacenti. L'armonia può manifestarsi in una duplice direzione: da una posizione forte, cioè prominente dal punto di vista percettivo, verso posizioni più deboli; viceversa, da posizioni poco prominenti verso gli altri segmenti vocalici. In letteratura, le cause dell'armonia vengono riportate al conseguimento di benefici d'ordine strutturale: semplificazione articolatoria (Pulleyblank, 2002), benefici percettivi (soprattutto se da posizioni deboli; cfr. Walker, 2005, 2006), semplificazione articolatoria + benefici percettivi (Cole & Kisseberth, 1994). Per una sintesi della questione, cfr. Jiménez & Lloret (c.d.s).

Sulla scorta di queste premesse, il presente lavoro intende presentare e discutere i risultati (ancora provvisori, in vista d'essere corroborati da una base più ampia di dati) d'un'indagine condotta sul friulano centrale, che provverebbero l'esistenza di processi riconducibili al primo dei due modelli d'assimilazione (da posizioni forti a posizioni deboli): il tratto [\pm RLA] (ma non il punto d'articolazione) della vocale accentata verrebbe esteso alle vocali non-accentate finali, in particolare alle medie /e, o/ (va detto che le uniche vocali ammesse in posizione finale sono /e, i, o/). La nostra indagine focalizzerà il comportamento delle medie. Va detto che il fenomeno cui si fa riferimento, rilevato dall'autore con metodo uditivo e acustico, non ha sinora trovato riscontro in letteratura, la quale si limita a rilevare la generale tendenza all'abbassamento ([ϵ , ø]) delle medie non-accentate in posizione finale (Canepari, 2006³; Miotti, 2002), senza però prestar attenzione alle significative differenze riscontrabili, con regolarità, in dipendenza dai condizionamenti visti.

Diacronicamente, la media anteriore rappresenta il normale esito di -A latina nei dialetti centrali (più innovativi; le varietà più marginali conservano /-a/), avvenuto per innalzamento. La media posteriore, in posizione finale, compare, invece, in pochissimi esempi, perlopiù d'origine veneta; l'esito normale per le vocali finali diverse da -A è infatti l'apocope (sulle vicende del vocalismo non-accentato finale nella Romania nordoccidentale e in particolare in seno all'italo-romanzo settentrionale, si rimanda a Loporcaro, 2005-2006).

Avremo dunque (esemplifichiamo con /-e/): [ϵ] dopo /'e, 'o/ ma [ϵ] dopo /'e, 'o/ ['fred ϵ , 'vor ϵ] 'fredda, lavoro' ma ['s ϵ d ϵ , 'm ø r ϵ] 'seta, mora'. Si tenga presente che pure /'e, 'o/ possono arrivare a [ϵ , 'o] (seppure in minor grado rispetto alle non-accentate finali).

In friulano centrale, il fenomeno sembrerebbe condizionato da fattori contestuali: solo le vocali finali in tonia (cioè alla fine d'enunciato intonativo) sono interessate dal processo descritto, laddove all'interno dell'enunciato stesso l'effetto tende a ridursi fino a scomparire (con una relativa tendenza alla riduzione/centralizzazione dei timbri vocalici).

I fenomeni friulani vengono poi confrontati con quelli descritti per altre varietà romanze, in particolare iberiche (varietà valenzane meridionali: Jiménez, 1998; Jiménez & Lloret, c.d.s.).

In valenzano, il processo è limitato a /-a/ preceduta da / ϵ , ø /, che agiscono sulla vocale finale innalzandola e propagando il tratto "punto d'articolazione" (coronale e labiale, rispettivamente): [is'terj ϵ], 'isteria', [is't ø rj ø] 'storia'.

Da notare che, in friulano, i processi canonici d'armonia vocalica, così come descritti sopra, sembrano coesistere con una tendenza di segno opposto, di dissimilazione timbrica, consistente nell'apertura (abbassamento), anziché nella chiusura (innalzamento), della vocale finale (almeno /e/), quando il nucleo della sillaba accentata è costituito da una vocale alta: ['frite, 'dute] anziché *['frite, 'dute] 'fritta, tutta'. Peraltro, fenomeni di dissimilazione timbrica a distanza, innescati dalla presenza d'una vocale alta, non sono sconosciuti in area italo-romanza. Anche nell'italiano neutro (descritto da Canepari, 1999²: 59) si ha /'t ø p/ ['t ø :p ø] ma /'t ip / ['ti:p ø], ma in veneziano e in trevigiano (cittadini, sia in dialetto che in italiano regionale), e in veneto-giuliano, si può arrivare addirittura a ['ti:p ø] (*ibid*: 398, 406). La dissimilazione riscontrata per il friulano potrà dunque esser messa in relazione con tendenze italo-romanze (o almeno con quelle documentate per i contigui dialetti veneti).

Il confronto friulano-valenzano mette in evidenza alcuni punti problematici d'ordine interpretativo (per quanto riguarda l'interpretazione articolatoria delle cause dei processi esaminati: la semplificazione articolatoria si manifesta all'interno di domini omogenei, con estensione dei tratti a vocali contigue, solo in valenzano, mentre ciò non si verifica necessariamente per il friulano, che conosce minori restrizioni in tal senso). Mentre in valenzano l'armonia interessa solo vocali contigue – nei proparossitoni, la vocale postaccentata interna blocca il

processo –, in friulano, al contrario, ciò non avviene: cfr. valenzano [ˈtɛtrika] (e non *[-kɛ]) ‘tetra’ vs friulano [ˈlɛtɛrɛ] ‘lettera’. Inoltre, se in valenzano il dominio dell’armonia è la parola prosodica (le vocali dei pronomi enclitici non ne vengono colpite), in friulano queste ultime possono partecipare al processo: valenzano [ˈpɛlɛla] (e non *[-lɛ]) vs friulano [ˈpɛlilɛ] ‘pelala’.

Protocollo sperimentale

Le parole bersaglio sono state inserite in posizione finale (tonia) e interna di frase cornice; sono state richieste almeno cinque ripetizioni di ciascuna struttura. I valori formantici (F1 e F2, in Hz) delle vocali finali (/e, o/) sono stati misurati, mediante Praat, in corrispondenza della porzione centrale del segmento vocalico, in una varietà di contesti consonantici e proiettati poi su piani cartesiani di tipo tradizionale. Nel corpus sono presenti le seguenti consonanti prevocaliche: /p, b, t, d, v, z, r,ʄ/. Gli informanti sono due soggetti di sesso femminile, entrambe parlanti native di friulano (varietà della “fascia sudorientale del basso Tagliamento” [Frau, 1984; Miotti, 2007] ai margini occidentali del friulano centrale, ma con caratteristiche, per quanto riguarda gli aspetti analizzati, sostanzialmente affini a quelle delle varietà propriamente centrali, come abbiamo potuto appurare mediante indagini preliminari).

RIFERIMENTI BIBLIOGRAFICI

- Canepari, L. (1999²) *Manuale di Pronuncia Italiana*. Bologna: Zanichelli.
- Canepari, L. (2006³) *Manuale di pronuncia*. München: Lincom.
- Cole, J. & Kisseberth, C. (1994) «An Optimal Domains Theory of Vowel Harmony», *Studies in the Linguistic Sciences* 34: 101-114 (<http://roa.rutgers.edu/>).
- Frau, G. (1984) *Friuli*. Pisa: Pacini.
- Jiménez, J. (1998) «Valencian vowel harmony» *Rivista di Linguistica*, 10: 137-161.
- Jiménez, J. & Lloret, M.-R. (c.d.s.) «Entre la articulación y la percepción: armonías vocálicas en la península Ibérica», in *Actes du XXV CILPR*, Innsbruck, 3-8 settembre 2007.
- Loporcaro, M. (2005-2006) “I dialetti dell’Appennino tosco-emiliano e il destino delle atone finali nel(l’italo-)romanzo settentrionale”, *L’Italia Dialettale*, 66-67: 69-122.
- Miotti, R. (2002) “Friulian”, *Journal of the International Phonetic Association*, 32:237-247.
- Miotti, R. (2007) “Le varietà di Dignano, Flaibano e Sedegliano nel contesto dei dialetti friulani. Aspetti fonologici” in Vicario, F. (a c. di) *Ladine loqui. IV Colloquium Retoromanistich*. San Denêl ai 26 e 27 di avost dal 2005. Udine: Società Filologica Friulana, 71-117.
- Pulleyblank, D. (2002) «Harmony Drivers: No Disagreement Allowed». In: Larson, J. & Paster, M. (a c. di), *Proceedings of the Twenty-eighth Annual Meeting of the Berkeley Linguistics Society*. Berkeley, California: Berkeley Linguistics Society, 249-267.
- Walker, R. (2005) «Weak Triggers in Vowel Harmony», *Natural Language and Linguistic Theory* 23: 917-989 (<http://roa.rutgers.edu/>).
- Walker, R. (2006) «Long-distance Metaphony: A Generalized Licensing Proposal», *Handout, Phonology Fest Workshop*, Indiana University, Bloomington, June 23, 2006. (<http://www-rcf.usc.edu/~rwalker/pubs.html>).

DURATA E STRUTTURE FORMANTICHE NEL PARLATO TOSCANO: UNA INDAGINE PRELIMINARE SU CAMPIONI DI PARLATO SEMISPONTANEO

Nadia Nocchi, Silvia Calamai
Università di Zurigo, Università di Siena
nadia_nocchi@yahoo.com, calamai@unisi.it

SOMMARIO

Negli ultimi anni molto è stato scritto sul vocalismo tonico di area occidentale, soprattutto per quanto riguarda stili di eloquio controllati. Tra le sue caratteristiche più salienti vi sono l'abbassamento delle vocali medio-basse /ε ɔ/ e la velarizzazione di /a/, tratti percepiti dagli stessi parlanti come *shibollet* della propria varietà (Calamai 2004a). Lo studio di natura percettiva di Calamai e Ricci (2005) ha confermato che la maggiore durata della vocale tonica manipolata è quasi sempre indice di 'livornesità' e che, nel parlato semispontaneo, /ε/ è vocale più 'bandiera' di /ɔ/. È bene precisare che gli aspetti acustici coinvolti non si limitano all'abbassamento della prima formante, ma interessano anche il dominio prosodico, poiché nel parlato meno controllato sono state riscontrate una maggiore durata per le vocali colpite da abbassamento e una maggiore modulazione della frequenza fondamentale (cfr. Marotta, Calamai & Sardelli, 2004). Del resto, già Luciano Giannelli aveva rimarcato come in 'pisano-livornese' si faccia "impiego stilistico della lunghezza vocalica" (Giannelli 2000: 66). Rispetto all'area occidentale, il vocalismo fiorentino non presenta particolari tipicità (Calamai 2004b), mentre risulta tutto da indagare il vocalismo di area aretina, dove anche a livello impressionistico si registrano differenze sostanziali rispetto al toscano centrale (Giannelli 2000; Nocentini 1989).

Il presente contributo si propone di osservare il rapporto tra durata e strutture formantiche in alcuni campioni di parlato semispontaneo (dati da *map task*) di soggetti giovani provenienti dalle città di Pisa, Firenze, Livorno, Arezzo. L'indagine si suddivide in due filoni principali: da un lato, l'ispezione del legame che intercorre tra durate e sottospecificazione acustica, dall'altro, l'analisi dell'evoluzione temporale dei movimenti formantici tipici di certe aree della regione e ancora mai indagati sperimentalmente.

Per quanto concerne il primo punto – ovvero l'osservazione del rapporto tra i valori di durata e i valori spettrali – le entrate vocaliche sono state suddivise, per una prima ricognizione, in base a fattori accentuali (tonica di enunciato, tonica lessicale, atona), posizionali (iniziale, interna, finale di turno o prima di pausa), semantici (parole funzionali vs. parole lessicali); mediante una serie di test di regressione lineare è stato indagato il rapporto tra durata e valori formantici (F1 e F2) di ciascuna entrata vocalica, per ciascuna località e per ciascuna delle tipologie *supra* descritte. In questo modo, si avrà anche per il parlato toscano un quadro aggiornato relativo al controverso rapporto tra andamenti spettrali e durata (Moon & Lindblom 1994; Savy & Cutugno 1997).

Per quanto concerne il secondo aspetto – l'analisi qualitativa dei tracciati formantici – la vocale viene considerata come una traiettoria e non come un punto statico, seguendo peraltro una tendenza ormai largamente diffusa negli studi relativi alle vocali, considerate sempre più spesso come segmenti dinamici (Lisker 1984; Di Benedetto 1989a, 1989b; Cerrato & Cutugno 1994; Hillenbrand *et alii* 1995). Per mezzo di uno *script* realizzato da Beat Siebenhaar (Università di Lipsia) sono stati misurati i valori di F0, F1, F2, F3 dal 5% fino al 95% della durata di quei segmenti del nostro *corpus* (soprattutto nomi di toponimi e indicazioni stradali) che limitassero – nella misura del possibile – la presenza di fattori di disturbo (influenza del contesto consonantico, della struttura accentuale e prosodica), in un confronto che prende in esame ciascuna località indagata.

Il fatto che il *corpus* sia diversificato in base al fattore 'luogo' permette di osservare, da un lato, una ridotta influenza del sostrato dialettale in relazione alla sottospecificazione acustica dipendente da accorciamenti di durata e, dall'altro, permette di raffinare ulteriormente il quadro relativo al vocalismo livornese che mostra nel parlato semispontaneo andamenti formantici estremamente instabili soprattutto per quanto concerne le vocali 'bandiera'.

RIFERIMENTI BIBLIOGRAFICI

- Calamai, S. (2004a), *Il vocalismo tonico pisano e livornese. Aspetti storici, percettivi, acustici*, Alessandria, Edizioni dell'Orso.
- Calamai, S. (2004b), "Vocali fiorentine e vocali pisane a confronto", in F. Albano Leoni, F. Cutugno, M. Pettorino & Savy R. (acd), *Atti del Convegno Nazionale Il parlato Italiano*, Napoli 13-15.II.2003, Napoli, D'Auria, CD-rom, 2004, E02: 1-25.
- Calamai, S. & Ricci, I. (2005), "Un esperimento di matched-guise in Toscana", *Studi Linguistici e Filologici on Line* (Dipartimento di Linguistica – Università di Pisa www.humnet.unipi.it/slifo.html), 3.1: 63-105.
- Cerrato, L. & Cutugno F. (1994), "Il problema della rappresentazione tempo/frequenza dei fenomeni vocalici dinamici", in F.E. Ferrero & E. Magno Caldognetto (acd), *Le vocali: dati sperimentali, problemi linguistici, applicazioni tecnologiche. Atti delle IIIe Giornate di Studio del Gruppo di Fonetica Sperimentale (AIA)*, Padova 19-20.XI.1992, "Collana degli Atti dell'Associazione Italiana di Acustica", 20: 61-71.
- Di Benedetto, M.G. (1989a), "Vowel Representation: Some Observations on Temporal and Spectral Properties of the First Formant Frequency", *The Journal of the Acoustical Society of America*, 86: 55-66.
- Di Benedetto, M.G. (1989b), "Frequency and Time Variation of the First Formant: Properties Relevant to the Perception of Vowel Height", *The Journal of the Acoustical Society of America*, 86: 67-77.
- Hillenbrand, J., Getty L.A., Clark M.J. & Wheeler K. (1995), "Acoustic Characteristics of American English Vowels", *The Journal of the Acoustical Society of America*, 97: 3099-3111.
- Lisker, L. (1984), "On Reconciling Monophthongal Vowel Percepts and Continuously Varying F Patterns", *Status Report on Speech Research, Haskins Laboratories*, 79/80: 167-174.
- Giannelli, L. (2000) [1976], *Toscana*, in Zamboni, A. (a cura di) *Profilo dei dialetti italiani*, Pacini, Pisa.
- Marotta, G., Calamai, S. & Sardelli, E. (2004), "Non di sola lunghezza. La modulazione di f0 come indice sociofonetica", in *Costituzione, gestione e restauro di corpora vocali. Atti delle XIVe Giornate di Studio del GFS* (A. De Dominicis, L. Mori, M. Stefani, a cura di), Università della Tuscia (Viterbo), 4-6.XII.2003, Roma: Esagrafica, 215-220.
- Moon, S.-J. & Lindblom B. (1994), "Interaction between Duration, Context, and Speaking Style in English Stressed Vowels", *The Journal of the Acoustical Society of America*, 96: 40-55.
- Nocentini, A. (1989), *Il vocabolario aretino di Francesco Redi*, Firenze, Elite.
- Savy, R. & Cutugno F. (1997), "Ipoarticolazione, riduzione vocalica, centralizzazione: come interagiscono nella variazione diafasica?", in F. Cutugno (acd), *Fonetica e fonologia degli stili dell'italiano parlato. Atti delle VIIe Giornate di Studio del Gruppo di Fonetica Sperimentale (AIA)*, Napoli 14-15.XI.1996, "Collana degli Atti dell'Associazione Italiana di Acustica", 24: 177-194.
- Siebenhaar, B. (online), *Online-Einführung in Praat*. Disponibile alla pagina:
<http://www.germanistik.unibe.ch/siebenhaar/subfolder/PraatEinfuehrung/index.html>

LO SPAZIO DI -U NEL DIALETTO DI MATELICA

Tania Paciaroni
Università di Zurigo
paciaron@rom.uzh.ch

SOMMARIO

La varietà parlata a Matelica appartiene ai dialetti dell'Italia mediana – l'area a sud-est della linea Roma-Ancona –, caratterizzata dal mantenimento della distinzione tra *-o* e *-u*. Tale distinzione ha base fondamentalmente etimologica, ma presenta estensione notevolmente diversa all'interno delle singole varietà, come documentato fin dai più antichi testi in volgare (v. Vignuzzi 1994: 343). L'analisi in Paciaroni / Loporcario (in stampa) mostra che anche tra le varietà odierne dell'area maceratese la distribuzione di *-u* (< -_) e *-o* (< -_, -_) è tutt'altro che compatta: mentre l'esito di *-o* è sempre [o], la realizzazione della *-u* oscilla tra il grado di apertura medio e quello alto. La stessa analisi mostra inoltre come quest'oscillazione e la conseguente estensione di *-o* siano da ricondurre a tipi differenti di mutamento. In alcune varietà l'estensione di *-o* è funzionalizzata al marcamento di genere (neo-)neutro vs. maschile (lo esemplificano in maceratese le forme [lo 'pjum:o] 'il piombo (metallo)' n. di contro a [lu 'pjum:u] 'il piombino (oggetto)' m. (pl. [li 'pjum:i])); in altre parlate, invece, la medesima estensione obbedisce ad un processo di armonia vocale che uniforma la *-u* alla qualità della vocale tonica (secondo lo sviluppo descritto da Merlo (1920; 1922) per la varietà di Cervara): ad es. a San Severino [u 'to:ro] 'il toro' m. anziché [u 'to:ru] (pl. [i 'to:ri]), [o 'fero] 'il ferro (metallo)' n., ['gros:o] 'grosso' m.=n. Anche nella varietà matelicese, contrariamente alla letteratura dialettologica che documenta ampiamente la saldezza di una distinzione etimologica *-u* / *-o* (dall'elenco di forme in Leopardi (1887: 69) a lavori recenti come il *Glossario* curato da Traballoni (2001/2002)), l'osservazione diretta avviata in quell'indagine mostra che lo spazio acustico della *-u* esibisce una grande variabilità.

Per descrivere in modo adeguato di quale variabilità si tratti, è stata effettuata un'indagine sul campo volta ad illustrare il vocalismo della varietà parlata a Matelica, con particolare riguardo alla realizzazione di *-o* e *-u* finali (v. per un primo saggio d'indagine Paciaroni 2007). Tenendo conto del problema del "paradosso dell'osservatore", è stato raccolto materiale sonoro esemplificativo di stili diversi lungo il *continuum* tra formale e informale. *In primis*, per elicitarne un parlato il più spontaneo possibile, sono state condotte con gli informatori conversazioni semiguide su temi come la coltivazione dei campi, la vita del paese ecc., poi è stato somministrato loro un questionario per ottenere le parole bersaglio in posizione interna di frase, infine è stato chiesto agli stessi parlanti di pronunciare le parole bersaglio in isolamento.

I soggetti scelti sono tre uomini (GM, BM, DM) di età compresa tra i 45 e i 70 anni nativi e residenti a Matelica.

L'analisi è stata compiuta utilizzando il software *Multispeech*. I parametri considerati sono stati i valori di F1 e di F2, e la durata del segmento vocalico. Nel quadro relativo ad ognuno dei tre stili sono stati presi in esame separatamente prima i valori medi dei singoli parlanti, poi i valori medi globali.

L'analisi conferma la variabilità fonetica di *-u* e dimostra la correlazione tra le proprietà acustiche della vocale da un lato, lo stile dell'enunciato e la velocità d'eloquio dall'altro. I risultati mostrano inoltre che la variabilità interessa la *-u* etimologica in misura diversa a seconda delle categorie lessicali in cui viene a trovarsi. Nelle parole funzionali (articolo, clitico e dimostrativo) le aree di *-u* e di *-o* sono chiaramente distinte in tutti e tre gli stili; diversamente, nell'aggettivo, nel participio e nel nome, l'area di *-u* rimane compatta e separata da quella di *-o* nelle parole in isolamento (in linea con la categorizzazione nel giudizio dei parlanti), mentre occupa uno spazio progressivamente più esteso nelle parole in frase e nel parlato spontaneo, ove per tutti e tre i parlanti si osserva una parziale sovrapposizione all'area di [o].

RIFERIMENTI BIBLIOGRAFICI

- Leopardi, Alfonso (1887): *Sub tegmine fagi. Sotto un tegame di fagioli*. Città di Castello: S. Lapi.
- Maturi, Pietro e Stephan Schmid (2001), *Allomorfia e morfo-fonetica: riflessioni induttive su dati dialettali campani*, in Federico Albano Leoni, Stenta Krosbakken, Rosanna Sornicola, Carolina Stromboli (a cura di), *Dati empirici e teorie linguistiche*, Roma, Bulzoni, 251-265.
- (2003), *Sulla diffusione areale di un fenomeno di variazione morfofonetica nei dialetti campani*, in Fernando Sanchez Miret (ed.), *Actas XXIII Congreso Internacional de Lingüística y Filología Románica. Vol. 1*, Tübingen, Niemeyer, 221-233.

- Merlo, Clemente (1920): *Fonologia del dialetto di Sora (Caserta)*. In: *Annali delle Università Toscane* 38, 117-283 [poi vol. a sé, Pisa: Mariotti, 1920].
- (1922): *Fonologia del dialetto della Cervara in provincia di Roma*. Roma: Società Filologica Romana.
- Paciaroni, Tania e Michele Loporcaro (in stampa), *Funzioni morfologiche dell'opposizione fra -u e -o nei dialetti del Maceratese*, in stampa in *Actes du CILPR XXV*, Innsbruck, 3-8 September 2007.
- Paciaroni, Tania (2007), *Sulla distribuzione di -u e -o nella flessione nominale e pronominale del maceratese e dei dialetti limitrofi*. Poster/comunicazione al XXXII Convegno annuale della Società Italiana di Glottologia (S.I.G.), Verona, 25-27 ottobre 2007.
- Traballoni, Michela (2001/2002), *Glossario del dialetto di Matelica*, Tesi di laurea, ms Università di Macerata.
- Vignuzzi, Ugo (1988): *Italienisch: Areallinguistik, VII. Marche, Umbrien, Lazio*. In: *LRL* 4, 606-642.
- (1994): *Il volgare dell'Italia mediana, in Storia della lingua italiana. III. Le altre lingue*, 329-372.

BALBUZIE E COARTICOLAZIONE

Caterina Pisciotto[°], Massimiliano Marchiori[#], Claudio Zmarich[°]

Istituto di Scienze e Tecnologie della Cognizione (ISTC), C.N.R., Sede di Padova[°]; Psicologo, libero professionista[#]
caterina.pisciotto@pd.istc.cnr.it[°], marchiori.massimiliano@yahoo.it[#], claudio.zmarich@pd.istc.cnr.it[°]

SOMMARIO

Alcune recenti ricerche sulla balbuzie suggeriscono che la coarticolazione, intesa genericamente come l'influenza di un fono su un altro ad esso adiacente, nei balbuzienti può essere diversa da quella riportata per i non balbuzienti. Queste ricerche partono da una ben radicata convinzione che la balbuzie origini dalle difficoltà nella realizzazione della transizione tra l'inizio della consonante e il nucleo vocalico della sillaba, specie quando quest'ultima è accentata e inizia la parola (Wingate, 1988). Tra le ricerche più recenti, Robb e Blomgren (1997), analizzando le sequenze CV percepite come fluenti in soggetti adulti balbuzienti e normoparlanti, hanno trovato che le transizioni di F2 dalla consonante alla vocale della sillaba erano caratterizzate da pendenze significativamente maggiori per i soggetti balbuzienti. Dal punto di vista articolatorio, esse suggeriscono modificazioni nella dimensione del tratto vocale che sono più ampie e/o più rapide del normale.

Esiste poi un altro filone di studi fonetici sulla balbuzie che si sofferma sulle influenze prosodiche nel linguaggio dei balbuzienti, che mostrano come i balbuzienti siano effettivamente in grado di realizzare le differenze prosodiche tra parole focalizzate e non focalizzate, ma aumentano la frequenza delle disfluenze sulle parole in focus (vedi per es. Bergmann, 1986; Marchiori, Zmarich, Avesani e Bernardini, 2005).

Nel tentativo di unificare gli spunti offerti da queste due impostazioni di studio sulla balbuzie, Zmarich e Marchiori (2005) avevano analizzato le influenze prosodiche della coarticolazione di sillabe CV in presenza o meno di accento lessicale sotto focus informativo o contrastivo, con lo scopo di verificare se nei balbuzienti l'influenza coarticolatoria di tipo anticipatorio di V su C è significativamente differente di quella dei non balbuzienti. E' da notare che le sillabe con accento contrastivo vengono prodotte con più sforzo e sono più resistenti ai processi coarticolatori rispetto alle sillabe senza accento lessicale e in quelle con accento lessicale ma senza accento intonativo (cfr. Zmarich, Avesani e Marchiori, 2007). Essi trovarono che i balbuzienti tendenzialmente coarticolano di meno rispetto al gruppo di non balbuzienti, cioè presumibilmente si sottopongono ad un maggior sforzo articolatorio, ma il risultato non era statisticamente significativo.

Questo lavoro prende le mosse dai suddetti studi di Zmarich e Marchiori (2005) e Marchiori, Zmarich, Avesani e Bernardini (2005), e si propone di determinare il grado di coarticolazione anticipatoria di V su C sulla prima sillaba della parola, pronunciata in modo percettivamente fluente, in due contesti prosodici che favoriscono opposti livelli di coarticolazione (minimo vs massimo, cfr. Zmarich, Avesani e Marchiori, 2007). Inoltre, la prima sillaba di una parola costituisce notoriamente per i balbuzienti una posizione critica perchè è sempre stata trovata essere associata in modo quasi esclusivo con la disfluenza (v. per es. Marchiori, Zmarich, Avesani e Bernardini, 2005).

Hanno preso parte all'esperimento quattro soggetti balbuzienti adulti e quattro non balbuzienti. Ciascun locutore ha letto 3 volte un *corpus* di 6 frasi composte di Nome-Verbo o Verbo-Nome nelle quali il verbo è una parola naturale, mentre il nome è una pseudo-parola trisillabica nella quale la posizione dell'accento lessicale è stata variata sistematicamente sulle tre sillabe: "viene dadada/dididi", "dadada/dididi viene". Ogni enunciato è costituito dalla lettura di una risposta informativamente adeguata ad una domanda presentata in cuffia ai locutori. In tal modo si sono ottenuti enunciati con *focus* ampio o "neutro" (*focus* distribuito sull'intera frase, ad es. "[viene dadada/dididi]_F"), con *focus* ristretto alla parola iniziale (ad es. dadada/dididi]_F viene") e con *focus* ristretto alla parola finale (ad es. "viene dadada/dididi]_F). Sia il nome che il verbo possono costituire il *focus* ristretto della frase.

La metrica scelta per investigare la coarticolazione CV nelle sillabe che sono state giudicate dagli autori come fluenti dal punto di vista percettivo (cioè non balbettate), è quella delle equazioni di luogo (*Locus equations*, cfr. Sussman et al. 1991), che ricava i valori di pendenza e di intercetta della retta di regressione lineare che interpola le coppie di valori di F2 (Hz) di un numero sufficiente di sillabe, misurati all'inizio della transizione CV e al centro di V. Secondo Sussman, le equazioni di luogo forniscono un indice che permette di quantificare il grado di coarticolazione C-V.

I risultati evidenziano come in condizioni massimamente critiche per il balbuziente (grado minimo di coarticolazione, grado massimo di suscettibilità alle disfluenze) i parlanti balbuzienti esibiscono un livello di coarticolazione significativamente maggiore rispetto a quello dei parlanti non balbuzienti. Questi risultati confermano lo studio di Subramanian, Yairi e Amir (2003), che esaminarono le transizioni di F2 nel parlato percettivamente fluente di bambini prescolari registrati subito dopo l'inizio della balbuzie. I soggetti furono poi seguiti per diversi anni rendendo quindi possibile distinguere i bambini che avevano smesso spontaneamente di balbettare da quelli che continuavano a farlo. I risultati dimostrarono che i movimenti articolatori, inferiti dalle transizioni di F2, erano più ristretti (con più coarticolazione) per i bambini destinati a cronicizzare. Forse questi

soggetti mostrano alcuni problemi proprio nella coordinazione fine richiesta per il controllo della normale produzione linguistica. Il fatto che questo studio sia stato condotto a partire dalle fasi iniziali d'insorgenza della balbuzie suggerisce la presenza di eventuali deficit già allo stadio formativo del disordine.

Il quesito che ci si pone ora è se questo livello eccessivo di coarticolazione è un sintomo diretto della balbuzie, oppure una reazione reattiva, di tipo secondario, alla balbuzie; in altre parole, bisogna chiedersi: i balbuzienti adulti (e i bambini balbuzienti destinati a cronicizzare) coarticolano di più per le difficoltà motorie causate dalla balbuzie, o per cercare di evitarla/controllarla?

Per rispondere a questo interrogativo, analizzeremo, attraverso l'analisi acustica, le produzioni disfluenti dei balbuzienti: se anche le ripetizioni e i prolungamenti di suono e le ripetizioni di sillaba saranno caratterizzate da un livello eccessivo di coarti colazione, sarà facile concludere che essa è allora un sintomo diretto della balbuzie.

BIBLIOGRAFIA

- Bergmann, G. (1986), Studies in stuttering as a prosodic disturbance, *Journal of Speech and Hearing Research*, 47, 778-82.
- Marchiori, M., Zmarich, C., Avesani, C., Bernardini, S. (2004), Focus e prosodia nelle produzioni verbali dei 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, EDK Editore, Brescia, 2005, 251-286.
- Robb, M. & Blomgren, M. (1997), Analysis of F2 transitions in the speech of stutterers and nonstutterers, *Journal of Fluency Disorders*, 22, 1-16.
- Subramanian, A., Yairi, E., Amir, O. (2003), Second formant transition in fluent speech of persistent and recovered preschool children who stutter, *Journal of Communication Disorder*, 36, 59-75.
- Sussman, H.M., McCaffrey, H., & Matthews, S. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization, *Journal of the acoustical society of America*, 90, 1309-1325.
- Wingate, M.E. (1988), *The structure of stuttering (a psycholinguistic analysis)*, Springer Verlag, New York.
- Zmarich, C., Marchiori, M. (2004), 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, EDK Editore, Brescia, 2005, 231-250.
- Zmarich, C., Avesani, C., Marchiori, M. (2006), Coarticolazione e Accentazione, in *Atti del III Convegno Nazionale dell'Associazione Italiana di Scienze della Voce (AISV)*, (V. Giordani, V. Bruseghini, Cosi P., a cura di), Trento, 29-30/11-1/12/2006, EDK Editore srl, Torriana (RN), 2007, 5-15.

STUDIO ELETTROPALATOGRAFICO DELL'ASSIMILAZIONE NEI NESSI DI NASALE + VELARE IN ITALIANO

Irene Ricci¹, Silvia Calamai², Chiara Bertini¹, Chiara Celata¹

¹ Scuola Normale Superiore, Pisa, ² Università di Siena

i.ricci@sns.it, calamai@unisi.it, c.bertini@sns.it, c.celata@sns.it

SOMMARIO

INTRODUZIONE: Recenti studi elettropalatografici [2, 4, ma vedi anche 6, 7] hanno mostrato che, in inglese, il fenomeno di assimilazione del punto di articolazione di una nasale finale di parola seguita da una occlusiva velare è graduale, condizionato da fattori esterni (es. stilistici), altamente variabile nel confronto tra soggetti e persino all'interno delle produzioni di un singolo parlante. In inglese la nasale velare che scaturisce dal fenomeno di assimilazione su giuntura risulta pertanto fundamentalmente diversa dalla nasale velare lessicale (fonologica). Obiettivo di questa ricerca è l'analisi della realizzazione dei nessi di nasale + occlusiva velare in italiano, ove il fenomeno è esclusivamente allofonico e può presentarsi tanto all'interno di parola (es. ancora) quanto su giuntura (es. in casa). Trattandosi di un processo contestuale [5], l'assimilazione del punto di articolazione della nasale all'occlusiva velare seguente è normalmente considerato un fenomeno obbligatorio e categorico, così come viene illustrato anche da un precedente studio EPG condotto su diverse varietà di italiano [3].

IPOSTESI, METODO, ANALISI: Scopo principale di questa indagine è verificare il livello di categoricità/gradualità del fenomeno in italiano, da un lato confrontandolo direttamente con quanto è stato trovato sull'inglese [2], dall'altro osservando l'effetto di variabili prosodiche/stilistiche che non venivano prese in considerazione in [3].

In particolare, viene introdotto un controllo statistico della variabilità intra- e inter-soggettiva, affiancando quindi all'analisi qualitativa consueta un tentativo di analisi quantitativa. Sempre con l'intenzione di ottenere un quadro più chiaro della variabilità intra-soggettiva, è stata introdotta una fase di adattamento al palato della durata di circa mezz'ora, prima dell'inizio delle registrazioni, durante la quale ai soggetti era proposta una pratica focalizzata [1].

Tre soggetti toscani di area toscana hanno prodotto dieci ripetizioni per ognuna delle sedici frasi sperimentali contenenti le parole target. Oltre ai nessi /nk/ e /ng/, vengono considerati /nt/ e /nd/ come nessi di controllo per la realizzazione alveolare della nasale. I fattori rilevanti sono la posizione dell'accento (es. *máncano* vs. *mancáto*), la presenza/assenza di un confine di parola (es. *mancato* vs. *fan caso*), la velocità di eloquio (ai soggetti viene chiesto di leggere le frasi sia a velocità normale, sia a velocità ridotta, come se stessero parlando a uno straniero).

L'analisi qualitativa mira a individuare gli schemi articolatori prototipici della nasale alveolare e della nasale velare assimilata. Per quanto riguarda l'analisi quantitativa, invece, oltre alla misurazione delle durate delle intere frasi e dei nessi, vengono estratti gli indici articolatori del Centro di Gravità (COG), della velarità (VELAR) e dell'anteriorità (ALVEOLAR), e su di essi è stata condotta analisi parametrica.

RISULTATI: Il risultato principale riguarda il fatto che il punto di articolazione della nasale quando seguita da occlusiva velare si differenzia notevolmente e praticamente senza eccezioni dal punto di articolazione della nasale seguita da alveolare, indicando così che il fenomeno di assimilazione è pervasivo e provoca uno spostamento radicale della regione in cui è realizzato il contatto linguo-palatale. I due schemi articolatori (alveolare vs. velare assimilato) si differenziano drasticamente e sulla totalità dei dati: contrariamente a quanto rilevato per l'inglese, non si hanno esempi, nei nostri dati, di realizzazioni alveolari della nasale seguita da occlusiva velare, né di realizzazioni "parzialmente assimilate".

Ciononostante, alcuni fattori condizionano il grado di posteriorità raggiunto dalla nasale quando seguita da una velare. La presenza di un confine di parola inibisce parzialmente il fenomeno dell'assimilazione, che produce valori di COG e di VELAR significativamente più alti quando il nesso è all'interno di parola. Inoltre, anche la velocità di eloquio conta: quando essa è rallentata, e i soggetti tendono a produrre realizzazioni iperarticolate, il livello di variabilità intra-soggettiva aumenta notevolmente (e di conseguenza, anche quello inter-soggettivo ne risente), indicando pertanto che l'assimilazione di velarità è un fenomeno più frequente e con caratteristiche più estreme nel parlato normoarticolato che non in uno stile controllato. In questo senso, tali risultati impongono anche di considerare con maggiore cautela l'apparenza di categoricità che il fenomeno possedeva in precedenti studi sull'italiano.

RIFERIMENTI BIBLIOGRAFICI

- [1] Bertini, Ricci (in stampa) Tempi di adattamento in elettropalatografia: primi dati articolatori, acustici e percettivi, Atti del IV Convegno Nazionale AISV 2007.
- [2] Ellis, Hardcastle 2002 Categorical and gradient properties of assimilation in alveolar to velar sequences: evidence from EPG and EMA data, *JoPh* 30, 373-396
- [3] Farnetani, Busà 1994 Consonant-to-consonant interactions in clusters: categorial or continuous processes? *QCSRF CNR Padova*, 220-245.
- [4] W.J. Hardcastle, Assimilation of alveolar stops and nasals in connected speech. In: J. Windsor Lewis, Editor, *Studies in General and English Phonetics in Honour of Professor J.D. O'Connor*, Routledge, London (1994), 49-67.
- [5] Ohala, J.J. (1990) The phonetics and phonology of aspects of assimilation. In J. Kingston, M.E. Beckman (eds). *Papers in Laboratory Phonology I: Between the grammar and physics of speech*, 258-275.
- [6] Stephenson 2003 An EPG study of repetition and lexical frequency effects in alveolar to velar assimilation, *Proceedings of the 15th International Congress of Phonetic Sciences (ICPHS-03)*, 1891-1894.
- [7] Stephenson, Harrington 2002 Assimilation of place of articulation: Evidence from English and Japanese, *Proceedings of the 9th Australian International Conference on Speech Science and Technology*, Melbourne, 2002.

VARIABILITÀ RITMICA DI VARIETÀ DIALETTALI DEL PIEMONTE

Antonio Romano, Paolo Mairano & Barbara Pollifrone
Laboratorio di Fonetica Sperimentale "Arturo Genre" - Università degli Studi di Torino
antonio.romano@unito.it, paolomairano@gmail.com, polbarbara@yahoo.it

SOMMARIO

Nel vasto e variegato panorama dei dialetti gallo-italici, le parlate del Piemonte costituiscono uno spazio tutt'altro che omogeneo (Berruto, 1974; Telmon, 1988, 2001). Al di là delle aree d'insediamento di minoranze linguistiche storiche (alcune delle quali in realtà caratterizzate dalla presenza di comunità plurilingui e, talvolta, da complesse stratificazioni di fenomeni areali che ne rendono difficile una delimitazione certa), anche le parlate del resto della regione sono soggette a fenomeni di contatto con quelle delle altre aree contigue e presentano proprietà fonetiche talvolta ritenute tipiche di altre regioni.

Com'è facile immaginare, studi vari e differenziati nel corso degli anni hanno esplorato numerose di queste proprietà in prospettiva geolinguistica. In questo quadro s'inseriscono le ricerche dei grandi cantieri atlantistici che proprio in questa regione hanno sede, ma anche studi specifici di fonetica acustica condotti su variabili diverse (Genre 1980, 1992). È in questi lavori che si trovano incidentalmente riflessioni su caratteristiche prosodiche che potrebbero concorrere alla caratterizzazione delle varie parlate e contribuire a una loro classificazione che includa proprietà sovrasegmentali.

Un contributo alla collocazione ritmica delle parlate piemontesi in rapporto ad altre varietà italo-romanze è stato recentemente proposto da Schmid (2004) che ha sfruttato i dati presenti nel disco allegato a Berruto (1974) per mostrarne l'appartenenza a un gruppo di lingue dalle caratteristiche più isoaccentuali (IA) rispetto ad altre varietà italo-romanze più isosillabiche (IS).

In particolare Schmid (2004), sulla scia di Dauer (1983) e Bertinetto (1989), ha verificato le relazioni tra le occorrenze di diversi tipi sillabici (in base a un'attenta osservazione degli inventari sillabici di ciascuna varietà osservata) e la corrispondenza tra questi sistemi e la classificazione suggerita dagli indicatori di Ramus *et alii* (1999), registrando per la parlata piemontese considerata il valore più alto (35)⁹.

Riferendoci allo stesso quadro metodologico, in questo studio ci proponiamo di passare in rassegna, con una tecnica sperimentale già estensivamente adottata in studi precedenti su altre lingue (v. Mairano & Romano, 2007), le caratteristiche ritmiche di alcune parlate piuttosto distanti tra loro, appositamente scelte alla periferia di questa regione linguistica. Si tratta di quelle di Roccaforte Ligure (AL), Briga Alta (CN), Exilles (TO) e Capanne di Marcarolo (AL) (studiate a partire dai dati raccolti e analizzati presso i nostri laboratori in contributo ai volumi 27, 28, 30 e 33 dell'*Atlante Toponomastico del Piemonte Montano*; v. *ATPM*, 2005-2008) e di quelle di Campertogno (VC) e Bagnolo Piemonte (CN) (i cui dati sono stati raccolti rispettivamente nell'ambito della recente monografia di Molino & Romano, 2008, e della tesi di laurea inedita di Piccato, 2007)¹⁰.

I risultati collocano agli estremi opposti le due varietà più vicine al ligure: quella di Roccaforte L. (fortemente caratterizzata per via dei suoi dittonghi discendenti) si situa infatti tra quelle più IA (alti V e C), mentre quella di Capanne di M. (che conserva meglio il vocalismo atono finale e evita gli allungamenti in sillaba chiusa) tra quelle IS (medio V e basso C). Campertogno, gravitante in area lombarda, si mostra quella col più alto C mentre Exilles e Briga A. (rispettivamente di area occitana e ligure) si caratterizzano per un C medio ma un alto V. Pur restando in area IA, la varietà di Bagnolo P. è invece quella che – lavorando con metriche

⁹ Si noti tuttavia come il fatto che il piemontese sia (insieme al milanese e ad altre parlate meridionali) tendenzialmente IA era già stato suggerito da Trumper *et alii* (1991; citato in Schmid, 2004: 111) e da Mayerthaler (1996, citato in Schmid, 2004: 112). Cfr. anche Mendicino & Romito (1991) e Romito & Trumper (1993).

¹⁰ La tecnica adottata si basa su valutazioni effettuate su campioni di parlato letto di durata compresa tra poco più di mezzo minuto e un minuto scarso: versioni locali de *La tramontana e il sole* raccontate da un locutore per punto (rispettivamente: *er vèntu de tramuntan-a e u su, a binda e 'l sù, l'aura fr_i e 'l sùrè, u vèntu de tramuntan-a e u su(l), al vènt e 'l sò, l'òra e 'l sul*). I brani sono segmentati in intervalli vocalici e consonantici grazie al programma PRAAT (per mezzo di un'accurata procedura che valuta, fonologicamente, la categoria di appartenenza del suono atteso e, foneticamente, quella naturale del suono realizzato). I file *TextGrid* prodotti sono analizzati automaticamente mediante un applicativo *Tcl/Tk* (v. Mairano & Romano, in prep.) che produce l'analisi dei diversi correlati e ne permette una rappresentazione grafica.

parziali di unità interpausale (escludendo quindi le unità più brevi) – più si avvicina a una posizione intermedia tra i due poli¹¹.

Le distinzioni non mancano di seguire il gradiente di altri fenomeni, come appunto quelli legati alle riduzioni postaccentuali: se infatti a Exilles *forte* perde totalmente la sua ultima sillaba (come accade, con dati incostanti, per Bagnolo), a Campertogno si ha ancora la perdita della sola vocale finale e a Capanne la conservazione (con Briga e Roccaforte ancora propense a perderla o a desonorizzarla più spesso).

RIFERIMENTI BIBLIOGRAFICI

- ATPM - Atlante Toponomastico del Piemonte Montano (2005-2008: 27-Roccaforte Ligure; 28-Briga Alta, 30-Exilles e 33-Capanne di Marcarolo).
- Berruto, G. (1974). Piemonte e Valle d'Aosta. In M. Cortelazzo (a c. di), "Profilo dei dialetti italiani", 1, Pisa, Pacini.
- Bertinetto, P.M. (1989). "Reflections on the dichotomy 'stress' vs. 'syllable-timing'". *Revue de Phonétique Appliquée*, Mons, 99-130.
- Bertinetto, P.M. & Bertini, C. (2008). "On modeling the rhythm of natural languages". *Proc. of the 4th International Conference on Speech Prosody (Campinas, Brazil, May 2008)*, 427-430.
- Dauer, R.M. (1983). "Stress-timing and syllable-timing reanalyzed". *J. of Phonetics*, 11, 51-62.
- Genre, A. (1980). "Le parlate occitano-alpine d'Italia", *Rivista Italiana di Dialettologia*, 4, 305-310.
- Genre, A. (1992). "Nasali e nasalizzate in Val Germanasca", *Rivista Italiana di Dialettologia*, 16, 181-224.
- Mairano, P. & Romano, A. (2007). "Inter-Subject Agreement in Rhythm Evaluation for Four Languages (English, French, German, Italian)". *Proc. of the 16th International Congress of Phonetic Sciences (Saarbrücken, Germany, August 2007)*, 1149-1152.
- Mairano, P. & Romano, A. (in prep.). "Un confronto tra diverse metriche ritmiche usando Correlatore 1.0". Articolo non pubblicato (in prep.).
- Molino, P. & Romano, A. (2008). *Il dialetto valsesiano nella media Valgrande*. Alessandria, Dell'Orso.
- Piccato, E. (2007). "La parlata di Bagnolo Piemonte". Tesi di Laurea (rel. A. Romano), Fac. di Lingue e Lett. Str. dell'Univ. degli Studi di Torino, inedita.
- Mendicino, A. & Romito, L. (1991). "«Isocronia» e «base di articolazione»: uno studio su alcune varietà meridionali". *Quaderni del Dip. di Ling. dell'Univ. della Calabria*, S. L. 3, 49-67.
- Ramus, F., Nespors, M. & Mehler, J. (1999). "Correlates of linguistic rhythm in the speech signal". *Cognition*, 73/3, 265-292.
- Romito, L. & Trumper, J. (1993). "Problemi teorici e sperimentali posti dall'isocronia". *Quaderni del Dip. di Ling. dell'Univ. della Calabria*, S. L. 4, 10, 89-118.
- Schmid, S. (2004). "Une approche phonétique de l'isochronie dans quelques dialectes italo-romans". In: Meisenburg, T., Selig, M. (eds.), *Nouveaux départs en phonologie*, Tübingen: Narr, 109-124.
- Telmon, T. (1988). "Areallinguistik II. Piemont". In G. Holtus, M. Metzeltin & Chr. Schmitt, *Lexicon der Romanistischen Linguistik*, Vol. IV, Tübingen, Niemeyer, 469-485.
- Telmon, T. (2001). Piemonte e Valle d'Aosta. In A.A. Sobrero (a cura di), *Profili linguistici delle regioni*, Bari, Laterza.

¹¹ Nell'interpretazione dei risultati ottenuti per queste varietà piemontesi, al momento sono state considerate solamente le metriche di Ramus *et alii* (1999), mentre i grafici corrispondenti al calcolo di altri correlati più recenti (*PVI* e *CCI*, v. Bertinetto & Bertini, 2008) necessitano ancora di una lettura più accurata.

STABILITÀ DEI PARAMETRI NELLO SPEAKER RECOGNITION: LA VARIABILITÀ INTRA E INTER PARLATORE – F0, DURATA E *ARTICULATION RATE*

Luciano Romito*, Rosita Lio*, Pier Francesco Perri[^], Sabrina Giordano[^]
Laboratorio di Fonetica Università della Calabria*, Dipartimento di Economia e Statistica Università della Calabria[^]
luciano.romito@unical.it, lio.rosita@libero.it

SOMMARIO

La tendenza della ricerca attuale in ambito di *Speaker Recognition* (SR) è volta a individuare informazioni quanto più oggettive possibili presenti nella voce umana ed analizzare la produzione di un parlatore senza occuparsi della sfera semantica, della produzione linguistica, o della costruzione sintattica e morfologica. In aggiunta i metodi noti come semiautomatici e parametrici si occupano di dati statici e non dinamici (al contrario di quanto avviene nei metodi completamente automatici). Tale scelta in primo luogo è giustificata dalla relativa facilità della misura e del trattamento di un ristretto numero di parametri (cfr. Barlow and Wagner 1998) e in secondo luogo perché la misura di dati statici è la naturale evoluzione di una tradizionale analisi linguistica (cfr. McDougall 2006).

Sono le variabili statiche quelle utilizzate per lo studio delle lingue, si pensi agli inventari fonologici, alle aree di esistenza delle vocali costruite su porzioni stazionarie (mid point o stady state), alle rotazioni consonantiche o alle regole fonologiche. Tale analisi prende lo spunto dalla necessità di differenziare due lingue, due dialetti o una lingua da un dialetto. Così, grande spazio nelle riviste, occupano concetti quali isoglosse o isofone utilizzati per identificare confini ideali tra due lingue o tra due dialetti. Retroflessione di esiti di LL latina, presenza vs assenza di occlusive aspirate, metafonìa ecc.

Quanto detto risulta funzionale per differenziare ma non per riconoscere, o addirittura identificare. Di fatto anche il concetto di isoglossa oggi viene sostituito dall'idea più analogica di corridoio di transizione, una larga fascia dove coesistono variabili differenti che caratterizzano entrambe le lingue o i dialetti contigui.

Un parlante nel produrre un messaggio, un atto comunicativo attraverso un meccanismo astratto (linguistico) organizza, target e goal che in seguito verranno tradotti in azioni che si realizzeranno in un progetto fonetico. Il meccanismo linguistico è essenzialmente l'insieme delle regole, della grammatica del parlante, è la lingua costituita da lessico, morfologia, opposizioni fonologiche, sintassi ecc. Tale meccanismo è fortemente influenzato dall'età, dal sesso, dal controllo fonologico, da fattori sociali quali l'origine geografica, lo stato economico, il contesto, la scolarizzazione ecc.. Nolan nel 1997 (p. 749) scrive: "In implementing the resources of their linguistic mechanism, speakers have to map them onto their individual anatomy. Whilst the requirements of communication may determine many of the details of speech articulation, we may hypothesize that there may be aspects of speech production where each individual is free to find his or her own articulatory solution. The speaker's behavior here is not 'learned' as part of the shared knowledge of the linguistic community; rather it is acquired, probably by trial and error".

Due differenti parlanti possono eseguire progetti fonetici differenti per lo stesso scopo linguistico. Le conseguenze acustiche di tale progetto possono aiutare molto nel differenziare i due parlanti, ma fatto estremamente importante, le conseguenze acustiche, non riguarderanno la parte statica del segnale.

Lo scopo di questo progetto ricerca, i cui primi risultati sono stati presentati al Convegno AISV 2006, è proprio quello di studiare la stabilità di alcuni variabili in una analisi dinamica dei parametri. Nel lavoro appena citato l'attenzione si è soffermata sull'andamento di F0 (frequenza fondamentale) in diversi contesti e in diverse modalità di produzione.

In questa sede, ci occuperemo della seconda fase del progetto di ricerca, cioè lo studio della dimensione temporale nel parlato. Sia l'esperimento precedente sia quello attuale si basano sul corpus di voci intercettate Primula.

(cfr. http://www.linguistica.unical.it/labfon/home_corpus_primula.html)

Così come già accennato in Romito & Lio 2008 riteniamo che una variabile sia portatrice di informazione in ambito di speaker recognition quando:

- a) mostra una alta variabilità interparlatore e una bassa variabilità intraparlato;
- b) è resistente al camuffamento;
- c) ha una alta frequenza di occorrenza;
- d) è robusta durante la trasmissione;
- e) è relativamente facile da identificare e misurare.

L'esperimento approntato, consiste innanzitutto nello stabilire i criteri per la misura dell'*Articulation Rate* (cfr. Romito, Galatà, Lio 2006) e ancor prima l'individuazione e la precisa caratterizzazione del gruppo di respiro; in secondo luogo verranno analizzati segnali prodotti in differenti situazioni come:

- 1) 10 ripetizioni della stessa frase letta in auto, al telefono, in aula, alla fermata dell'autobus, in camera silente;
- 2) lettura di 10 frasi differenti in auto, al telefono, in aula, alla fermata dell'autobus, in camera silente;
- 3) lettura delle frasi di cui al punto 1) e 2) con tono di voce basso, normale e alto;
- 4) parlato spontaneo in luoghi e condizioni differenti come auto, camera silente e lezione universitaria.
- 5) verranno anche comparate registrazioni in ambienti rumorosi rispetto ad ambienti silenziosi.

RIFERIMENTI BIBLIOGRAFICI

- Barlow and Wagner, (1998), "Measuring the dynamic encoding of speaker identity and dialect in prosodic parameters", in R. H. Mannell and J. Robert-Ribes (eds) *Proceedings of the 5th International Conference on Spoken Language Processing*, 30 Nov – 4 Dec, Sydney, Australian Speech Science and Technology Association, pp. 81-4.
- McDougall Kirsty, (2006) "Dynamic features of speech and the characterization of speaker: towards a new approach using formant frequencies", *Speech, Language and the law*, Equinox Publishing Ltd, 13 (1).
- Nolan F., (1997) "Speaker recognition and forensic phonetics", in W.J. Hardcastle and Laver (eds), *The Handbook of Phonetic Sciences*, Cambridge: Cambridge University Press 744-67, 1997
- Romito L., Galata' V., Lio R., (2006), "Fluency Articulation and Speech Rate as new parameters in the Speaker Recognition", in *Atti del convegno III Congreso de Fonética Experimental*, Santiago de Compostela, 26-24 ottobre, 2005, Xunta de Galicia:Santiago de Compostela, pp. 537-549.
- Romito L., Lio R., (2008), "Stabilità dei parametri nello Speaker Recognition: la variabilità intra e interparlatore – La frequenza Fondamentale". *Atti del convegno IV Convegno AISV* (Associazione Italiana Scienze della voce, Unical, Campus di Arcavacata di Rende (CS), 3-5/12, 2007, EDK Editore s.r.l.:RN, Vol. IV.

ENFASI E CONFINI PROSODICI IN DUE STILI DI ELOQUIO EMOZIONALE

Pier Luigi Salza, Enrico Zovato, Morena Danieli
Loquendo SpA, Torino
PierLuigi.Salza@loquendo.com,

SOMMARIO

Introduzione

La nuova frontiera delle tecnologie di sintesi del parlato è la capacità di generare segnali vocali con caratteristiche espressive in grado di variare in modo analogo a quanto avviene nella voce umana. Molti studi hanno evidenziato correlazioni significative tra lo stile di eloquio e le variazioni di alcuni parametri prosodici e spettrali (si vedano, ad es., Scherer [1], Johnstone [2], Magno Caldognetto [3]). Nel presente lavoro si illustra lo studio di alcuni fenomeni di natura ritmica e intonativa, relativamente a due stili espressivi che sono stati riprodotti mediante registrazioni effettuate in laboratorio da parte di parlanti madrelingua inglesi. In particolare, si sono studiati: il *phrasing*, ovvero le strategie utilizzate dai parlanti per suddividere i testi letti in unità tonali (o sintagmi intonativi); i fenomeni di enfasi eventualmente realizzati, intendendo con enfasi quei fenomeni di prominenza (a livello intonativo, di durata e di intensità) tesi ad evidenziare alcune parole nel contesto della frase. A questo scopo si sono annotati, in ciascun enunciato, i confini prosodici e le parole (o i gruppi di parole) prominenti. L'obiettivo di questo lavoro è cercare di individuare le correlazioni tra i fenomeni prosodici presi in considerazione e lo stile di eloquio, in modo da riprodurre, nel sistema di sintesi vocale, analoghi meccanismi per una più accurata caratterizzazione in senso espressivo dei segnali generati. In particolare, si potrebbero introdurre, nel modulo di assegnazione automatica del *phrasing*, regole dipendenti dallo stile e, nel contempo, generare segnali con realizzazioni acustiche dei fenomeni di prominenza, anch'essi legati in modo contestuale allo stile adottato. Gli stili emozionali studiati in questo progetto sono due stili aventi opposta valenza e differenti livelli di attivazione. Prendendo a riferimento modelli basati su categorie, potremmo definire *triste* lo stile a valenza negativa e *allegro* quello a valenza positiva. La scelta di ridurre lo studio a due soli stili è stata dettata da motivazioni pratiche legate anche ai domini di applicazione dei prototipi che sono oggetto di sviluppo nell'ambito del progetto citato. Queste applicazioni mirano alla realizzazione di un assistente virtuale capace di conversare con l'utente in modo *affettivo*, riproducendo, tramite la voce e altre modalità, comportamenti emozionali legati al contesto del dialogo. L'agente deve pertanto essere capace di assumere un atteggiamento positivo o negativo a seconda delle reazioni dell'utente.

Corpus di dati vocali e modalità dell'esperimento

Al fine di ottenere del materiale vocale con due stili espressivi ben caratterizzati, sono stati registrati dei testi opportunamente scelti in modo da indurre nel lettore il corretto atteggiamento emotivo. Un esperto madrelingua inglese ha selezionato, da romanzi e racconti, due gruppi di brani con contenuto rispettivamente triste e allegro. Le registrazioni si sono svolte facendo leggere i brani con continuità e chiedendo ai parlanti di leggere in modo naturale e conforme ai contenuti. Gli stessi brani sono poi stati registrati, in sessioni separate, forzando i parlanti ad adottare uno stile di lettura neutro, a prescindere dal contenuto. Nel complesso sono stati acquisiti dati di tre parlanti, due inglesi americani e un inglese britannico. Le registrazioni sono state eseguite in sala di registrazione e sono stati prodotti segnali campionati a 44.1 kHz con codifica PCM a 16 bit. Per comodità di elaborazione, tutti i testi sono stati suddivisi in frasi di lunghezza variabile, ma non eccedenti le 50 parole. Il corpus, per ciascun locutore, è composto di 36 frasi espressive (14 per lo stile allegro e 22 per lo stile triste) e delle corrispondenti versioni neutre. Da una prima analisi, si è notato, tuttavia, che non tutte le frasi erano caratterizzate in modo espressivo, pertanto sono state condotte delle valutazioni percettive da parte di più soggetti per selezionare ulteriormente il materiale. A questo scopo è stata predisposta un'interfaccia basata su web in cui si richiedeva al soggetto di ascoltare le frasi registrate, neutre e espressive, presentate in ordine casuale, e di valutarne lo stile espressivo sulla base di una scala a sette punti che andavano da *molto triste* a *molto allegro*, con *neutro* in posizione centrale. Si è così ricavato un sottoinsieme di 90 frasi effettivamente espressive.

Elaborazione e annotazione dei dati

Ai fini di questo studio i segnali registrati delle 90 frasi selezionate sono state acquisiti, analizzati e annotati a livello fonetico e prosodico. I dati sono stati intanto elaborati mediante strumenti automatici che hanno fornito la segmentazione in unità fonetiche e il calcolo dei parametri prosodici quali la frequenza fondamentale, la durata e l'intensità sonora. Le trascrizioni fonetiche e i relativi allineamenti temporali sono stati quindi controllati da esperti fonetisti. La localizzazione dei confini prosodici e l'analisi dei fenomeni di enfasi è stata svolta, mediante l'ascolto dei segnali e la visualizzazione delle forme d'onda e delle curve di intensità e di frequenza fondamentale, da un pool di esperti fonetisti e madrelingua. I fenomeni annotati sono stati oggetto di una

preliminare analisi statistica al fine di individuare, nei tre locutori, le correlazioni più significative rispetto agli stili considerati, ed eventuali omogeneità di comportamento tra i parlanti stessi.

Primi risultati e analisi statistica dei dati

Dai dati raccolti e selezionati emerge un'evidente diversità nella frequenza di occorrenza di confini prosodici ed enfasi nei due stili, per ciascun parlante. Le differenze tra stati emotivi e realizzazioni di confini prosodici ed enfasi sono state analizzate con l'analisi della varianza, ANOVA a misure ripetute: 2 (Stato emotivo: Allegro vs Triste) X 2 (Tipo di Fenomeno: Confine vs Enfasi) X 2 (Parlanti: ciascuna coppia di parlanti). Per ciascun parlante sono state scelte a caso dieci tra le frasi annotate per ogni stato emotivo, avendo così a disposizione per i conteggi delle frequenze 30 frasi *allegre* (N=30) e 30 frasi *tristi* (N=30). L'analisi *by subjects* ha fatto emergere un effetto principale significativo dello stato emotivo: la variazione dello stato emotivo si accompagna a una variazione nell'inserimento di confini o enfasi [$F_{allegro(1,30)} = 78.35; p < .001$; $F_{triste(1,30)} = 77.01; p < .005$], senza che emergano differenze significative tra i parlanti [$F_{subj} = 12,35; p < .001$]. Tuttavia, l'analisi rivela anche un'interazione significativa tra Stato Emotivo X Tipo di Fenomeno X Parlante [$F_{(1,30)} = 4.98; p < .05$], suggerendo che i fattori analizzati interagiscano in modo diverso. Per indagare meglio le reciproche interazioni sono stati eseguiti dei confronti pianificati. Tali confronti suggeriscono che l'inserzione di un confine prosodico correla in modo significativo con lo stato emotivo triste, ma non con quello allegro. Questo risultato si riscontra sia nelle analisi *by subjects* [$F_{subj} = 103.61; p < .001$, errore: $F_{subj} = 29,40; p < .001$] che in quelle *by items* [$F_{items} = 12.38; p < .001$, errore: $F_{items} = 6.67; p < .005$]. Viceversa, lo stesso tipo di controllo statistico supporta la significatività della frequenza di enfasi in concomitanza con uno stato emotivo allegro [$F_{subj} = 110.18; p < .001$, errore: $F_{subj} = 30,40; p < 0.001$; $F_{items} = 15.36; p < 0.001$, errore: $F_{items} = 4.85; p < 0.005$].

Osservazioni e futuri sviluppi

Sarà senz'altro interessante confrontare questi dati con analoghe annotazioni effettuate sulle frasi neutre. Lo sviluppo di questo lavoro prevede anche l'analisi dei dati di tipo acustico, in modo da rilevare correlazioni significative tra un eventuale differente uso delle variazioni di intonazione, durata e intensità nella realizzazione dell'enfasi e dei confini prosodici (in tal caso la durata si riferisce alla presenza o meno di pausa), nei due stili di eloquio emotivo. Un fenomeno, infine, che si intreccia, in modo indipendente, con i fenomeni indagati e che comporterà un'analisi a sé, è la velocità di elocuzione, a fronte di un comportamento soggettivo peculiare che pare caratterizzare in modo evidente, riguardo al *tempo*, almeno uno dei tre locutori.

Ringraziamenti

Questa ricerca è stata parzialmente finanziata dalla Commissione Europea nell'ambito del progetto europeo FP6 IST34434 "Companions" [4]. Gli autori sono riconoscenti a Kim Bao Nguyen, Simon Parr e Cristina Segatto.

RIFERIMENTI BIBLIOGRAFICI

- [1] Scherer, K.R. (2003). Vocal communication of emotion: A review of research paradigms, *Speech Communication*, n. 40.
- [2] Johnstone, T. & Scherer, K.R. (1999). The effects of emotions on voice quality, in *Proceedings of XIV Int. Congress on Phonetic Science*.
- [3] Magno Caldognetto E. (2002). I correlati fonetici delle emozioni. In Bazzanella C., Kobau P. (Eds). *Passioni, Emozioni, Affetti*. Milano: McGraw-Hill. 197-214.
- [4] <http://www.companions-project.org>

L'IPOTESI DELLE CLASSI RITMICHE E I DIALETTI ITALO-ROMANZI

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

SOMMARIO

L'idea che le lingue del mondo siano classificabili in base a un parametro ritmico ispira la ricerca in fonetica e fonologia da più di mezzo secolo (v. Bertinetto 1989 per una sintesi). In un primo momento è stata avanzata l'ipotesi che i due tipi fondamentali, di lingue ad isocronia accentuale e ad isocronia sillabica, si rispecchino direttamente nell'organizzazione temporale del parlato: nelle lingue stress-timed i gruppi accentuali si disporrebbero su intervalli di uguale durata, mentre nelle lingue syllable-timed l'unità isocrona verrebbe fornita appunto dalla sillaba. Com'è noto, questa ipotesi forte è stata falsificata da una serie di ricerche empiriche, per cui il concetto di 'isocronia' è stato reinterpretato in chiave fonologica. In questa seconda fase, piuttosto che a una dicotomia netta sul piano temporale si è pensato a un continuum definito da alcune proprietà prototipiche, tra le quali giocano un ruolo di primo ordine le restrizioni fonotattiche.

In anni recenti la cosiddetta 'ipotesi delle classi ritmiche' (Ramus et al. 1999) ha inaugurato una terza fase che cerca di saldare il legame tra la concezione fonetica e quella fonologica, spostando il focus sulle durate degli intervalli vocalici e consonantici nonché sulla loro variabilità. La Rhythm Class Hypothesis ha ricevuto obiezioni di vario genere, ma ha portato anche all'elaborazione di ulteriori algoritmi che tengono conto della velocità di eloquio, quali i Pairwise Variability Indices (Grabe & Low 2002) o i coefficienti di variazione di Dellwo & Wagner (2003). Recentemente Bertinetto & Bertini (2008) hanno infine presentato un modello alternativo che concepisce le differenze ritmiche tra le lingue in termini di 'controllo' e di 'compensazione' delle durate segmentali.

Che i dialetti italo-romanzi si differenzino per le loro caratteristiche ritmiche lo si è sostenuto a varie riprese. In base alle misurazioni delle durate vocaliche, Trumper et al. (1991) collocano le parlate venete e toscane al polo iso-sillabico di un continuum dialettale, assegnando invece dialetti pugliesi al polo iso-accentuale. In sintonia con la reinterpretazione fonologica dell'isocronia, Mayerthaler (1996) delinea un continuum dei dialetti italo-romanzi, muovendo dai processi fonologici che in diacronia hanno 'migliorato' o 'peggiolato' la struttura sillabica. Tale continuum mostra una complessità fonotattica crescente secondo una distribuzione geografica che corrisponde a quattro aree dialettali tradizionali: dall'area meridionale estrema, i cui dialetti presentano delle sillabe relativamente semplici, si procede attraverso due aree di tipo 'misto' (i dialetti alto-meridionali e mediani) verso l'area gallo-italica costituita da dialetti con strutture sillabiche molto marcate.

Con un approccio complementare, Schmid (1997) elenca gli inventari dei tipi sillabici (CV, CVC, ecc.) di sei varietà italo-romanze, ordinandole di nuovo in un continuum che muove da un polo tendenzialmente iso-sillabico (siciliano, veneziano, italiano) a un polo tendenzialmente iso-accentuale (feltrino, romagnolo, friulano). In un lavoro successivo Schmid (2004) amplia l'indagine alla considerazione di altri sei dialetti, confrontando gli inventari dei tipi sillabici con un'analisi secondo il modello di Ramus et al. (1999). Dall'osservazione dei diagrammi cartesiani ottenuti in base alle misurazioni di dieci enunciati per ogni varietà si evince che i dialetti con una maggiore complessità sillabica (milanese, torinese, bitontino) sono posizionati verso l'area che potremmo denominare 'iso-accentuale', mentre i dialetti con una struttura fonotattica più semplice (napoletano, veneziano, pisano) si orientano piuttosto verso il polo iso-sillabico. In un'analisi recente degli stessi dati secondo il modello di Grabe & Low (2002), il napoletano si stacca invece dal gruppo iso-sillabico per entrare nell'area iso-accentuale (Schmid 2008).

Allo stato attuale risulta quindi che i dialetti italo-romanzi esibiscono notevoli differenze ritmiche, che non si lasciano tuttavia rappresentare come continuum areale. La composizione dei gruppi dialettali varia – almeno in parte – secondo l'algoritmo applicato, dato che i modelli colgono aspetti diversi dell'organizzazione temporale. Inoltre va sottolineato che il ritmo è determinato non soltanto dalle strutture fonotattiche, ma anche dalle regole allofoniche che operano in superficie.

Con il presente contributo si intende offrire un bilancio dello stato della ricerca sulla tipologia ritmica dei dialetti italo-romanzi, e ci si propone di allargare il quadro dal punto di vista sia empirico che metodologico. A tal fine si analizzano dati da altri dialetti, in particolare dal feltrino e dal friulano (nonché dai due dialetti ritenuti 'estremi', ovvero dal siciliano sul versante iso-sillabico e dal romagnolo sul versante iso-accentuale). Inoltre, tutti i dati vengono sottoposti a ulteriori calcoli, applicando anche gli algoritmi Varco (Dellwo & Wagner 2003) e CCI (Bertinetto & Bertini 2008).

RIFERIMENTI BIBLIOGRAFICI

- Bertinetto, P.M. (1989). Reflections on the dichotomy 'stress' vs. 'syllable-timing', *Revue de Phonétique Appliquée*, 91-92-93, 99-130.
- Bertinetto, P.M. & Bertini, C. (2008). On modeling the rhythm of natural languages, in *Proceedings of the 4th International Conference on Speech Prosody* (P. Barbosa et al., editors), São Paulo: Capes, 427-430.
- Dellwo, V. & Wagner, P. (2003). Relations between language rhythm and speech rate, in *Proceedings of the 15th International Congress of Phonetics Sciences* (M.J. Solé, D. Recasens & J. Romero, editors), Barcelona: Futurgraphic, 471-474.
- Grabe, E. & Low, E.L. (2002). Durational variability in speech and the rhythm class hypothesis, in *Papers in Laboratory Phonology 7* (C. Gussenhoven & N. Warner, editors), Berlin: Mouton de Gruyter, 515-546.
- Mayerthaler, E. (1996). Stress, syllables, and segments: their interplay in an Italian dialect continuum, in *Natural Phonology: The State of the Art* (B. Hurch & R. Rhodes, editors), Berlin: Mouton de Gruyter, 201-221.
- Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal, *Cognition*, 73, 265-292.
- Schmid, S. (1997). A typological view of syllable structure in some Italian dialects, in *Certamen Phonologicum III* (P.M. Bertinetto, L. Gaeta, G. Jetchov & D. Michaels, editors), Torino: Rosenberg & Sellier, 247-265.
- Schmid, S. (2004). Une approche phonétique de l'isochronie dans quelques dialectes italo-romans, in *Nouveaux départs en phonologie. Les conceptions sub- et suprasegmentales* (T. Meisenburg & M. Selig, editors), Tübingen: Gunter Narr, 89-109.
- Schmid, S. (2008). Measuring the rhythm of Italian dialects (Poster presented at the Workshop on Empirical Approaches to Speech Rhythm, University College London, 28th March 2008).
- Trumper, J., Romito, L. & Maddalon, M. (1991). Double consonants, isochrony and raddoppiamento fonosintattico: some reflections, in *Certamen Phonologicum II* (P.M. Bertinetto, M. Kenstowicz, & M. Loporcaro, editors), Torino: Rosenberg & Sellier, 329-360.

DOES A TALKER'S OWN RATE OF SPEECH AFFECT HIS/HER PERCEPTION OF OTHERS' SPEECH RATE?

Sandra Schwab
Université de Genève
sandra.schwab@unige.ch

SUMMARY

Speech rate –determined by articulation rate and by number and duration of pauses (see Grojean & Deschamps, 1975 for a detailed description)– has been widely studied from various points of view for the past 50 years. Among the numerous studies, many have dealt with speech rate perception. Research in this field has shown that the subjective rate estimation grows more quickly than the objective physical measurements, and that it rises in a non-linear way (Lane & Grosjean, 1973). Speech rate perception can indeed be described by Stevens' power function law (Cartwright & Lass, 1975), which assumes that sensation is proportional to the physical intensity raised to a given power (Stevens, 1957).

Many studies have investigated the factors that may affect the perception of speech rate (e.g. Grosjean & Lane, 1976; Feldstein & Bond, 1981; Kohler, 1986; Greene, 1987; Crown & Feldstein, 1991), but very few studies have examined the role that the talker's own rate might play in his/her perception of others' rate. For example, Lass & Cain (1972) investigated the hypothesis that a speaker's preferred speech rate depended on his actual speech rate. They showed a good correlation ($r = 0.61$) between speakers' preferred and actual speech rates: speakers who produced slow speech rates preferred listening to slow speech rates, whereas fast speakers tended to prefer fast speech rates. This conclusion raises the question whether speech rate production affects not only speech rate preference, but also speech rate perception. To our knowledge, very few studies have tried to answer this question. Gósy (1991) formulated the hypothesis that "*the speaker's own speech tempo determines his judgements concerning that of other people: the faster his own speech the less fast he perceives that of others*" (p. 101). Gósy showed that speakers with different speech rates (very slow, slow, moderate, fast, very fast) did not perceive speech rate in a similar way. In the same direction, Koreman (2006) hypothesized that listeners' own speaking habits may affect their perception of speech rate. Nevertheless, his results failed to show an effect of the listener's rate on his/her perception of rate.

In sum, considering the lack of totally conclusive results on the role that the talker's rate might play in rate perception, the objective of this research is to explore more deeply the hypothesis that speakers with different speech rates do not perceive speech rate in a similar way. To this end, we conducted both production and perception experiments.

In the production experiment, forty French speaking participants were asked to read a passage at their usual (normal) rate, as well as at fast and slow rates. For each participant at each rate, we measured the total, articulation and pause time, and obtained speech and articulation rates. According to the speech rate distribution, a representative subset of twenty-eight speakers was selected to participate in the perception experiment. In this experiment, participants were asked to listen to and estimate various samples at different speech rates (normal, fast and slow), using a magnitude-estimation task (Stevens, 1957).

Results firstly showed a negative correlation between rate estimation and own rate at normal and slow rates (respectively, $r = -0.45$, $r = -0.39$, $p < 0.05$), but no correlation at fast rate ($r = -0.11$, ns): speakers with slow speech rate tended to overestimate the sample speech rates in comparison with fast speakers (at normal and slow rates). Secondly, and more interestingly, a regression analysis revealed that the own rate has a moderator effect on rate estimation, at all rates (normal: $t(781) = -5.67$, $p < 0.001$; fast: $t(781) = -2.06$, $p < 0.05$; slow: $t(781) = -6.46$, $p < 0.001$): the faster a listener speaks, the less his/her rate estimations raise as a function of heard rates, especially at normal and slow rates.

In sum, the present research has shown that a relationship does exist between speech rate production and perception, the former determining the latter: slow speakers tend to overestimate rates in comparison with fast speakers. Nevertheless, a question that may arise from these results concerns the direction of the relationship. At the segmental level, the direction of the link between production and perception has been considered in both ways. Indeed, following Perkell et al. (2004)'s hypothesis, speech perception affects speech production, while according to other researchers (Paliwal, Lindsay, Ainsworth, 1983) and defenders of the *Motor Theory of Speech Perception* (Liberman & Mattingly, 1985), speech production regulates speech perception. Following Gósy (1991) and Koreman (2006), we hypothesized that rate production regulates rate perception, but it would be worth considering the reverse possible interpretation.

REFERENCES

- Cartwright, L.R. & Lass, N.J. (1975). A psychophysical study of rate of continuous speech stimuli by means of direct magnitude estimation scaling. *Language and Speech*, 18, 358-365.
- Crown, C.L. & Feldstein, S. (1991). The perception of speech rate from the sound-silence patterns of monologues. *Journal of Psycholinguistic Research*, 20, 47-63.
- Feldstein, S. & Bond, R. (1981). Perception of speech rate as a function of vocal intensity and frequency. *Language and Speech*, 24, 387-395.
- Gósy, M. (1991). The perception of tempo. In M. Gósy (Ed.). *Temporal Factors in Speech. A collection of Papers* (pp. 63-107). Budapest: Research Institute for Linguistics, HAS.
- Green, K.P. (1987). The perception of speaking rate using visual information from a talker's face. *Perception & Psychophysics*, 42, 587-593.
- Grosjean, F. & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français: Vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica*, 31, 144-184.
- Grosjean, F. & Lane, H. (1976). How the listener integrates the components of speaking rate. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 538-543.
- Kohler, K.J. (1986). Parameters of speech rate perception in German words and sentences: Duration, F0 movement, and F0 level. *Language and Speech*, 49, 115-139.
- Koreman, J. (2006). Perceived speech rate: The effect of articulation rate and speaking style in spontaneous speech. *Journal of the Acoustical Society of America*, 119, 582-596.
- Lane, H. & Grosjean, F. (1973). Perception of reading rate by speakers and listeners. *Journal of Experimental Psychology*, 97, 141-147.
- Lass, N.J. & Cain, C.J. (1972). A correlational study of listening rate preferences and listeners' oral reading rates. *Journal of Auditory Research*, 12, 308-312.
- Liberman, A.M. & Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.
- Paliwal, K.K., Lindsay, D. & Ainsworth, W.A. (1983). Correlation between production and perception of English vowels. *Journal of Phonetics*, 11, 77-83.
- Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Stockmann, E., Tiede, M. & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *Journal of the Acoustical Society of America*, 116, 2338-2344.
- Stevens, S.S. (1957). On the psychophysical law. *Psychological Review*, 64, 153-181.

PERCEZIONE E PRODUZIONE DEI FONEMI DELL'INGLESE AMERICANO IN PARLANTI CON UN SISTEMA PENTAVOCALICO

Bianca Sisinni, Mirko Grimaldi
Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL) – Università del Salento
grimaldi.mirko@alice.it

SOMMARIO

Con questo lavoro ci proponiamo di studiare i processi di produzione e percezione durante l'acquisizione dei fonemi vocalici dell'Inglese Americano (AE) in un gruppo di studenti universitari della Facoltà di Lingue dell'Università del Salento (parlanti nativi dell'Italiano Salentino (IS) con un sistema a 5 vocali e tre gradi di apertura).

In letteratura è ancora dibattuta la questione se lo sviluppo percettivo della L2 preceda lo sviluppo della produzione, oppure se una idonea percezione non sia condizione necessaria per una corretta produzione (revisioni in Listerri 1995; Leather 1999; Escudero 2005; Hansen Edwards-Zampini 2008). L'idea prevalente ritiene che per spiegare appieno l'acquisizione fonologica di una L2 bisogna prima spiegare il modo in cui i parlanti della L2 riescono a sviluppare una percezione appropriata e quindi una rappresentazione cognitiva dei segmenti della L2: la produzione corretta sarebbe una diretta conseguenza della corretta rappresentazione astratta.

In quest'ottica abbiamo incrociato gli strumenti di analisi dello *Speech Learning Model* (SLM) di Flege (1995) e del *Perceptual Assimilation Model* (PAM) di Best (1995), anche per capire se gli studenti di Lingue risultano assimilabili alla categoria dei *naïve L2 listeners* oppure a quella dei *fluent L2 listeners* (Best & Tyler 2006).

Sono state prese in esame 18 studentesse universitarie (SU) – età media 20,4 anni, con alle spalle una media di 10 anni di studio scolastico dell'inglese anche con insegnanti madrelingua – al secondo mese del corso di inglese tenuto da un lettore madrelingua USA.

Delle SU, tramite lettura di frasi-cornice (*Dico CVCCV adesso*, sei ripetizioni per fonema), sono state rilevate le aree di esistenza (proiettate su assi cartesiani) della L1. Allo stesso modo sono state rilevate le aree di esistenza di tre parlanti nativi (NS) di sesso femminile dell'AE (frase cornice: *I say CVC now*), come rappresentato nella Fig. 1 (la struttura sillabica della parola bersaglio nelle frasi cornice dell'IS è giustificata dalla necessità di simulare la struttura sillabica della L2). Si è quindi proceduto a rilevare 11 fonemi dell'AE (/i:/, /ɪ/, /ε/, /æ/, /ɜ:/, /ʌ/, /ɑ:/, /ɑ:/, /ɔ:/, /ʊ/, /u:/: Ladefoged 2005) prodotti dalle SU, utilizzando il metodo della *delayed repetition* (Flege et al. 2003, Piske et alii 2001), in cui un NS produce una frase cornice che le SU devono ripetere dopo l'ascolto di elementi distrattori e *interventing speech material* (garantendo così una ripetizione del fonema bersaglio distanziata dall'ascolto del fonema nativo), come in (1), cf. Fig. 2:

(1) NS : *I say CVC now* (FRASE-CORNICE). *Could you repeat, please?* (DISTRATTORE)
(PAUSA – BIP – PAUSA – BIP) (DISTRATTORI)

SU : *Of course I could* (INTERV. SPEECH MAT). *I say CVC now*. (FRASE-CORNICE)

F0, struttura delle frasi-cornice di L1 ed L2, contesti consonantici, qualità delle sillabe sono stati mantenuti il più possibile costanti, per un totale di 54 stimoli per il corpus di IS e 66 stimoli per il corpus di AE. Le registrazioni sono state effettuate con il sistema CSL 4500 (frequenza di campionamento di 22.05 Khz). La segmentazione è stata eseguita con Praat 4.6.29, calcolando per ogni fonema la durata totale (in secondi) e le F0, F1, F2 e F3 nel punto medio (0,02 s).

La capacità di percezione è stata invece testata attraverso la somministrazione di due classici test percettivi: l'*identification test* (IT) e l'*oddity discrimination test* (ODT) (Flege & MacKey 2004, Tsukada et al. 2005). Gli stimoli per questi test sono stati ricavati dalle produzioni delle NS per un totale di 36 stimoli uditivi. Le parole sono state normalizzate in intensità e segmentate con Praat 4.6.29. Le SU sono state testate singolarmente in una stanza insonorizzata e gli stimoli sono stati somministrati tramite PC. L'ODT è stato fatto eseguire anche ad un gruppo di 10 parlanti native dell'AE per ottenere dati relativi a un gruppo di controllo.

I risultati dei due test percettivi – che forniscono da un lato le modalità di assimilazione dei fonemi dell'AE ai fonemi dell'IS e dall'altro la capacità di discriminazione dei contrasti fonologici individuati – sono stati analizzati alla luce della griglia interpretativa proposta dal modello PAM. In base al modo con cui i fonemi non nativi vengono assimilati ai fonemi nativi, il PAM prevede sei diversi tipi di assimilazione di coppie non native ai fonemi nativi. Per ogni livello è possibile predire una determinata capacità di discriminazione da parte dei *naïve L2 listeners*, che può spaziare da un grado molto basso al grado di *native-like*.

Come notiamo in Fig. 3, i risultati dell'ODT evidenziano che il gruppo di SU ha raggiunto differenti livelli di accuratezza nel processo di discriminazione. Dato che questi livelli di discriminazione individuati trovano una interpretazione coerente all'interno del PAM, possiamo concludere che questo modello può essere applicato

anche delle SU, che paiono comportarsi percettivamente come *naïve L2 listeners*, nonostante un lungo background di L2 (sebbene esclusivamente scolastico).

Da un punto di vista produttivo, sulla base delle analisi statistiche condotte sui dati acustici (che verranno meglio illustrate in sede di presentazione) si può ipotizzare che i soggetti analizzati si trovino in una fase di *interlingua* in cui la capacità di realizzare i goal articolari previsti dal sistema della L2 risulta ancora mediata dalle caratteristiche implicite nella L1, anche se ciò non esclude la possibilità di realizzare fonemi come /ʊ/ ed /ʌ/ con valori formantici vicini a quelli dei NS (cf. Fig. 2).

Infine il confronto dei dati in produzione con quelli in percezione – e in particolare con i dati ottenuti nell'IT – pare supportare l'idea che una corretta rappresentazione astratta preceda la capacità di produrre coerentemente i fonemi di una L2 (Flege et alii, 1999; Bion et alii, 2006). Infatti, la produzione 'corretta' da parte delle SU è avvenuta esclusivamente per due fonemi di L2, /ʊ/ ed /ʌ/, facenti parte di un set più ampio di fonemi dell'AE percepiti come molto dissimili dai fonemi di L1. Questo potrebbe far supporre uno stadio iniziale in cui i parlanti sviluppano prima la capacità di discriminare percettivamente alcuni fonemi di L2 rispetto ai fonemi della L1, per poi iniziare gradualmente ad estrarre le informazioni articolatorie necessarie alla produzione di alcuni dei fonemi presenti nel campo percettivo discriminato. Fatto questo che preluderebbe ad una successiva estrazione delle ulteriori informazioni articolatorie per la produzione dei restanti fonemi.

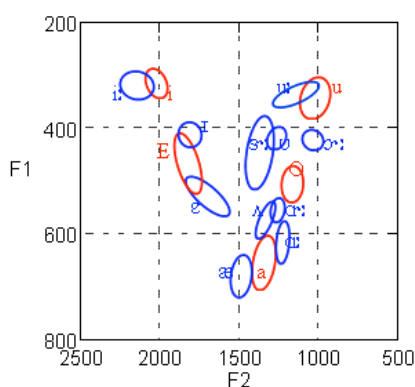


Figura 1: I fonemi dell'AE (blu) e dell'IS (rosso) (blu)

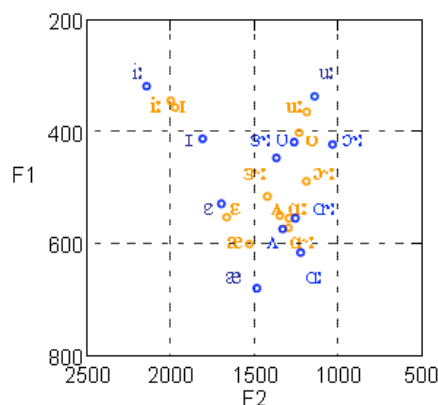


Figura 2: I fonemi di L2 delle SU (arancio) e dell'AE (blu)

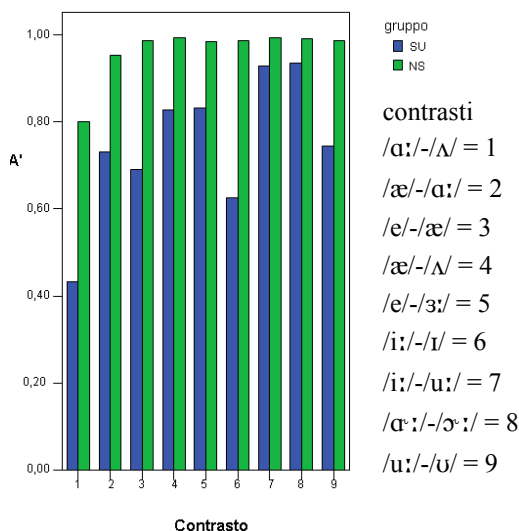


Figura 3: risultati ODT delle SU (blu) e delle 10 NS di controllo (verde)

STATICO VS DINAMICO, UN POSSIBILE RUOLO DELLA SILLABA NEL RICONOSCIMENTO AUTOMATICO DEL PARLATO

Serena Soldo, Bogdan Ludusan

Dipartimento di Scienze Fisiche, NLP-Group, Università degli Studi di Napoli "Federico II"

soldo@na.infn.it, ludusan@na.infn.it

SOMMARIO

Il continuum fonico su cui un sistema automatico di riconoscimento deve lavorare viene normalmente segmentato in piccole porzioni sulle quali algoritmi basati su tecniche statistiche operano sia per l'identificazione dell'informazione linguistica in essi contenuta, sia per ricostruire a posteriori il contenuto complessivo dell'enunciato contenuto nel segnale acustico. Mentre tradizionalmente fino a pochi anni fa le dimensioni della porzione minima di analisi si aggirava intorno a dimensioni che linguisticamente potremmo definire subfoniche, sempre più spesso, ormai, i sistemi di riconoscimento del parlato fanno uso di analisi di segmenti di parlato superiori ai 150-200 ms. Questa tendenza indica probabilmente l'uso di parametri prosodici oltre che segmentali. Fra i parametri per la descrizione di segmenti lunghi che è possibile usare si incontrano quelli legati a proprietà ritmiche del parlato. Recentemente sono stati portati avanti lavori per dimostrare che tali parametri possono essere estratti automaticamente con algoritmi indipendenti dalla lingua ([1] [2] [3]).

Sebbene la definizione 'classica' di sillaba (ma i linguisti sanno bene quanto trovare una definizione condivisa da tutti sia difficile) solitamente utilizzata in letteratura tende a mettere in evidenza le caratteristiche dinamiche del segnale vocale come ad esempio la coarticolazione, in questo lavoro si è cercato di prendere in considerazione una ipotesi alternativa. L'idea è quella di vedere la sillaba come una rappresentazione statica di un pezzo di parlato, una sorta di istantanea che contenga in sé unitariamente informazione che solitamente si ritiene di tipo tempo-variabile, che si estende su un determinato intervallo di tempo. Alla luce di questo tipo di rappresentazione, la variabile indipendente rispetto alla quale i fenomeni che osserviamo evolvono e sulla quale basare un sistema di riconoscimento del parlato non risulterà più essere il tempo ma la sequenza di unità sillabiche. Supponendo di essere in grado di individuare con precisione gli estremi dell'intervallo su cui si estende ciascuna sillaba, si può dunque pensare di "fotografarla" estraendone le caratteristiche nei punti salienti.

È da osservare che questo genere di rappresentazione della sillaba è completamente originale e mai proposto in letteratura. Lo scopo di questo lavoro è proprio quello di capire se si tratta di una tecnica in grado di fornire buoni risultati e, eventualmente, di evidenziarne i punti deboli.

Come è noto, una sillaba è costituita da almeno una vocale (che ne costituisce il nucleo) e può al massimo essere formata da tre parti, il nucleo vocalico testé definito, la testa e la coda.

Il primo passo del nostro lavoro è consistito nella trasformazione di ogni porzione di segnale corrispondente ad una sillaba in un set di parametri (d'ora in poi *features*) da dare in ingresso ad un sistema di riconoscimento. Per rappresentare ciascuna sillaba abbiamo quindi scelto di concentrare l'estrazione delle *features* solo sul centro di ciascuna delle tre parti. In particolare sono stati estratti tre vettori di parametri per la testa, tre per il nucleo e tre per la coda. Si è scelto di utilizzare per la rappresentazione i 13 coefficienti MFCC (*Mel Frequency Cepstral Coefficients*). Ciascuna sillaba, alla luce di queste scelte, risulta essere rappresentata da una matrice di dimensioni 9x13.

Per valutare l'efficacia di questa rappresentazione, si è deciso di provare a classificare le sillabe così rappresentate tramite un classificatore lineare SVM (Support Vector Machines). Una SVM è un classificatore supervisionato che prende in input un vettore di *features* (viste come un punto in piano n-dimensionale) e ne restituisce la classe di appartenenza sulla base del margine geometrico tra questo punto e le classi astratte corrispondenti alle unità da riconoscere (viste come degli iperpiani) individuate durante la fase di addestramento.

Il corpus utilizzato per questo lavoro è una parte del corpus SPEECON, sequenze foniche contenenti la lettura di numeri in lingua italiana tra 0 e 999999 pronunciati da soli speaker maschi. I motivi per cui si è scelto un corpus con tali caratteristiche sono vari: contiene un numero di parole non troppo grande, per evitare tempi computazionali eccessivi, e non troppo piccolo, per mantenere le caratteristiche generiche di un corpus con un dizionario più ampio; con l'uso di poche sillabe è possibile costruire parole abbastanza lunghe e con le stesse caratteristiche di variabilità di pronuncia presenti in altri corpus. In base alla divisione sillabica fonologica, le parole contenute nei file del corpus sono state suddivise in 8631 sillabe distinte che vanno a coprire l'intero insieme di 42 classi di sillabe presenti nel dizionario. Inoltre è stata aggiunta una classe per identificare i segmenti considerati silenzio o rumore.

Abbiamo quindi addestrato una SVM multiclasse sulla base di un training set estratto dal corpus, e poi testato l'efficacia della nostra rappresentazione sul test set. Le prestazioni ottenute dalla classificazione delle sillabe rappresentate con questo primo approccio è pari all'85% circa.

L'analisi degli errori della classificazione ha evidenziato che la scelta di soli nove frames per la rappresentazione di qualsiasi sillaba non è forse la più adatta. Quindi una delle prime ipotesi prese in

considerazione è stata quella di cambiare il numero di frames utilizzati. Le possibili tecniche da utilizzare a questo scopo sono due: il numero di frames considerati per ciascun segmento può variare al variare della dimensione del segmento stesso; oppure il numero di frames può essere fissato a priori ma con un valore più alto, scegliendo una maggiore o minore sovrapposizione delle finestre di avanzamento durante l'estrazione delle features in modo da adattarsi a ciascun segmento. Nella realizzazione della prima tecnica la sovrapposizione tra le finestre durante l'estrazione delle features è stata mantenuta fissa e quindi per ogni segmento sono stati estratti un numero di vettori variabile in base alla lunghezza del segmento stesso. Per questa tecnica sono state fatte delle prove considerando finestre di 128 e 256 campioni (corrispondenti a circa 8 e 16 msec), ma entrambe hanno prodotto risultati molto scarsi; per tale motivo questo sistema è stato accantonato in favore della seconda tecnica, molto più promettente. La seconda tecnica invece mira ad ottenere un numero di frames fisso da ciascun segmento; questo è stato ottenuto fissando l'ampiezza della finestra (nel nostro caso 256 campioni, corrispondenti a 16 msec) e variando l'ampiezza della sovrapposizione opportunamente. I risultati ottenuti dalla classificazione delle sillabe rappresentate con questa nuova tecnica sono stati decisamente migliori. In particolare, abbiamo fatto variare il numero di frames per ogni sillaba tra 9 e 21; all'aumentare del numero di frames le prestazioni sono migliorate fino al 87.2% (in corrispondenza di 19 frames per sillaba). Aumentando ulteriormente questo valore non si sono verificati ulteriori miglioramenti, per cui abbiamo fissato a 19 il numero di frames ideale per questo genere di rappresentazione.

Un altro dato importante è stato ottenuto valutando gli N-best della classificazione. Estrahendo dalla SVM le percentuali di appartenenza dei segmenti a ciascuna classe possibile abbiamo notato che nel 96% dei casi la classe giusta rientra tra le prime 3 più probabili e nel 99% dei casi essa è nelle prime 10.

In conclusione questo lavoro mira ad indagare nuove tecniche di estrazione delle features per la rappresentazione delle sillabe. Abbiamo mostrato come un approccio teso a estrarre informazioni sulle caratteristiche statiche del segnale, piuttosto che quelle dinamiche, può fornire buoni risultati.

RIFERIMENTI BIBLIOGRAFICI

- [1] F.Tamburini, C.Caini. *An Automatic System for Detecting Prosodic Prominence in American English Continuous Speech*, «INTERNATIONAL JOURNAL OF SPEECH TECHNOLOGY», 2005, 8, pp. 33 – 44
- [2] Petrillo M. *Algoritmi per la divisione del segnale verbale in unità sillabiche*. 2000, Tesi di Laurea presso l'Università degli Studi Di Napoli "Federico II".
- [3] Ludusan B., Soldo S., *Sonority based syllable segmentation*, comunicazione in questa conferenza.

ALCUNE NOTE SULLE OPPOSIZIONI DI QUANTITÀ VOCALICA

Arianna Uguzzoni
Alma Mater Studiorum – Università di Bologna
arianna.uguzzoni@unibo.it

SOMMARIO

Negli ultimi tempi si sono intensificati studi e dibattiti riguardo alla caratterizzazione e alla classificazione delle lingue che fanno un uso distintivo delle differenze di durata (Becker, 1998; Auer et al, 2002). A scopo di maggiore chiarezza in questa sede mi limito ad alcuni aspetti e problemi connessi specificamente con la quantità vocalica di alcune lingue d'Europa.

Entrando in medias res, una linea demarcativa, su cui non c'è contrasto fra gli studiosi, consiste nella relazione di indipendenza o di dipendenza tra le opposizioni di quantità e l'accento lessicale.

Il finlandese è un esempio tipico di lingua europea odierna in cui chiaramente vige l'indipendenza dall'accento: vocali brevi e vocali lunghe sono usate tanto in sillaba accentata quanto in sillaba non accentata (a).

Del tutto diversa è la situazione delle lingue in cui le opposizioni di quantità vocalica sono rigorosamente subordinate all'accento lessicale. Nella maggior parte delle lingue europee che distinguono fonologicamente vocali brevi e vocali lunghe vige la restrizione per la quale ciò è limitato alle sillabe accentate: sia in area germanica (es. tedesco, danese), sia in area romanza (es. parlate franco-provenzali, parlate italo-romanze) (b).

Considerando altri fattori linguistici è possibile enucleare ulteriori dicotomie. In alcuni casi esse da un lato rafforzano la contrapposizione tra (a) e (b), dall'altro operano all'interno delle lingue esemplificate in (b). Mi riferisco alla proprietà distribuzionale riguardante le sillabe accentate aperte: in finlandese e in danese si trovano vocali sia brevi sia lunghe, mentre in tedesco sono ammesse soltanto vocali lunghe.

La differenza tipologica tra danese e tedesco risulta corroborata se a ciò che si è appena visto, si aggiungono due considerazioni. (1) il danese presenta vocali accentate brevi in finale assoluta, cosa che invece è esclusa dalla fonotassi tedesca. (2) si ipotizza che in parole con struttura /^hcvcv/ la sillabazione del danese contemporaneo sia /^hcv-cv/, e sia invece /^hcvc-v/ in tedesco: qui la consonante interna aderisce strettamente alla vocale breve precedente e rende chiusa la sillaba accentata (Martinet, 1969).

L'utilizzazione opposizionale della durata vocalica assume modalità diverse, come mostrano le condizioni linguistiche che ho presentato sopra a scopo esemplificativo. E' quindi legittimo e auspicabile approfondire la ricerca delle differenze sistematiche che consentono di individuare tipi, sottotipi, intertipi. In questa cornice potrebbero trovare una collocazione plausibile anche le lingue romanze citate (Uguzzoni et al., 2003).

In base alle mie conoscenze sono prevalse finora posizioni radicali che, in merito alle opposizioni di quantità vocalica, propongono una differenziazione tra due sole categorie: "quantity contrasts" e "syllable-cut contrasts". Pertanto le lingue vengono classificate o come "quantity languages" o come "syllable-cut languages" ("Silbenschnittsprachen").

Sono considerate lingue con opposizioni di taglio sillabico il tedesco (precisamente il tedesco settentrionale), l'olandese, l'inglese, nelle quali è cruciale la differenza tra i modi in cui si legano e interagiscono la vocale accentata e la consonante successiva: il taglio brusco ("abrupt cut", "scharfer Schnitt") da una parte e il taglio piano ("smooth cut", "sanfter Schnitt") dall'altra. Secondo alcuni in questo quadro è decisamente centrale la presenza della consonante successiva alla vocale centrale (Becker, 1998).

E la durata vocalica? In questa prospettiva, che è decisamente prosodica, alla durata segmentale viene attribuito uno status subordinato alla opposizione di taglio sillabico: brevità e lunghezza delle vocali accentate sono interpretate come fenomeni collaterali e semplici concomitanti fonetici dei due modi di taglio sillabico (Uguzzoni, 2002).

Un'altra via di ricerca che meriterebbe di essere percorsa in maniera sistematica e approfondita riguarda lo studio dei processi evolutivi. Come si sono formate le proprietà che in parte suddividono in parte accomunano le lingue odierne dal punto di vista delle modalità dell'uso linguistico delle differenze di durata vocalica?

Per l'Europa ci sono interessanti premesse per ricostruire alcune tappe significative della storia linguistica che è approdata allo stato attuale. Penso in particolare a studi diacronici su lingue germaniche e su lingue romanze: essi potrebbero gettare luce su problemi che non sono ancora pienamente risolti restando soltanto sul piano sincronico.

RIFERIMENTI BIBLIOGRAFICI

- Auer, P.; Gilles, P.; Spiekermann, H. (a c. d.), 2002. *Silbenschnitt und Tonakzente*. Tübingen: Niemeyer.
- Becker, T., 1998. *Das Vokalsystem der deutschen Standardsprache*. Frankfurt: Peter Lang.
- Martinet, A., 1969. Coupe ferme et coupe lâche. In P.Valentin e G. Zinke (a c. d.), *Mélanges pour Jean Fourquet*, Paris: Klincksieck, 221-226.
- Uguzzoni, A., 2002. Fester vs. loser Anschluss. Appunti per una storia di un concetto secolare. *Lingue e Linguaggio*, 1, 327-340.
- Uguzzoni, A.; Azzaro, G.; Schmid, S., 2003. Short vs long and/or abruptly vs smoothly cut vowels. New perspectives on a debated question. In *Proceedings of the XV International Congress of Phonetic Sciences*, Barcelona, 3, 2717-2220.

THE ANTICIPATORY PERCEPTION BASED ON EVENTS (*APE*) HYPOTHESIS

¹Béatrice Vaxelaire, ¹Rudolph Sock, ¹Fabrice Hirsch, ¹Véronique Ferbach-Hecker, ²Johanna-Pascale Roy,
¹Fayssal Bouarourou

¹Université de Strasbourg – UdS, Institut de Phonétique de Strasbourg (IPS) & Composante Parole et
Cognition – E.A. 1339 Linguistique, Langues et Parole – LiLPa; ²Université Laval Québec, Laboratoire de phonétique,
Département de langues, linguistique et traduction
Beatrice.Vaxelaire@umb.u-strasbg.fr, sock@umb.u-strasbourg.fr, fabrice_hirsch@yahoo.fr

ABSTRACT

This abstract summarises sensory-motor timing results obtained by our team on anticipatory phenomena in speech production and perception. The *APE* Hypothesis proposed here is based on *X-ray* (see, e.g., Vaxelaire *et al.*, 1999 a, b ; Roy *et al.*, 2003 ; Vaxelaire *et al.*, 2007), *kinematic* (see, e.g., Hardcastle *et al.*, 1996 ; Hecker *et al.*, 2001 ; Sock & Vaxelaire, 2004) and *acoustic data* (Hirsch *et al.*, 2003). Although data for consonantal anticipatory sensory-motor relations have been studied (see, e.g., Vaxelaire *et al.*, 2003), the focus here will be on anticipatory vocalic gestures, together with their visual and auditory effects. In general, sensory-motor anticipatory relations are examined within Vowel¹-Consonant(s)-Vowel² sequences (V¹C⁽ⁿ⁾V²), where V¹ is an unrounded vowel, C⁽ⁿ⁾ is one or more consonants and V² is always a protruded vowel.

Globally, results show that anticipatory visual identification of the rounded vowel systematically precedes its auditory identification. There is no auditory identification before release of a stop consonant. Such identification takes place at release, which coincides more or less with peak velocity. However, auditory vowel identification is robust only *after* peak velocity.

The situation is somewhat different when the obstruent interval is one or more fricatives, or a sequence of plosives including one or more fricatives.

Auditory identification takes place at peak acceleration, which coincides more or less with a steep slope in the inferior limit of the fricative noise, in its trajectory towards F3 of the protruded vowel. However, robust identification of the vowel happens only *after* peak acceleration.

Our results show that extension of anticipatory gestures may increase as the obstruent interval increases, thus confirming predictions of some aspects of the Movement Expansion Model (Abry & Lallouache, 1995).

The fricative context favours protrusion anticipatory expansion, together with its visual (Roy, 2004) and auditory efficiencies (Hecker *et al.*, 2001). We posit therefore that it is more appropriate to treat plosives separately from fricatives.

Consonantal anticipatory gestures may emerge in the preceding vowel, in a CVC^{+lab} sequence. In such cases, their efficiency would be visual but not auditory.

Data in absolute values are observed, but in the final analysis their exploitation is limited, due mainly to the temporal elasticity of speech signals (Gaitenby, 1965).

Consequently, the procedure adopted in order to analyse anticipatory phenomena is the following: Articulatory, kinematic and acoustic anticipatory events which underlie the perceptual efficiency of the upcoming vowel are detected. Then, the extension of the anticipatory gesture, together with its sensory efficiencies, are quantified by calculating the *percentage of time* taken by the interval which separates the decisive event and the acoustic emergence of the target vowel (onset of a stable formant structure), within the obstruent phase. Such a normalisation procedure allows reasoning in relative terms, thus taking into account potential compressions and expansions of speech signals, which determine the advent of the various articulatory, kinematic and acoustic events.

Constraints elaborated for anticipatory gestures, in a sensory-motor perspective, thus take into account, as a priority, the advent of articulatory, kinematic and acoustic events, which appear on the anticipatory phase of the gesture, hence providing landmarks to rationalise the sensory extension of the anticipatory gesture. The auditory and/or visual efficiencies of specific articulatory, motor and acoustic anticipatory events allow reinforcing the idea of a tight coupling between the articulatory or motor level, the acoustic level and the sensory level, be it visual or auditory. Indeed, anticipatory perceptual behaviour of subjects seems to be harnessed to these articulatory and acoustic events.

REFERENCES

- ABRY C. LALLOUACHE T. (1995) Le MEM : un modèle d'anticipation paramétrable par locuteur. Données sur l'arrondissement en français. Les cahiers de l'ICP. Bulletin de la Communication Parlée 3, 85-99.
- GAITENBY J (1965) The elastic word. Haskins Laboratories, Status Report, Speech Research 2, 1-12.
- HARDCASTLE W.J. VAXELAIRE B. GIBBON F. HOOLE P. NGUYEN N. (1996) EMA/EPG study of lingual coarticulation in /kl/ clusters. 4th Speech Production Seminar, 21-24 mai 1996, Autrans, France, 53-56.
- HECKER V. VAXELAIRE B. CATHIARD M. SAVARIAUX C. SOCK R. (2001) How movement expansion influences auditory perceptual extent. Probing into the Movement Expansion Model. Orage 2001 - Oralité et Gestualité : Communication Multimodale. Congrès International à Aix-en-Provence, juin 2001., Editions L'Harmattan, Paris, 450-456.
- HIRSCH F. SOCK R. CONNAN P.-Y. BROCK G. (2003) Auditory effects of anticipatory rounding in relation with vowel height in French. Proceedings of the XVth International Congress of Phonetic Sciences, Barcelona, 1445-1448.
- ROY J.-P. (2004) Etude de la perception des gestes anticipatoires d'arrondissement par les sourds et les malentendants. Doctorat des Sciences du Langage. Institut de Phonétique de Strasbourg – E.A. 3403. Université Marc Bloch, 296 p.
- ROY J.-P. SOCK R. VAXELAIRE B. HIRSCH F. (2003) Auditory effects of anticipatory and carryover coarticulation. In Proceedings of the 6th International Seminar on Speech Production, Sydney Australia, 243-248.
- SOCK R. VAXELAIRE B. (2004) Le diable perceptif dans les détails sensori-moteurs anticipatoires. In L'anticipation à l'horizon du présent. SOCK R. VAXELAIRE B. (Eds.) Mardaga, Sprimont, Belgique, 141-157.
- VAXELAIRE B. SOCK R. ROY J.-P. ASCI A. HECKER V. (2003) Audible and inaudible anticipatory gestures in French. 15th International Congress of Phonetic Sciences, Barcelona, 3 – 9 August, 447 - 450.
- VAXELAIRE B. SOCK R. HIRSCH F. ROY J.-P. (2007) Anticipatory laryngeal movements. An X-ray investigation. Proceedings of the XVIth International Congress of Phonetic Sciences, Saarbrücken, 525-528.

FONOLOGIA E PERCEZIONE. RIFLESSIONI SU UN MODELLO DI PERCEZIONE DEI SUONI NON-NATIVI

Mario Vayra* e Cinzia Avesani**

*Dipartimento di Studi Linguistici e Orientali, Università di Bologna; **Istituto di Scienze e Tecnologie della Cognizione, CNR

mario.vayra@unibo.it, avesani@pd.istc.cnr.it

SOMMARIO

L'influenza che la lingua nativa esercita sul parlato si manifesta in modo evidente nelle difficoltà che gli adulti, indipendentemente dal livello culturale, incontrano nell'eliminare l'"accento" della lingua materna allorché parlano una seconda lingua acquisita prima del quinto/sesto anno d'età (per una rassegna, cfr. Flege, 1991; per un inquadramento generale, e sull'importante nozione di *interlingua*, Costamagna e Giannini, 2003). E' meno ovvio, invece, ad un osservatore esterno – benché sia ben noto agli ascoltatori non-nativi - che quando gli adulti si trovano ad ascoltare una lingua non-nativa incontrano spesso difficoltà nel percepire le differenze fonetiche tra contrasti, vocalici o consonantici, che nella propria lingua non abbiano una funzione fonologica (cfr., ad esempio, Lisker e Abramson, 1970).

Contrariamente a quanto emerge dallo scenario della percezione interlinguistica nell'adulto, il bambino approssimativamente sotto i sei mesi non è ancora condizionato dalla fonologia della lingua nativa e si rivela in grado di discriminare differenze fonetiche native e non-native, con identica sorprendente facilità. I primi a osservare un mutamento evolutivo nella capacità del bambino di discriminare contrasti fonetici non-nativi furono Werker e colleghi (Werker e Tees, 1984). Questi ricercatori scoprirono che mentre i bambini di 6-8 mesi sono in grado di discriminare contrasti fonetici di tipo consonantico sia nativi sia non-nativi, i bambini di 10-12 mesi non discriminano che contrasti nativi. Tale schema evolutivo è stato confermato per numerosi altri contrasti fonetici non-nativi, incluse le vocali, che possono esibire uno "slittamento" percettivo più precoce (Best, 1994b; Best e McRoberts, 2003; Werker, 1989). Questi risultati sollevano alcune questioni generali. Essi mostrano infatti che intorno alla svolta del primo anno il bambino ha acquisito una conoscenza notevole della fonologia della lingua materna. Ma, certamente, il bambino di un anno non ha ancora raggiunto la competenza fonologica dell'adulto. Allora, qual è esattamente la natura dell'*informazione* linguistica che gli adulti sono riusciti a estrarre in relazione alla lingua materna, e che pone restrizioni tanto forti sulla percezione di contrasti non-nativi? L'esame di come un insieme di tipi diversi di contrasti fonetici non-nativi venga percepito da ascoltatori di lingue diverse, caratterizzate da sistemi fonologici anche disparati, si rivela illuminante circa la possibile risposta a questa domanda.

La nostra prospettiva teorica si centra sul *Perceptual Assimilation Model (PAM)*, un modello della percezione linguistica non-nativa sviluppato da C. Best (cfr., ad esempio, Best, 1993; 1994a; 1994b; 1995; 1996; 2007). Il modello si basa sull'osservazione empirica che i contrasti fonetici non-nativi non pongono tutti lo stesso grado di difficoltà percettiva all'ascoltatore maturo, e assume che la capacità di discriminarli dipenda dalle similarità e dalle dissimilarità *fonetico-articulatorie* che l'ascoltatore percepisce fra foni non-nativi e il sistema fonetico-fonologico nativo. La novità di PAM, che lo differenzia da altri modelli di percezione linguistica non-nativa, risiede nella definizione di cosa conti come "percettivamente simile". PAM definisce la similarità percettiva nel quadro teorico della Articulatory Phonology (cfr., ad esempio, Browman e Goldstein, 1986, 1989, 1992) e della Direct Realist Theory of Speech Perception (Gibson, 1979; Fowler, Rubin, Remez e Turvey, 1980). La percezione dei suoni del linguaggio si basa, secondo questa prospettiva, non tanto sulle proprietà *acustiche* dello stimolo prossimale, quali ad esempio la sua salienza, quanto sulla fonte distale dello stimolo, ovvero sugli eventi *articulatori* che danno forma all'*informazione* contenuta nei segnali. In questo caso la fonte distale è costituita dai gesti articolatori.

In questo studio valuteremo le predizioni del *Perceptual Assimilation Model* sia nella sua versione originaria, che in una versione più recente che fa propria una nuova ipotesi derivata dalla Fonologia Articolatoria. Tale ipotesi, nota come Articulatory Organ Hypothesis (Goldstein, 2003; Goldstein e Fowler, 2003; Studdert-Kennedy e Goldstein, 2003) è stata originariamente sviluppata per spiegare il processo di "sintonizzazione" (*attunement*) sulle strutture fonetico-articulatorie della lingua materna, e si fonda sul ruolo svolto dai gesti degli "organi" articolatori (labbra, apice della lingua, dorso della lingua, radice della lingua, velo e glottide) nella struttura fonologica della lingua materna e nelle assimilazioni percettive dei contrasti non-nativi. Secondo questa nuova versione di PAM, un ascoltatore avrà una maggiore difficoltà nel discriminare contrasti fonetici distinti da gesti prodotti con lo stesso articolatore primario (contrasti "intra-organo") rispetto a contrasti fonetici distinti da gesti prodotti con articolatori diversi (contrasti "tra-organi").

Abbiamo testato questa ipotesi in due esperimenti di percezione con soggetti monolingui funzionali, le cui lingue native (italiano e danese) hanno fonologie che consentono di predire assimilazioni percettive diverse degli stessi contrasti non-nativi (inglesi). I risultati evidenziano i limiti di una versione forte della Articulatory Organ

Hypothesis quando applicata alla percezione di ascoltatori adulti, e suggeriscono una possibile revisione di PAM.

RIFERIMENTI BIBLIOGRAFICI

- Best, C. T. (1993), Emergence of language-specific constraints in perception of non-native speech contrasts: A window on early phonological development, in de Boysson-Bardies, B., S. de Schonen, P. Jusczyk, P. MacNeilage, e J. Morton (a cura di) *Developmental neurocognition: Speech and face processing in the first year of life*, Dordrecht, Netherlands: Kluwer Academic Publishers, pp. 289-304.
- Best, C. (1994a), Emergence of native-language phonological influences in infants: A perceptual assimilation model, in Goodman, J. C. e H.C. Nusbaum (a cura di), *The Development of speech perception: The transition from speech sounds to spoken words*, Cambridge, MA: MIT Press, pp.167-224.
- Best, C. T. (1994b), Learning to perceive the sound pattern of English, in Rovee-Collier, C. e L. Lipsitt (a cura di) *Advances in Infancy Research* (vol. 9), Norwood NJ: Ablex Publishing Corporation, pp. 217-304.
- Best, C. T. (1996), Nonnative speech perception as a window on the native phonological system and its development, in *Journal of the Acoustical Society of America*, 99, 2602.
- Best, C. T., e Tyler, M. D. (2007), Nonnative and second-language speech perception: Commonalities and complementarities, in Bohn, O.-S. e M.J. Munro (a cura di) *Language Experience in Second language Speech Learning. In honor of James Emil Flege* Amsterdam: John Benjamins, pp.13-34.
- Browman, C. P. e Goldstein, L. (1986), Towards an articulatory phonology, in *Phonology Yearbook*, Vol. 3, pp. 219-252.
- Browman, C. e Goldstein, L. (1989), Articulatory gestures as phonological units, in *Phonology*, Vol. 6, pp. 201-251.
- Browman, C. P. e Goldstein, L. (1992), Articulatory Phonology: An Overview, in *Phonetica*, Vol. 49, pp. 155-180.
- Goldstein, L. (2003), Emergence of discrete gesture, in *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 4-9 August 2003, 85-88.
- Goldstein, L. e Fowler, C.A. (2003), Articulatory phonology: A phonology for public language use, in Schiller, N.O. e A.S. Meyer (a cura di), *Phonetics and Phonology in Language Comprehension and Production*, Mouton de Gruyter, pp. 159-207.
- Fowler, C., Rubin, P., Remez, R. e Turvey, M. (1980), Implications for speech production of a general theory of action, in Butterworth, B. (a cura di), *Language Production*, New York, Academic Press: 373-420.
- Gibson, James J. (1979) *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.
- Studdert-Kennedy, M. e Goldstein, L. (2003), Launching language: The gestural origin of discrete infinity, in Morten, C. e S. Kirby (a cura di), *Language Evolution*, Oxford: Oxford University Press, pp. 235-254.

TEMPI E MODI DI CONSERVAZIONE DELLE R ITALIANE NEI FRIGORIFERI CLIPS

Alessandro Vietti¹, Lorenzo Spreafico¹, Antonio Romano²

¹Centro di Competenza Lingue – Libera Università di Bolzano; ²Laboratorio di Fonetica Sperimentale "Arturo Genre",
Università degli Studi di Torino

alessandro.vietti@unibz.it, lorenzo.spreafico@unibz.it, antonio.romano@unito.it

SOMMARIO

Scopo della ricerca è offrire una caratterizzazione acustica di alcune realizzazioni di /r/ nell'italiano contemporaneo ricavabili dai dati *CLIPS* e da un campione raccolto a Bolzano.

Nel panorama degli studi sulle varietà d'italiano, se si escludono le pur acute osservazioni di Canepari (ad es. 1986, 1999), le ricerche sulle modalità di realizzazione di /r/ sono al momento relativamente poco avanzate: studi acustici preliminari hanno soltanto sottolineato caratteristiche salienti di rese piuttosto standard (cfr. tra gli altri Vagges *et alii*, 1978) oppure osservato dati dialettali specifici nell'ambito di studi con finalità più ampie (Soriano, 2003; Felloni, 2006).

Allorché per altri domini linguistici l'argomento, già esplorato preliminarmente, incomincia ad essere affrontato più estensivamente (cfr. ad es. Meyer-Eppler, 1959; Delattre, 1971; Schiller, 1988; Recasens, 1991; Espy-Wilson *et alii*, 1997; Solé, 1999; Wiese, 2001; Docherty & Foulkes, 2001; Engstrand *et alii*, 2007) per quello italiano disponiamo del solo quadro di riferimento offerto da Romano (2002; in prep.) che ha osservato alcune delle realizzazioni possibili col metodo dei *loci* acustici e nel quadro della teoria della perturbazione. Questo riferimento può tuttavia risultare inadeguato quando si tratti di rendere conto di articolazioni multiple e di strategie di realizzazione che, bisognose di verifiche articolatorie, sfuggano a rappresentazioni certe in quest'ottica. Ancora maggiori le incertezze che sorgono quando l'osservazione si basa su materiali in cui diversi fenomeni concorrenti si sovrappongono a quelli indotti da variabili ambientali o sociolinguistiche: alla varietà di realizzazioni può corrispondere infatti anche una notevole variabilità acustica.

L'analisi che qui proponiamo si basa sui dati di italiano letto ricavati dalle liste di parole contenute nei materiali *CLIPS* e da quelle usate in un'indagine sull'italiano a Bolzano (cfr. Vietti & Spreafico, 2008). In particolare ci siamo concentrati sulle sequenze /rV/ della parola *frigorifero*.

In termini generali, nei materiali *CLIPS*, nei quali è rappresentata da 120 realizzazioni (da parte di 8 parlanti per le 15 città del campione), abbiamo rilevato ad es. per la realizzazione della seconda /r/ di *frigorifero* (in attacco di sillaba accentata) una realizzazione dominante monovibrante (vibrata apico-alveolare, presente nel 40,8%; ben esemplificata nei dati di Palermo o, in presenza di una sospetta componente secondaria di uvularità, di Genova) che, rispetto a quella riportata nei dati di varietà iberiche (Recasens, 1991; Solé, 1999), si caratterizza per una maggiore durata e per una certa rigidità energetica prima e dopo che la fanno percepire (seppur non polivibrante) come più forte di una normale monovibrante.

Quanto alle numerosi varianti, i dati si presentano con una diversa distribuzione, con pochi casi di presunta cancellazione (3,3%) o, comunque, con sole forme di rotacismo vocalico (1,7%; con esempi isolati, da Napoli a Bergamo). Rese approssimanti pure sono state riconosciute nel 6,7% dei casi (in luoghi d'articolazione diversi), mentre di gran lunga più frequenti si sono rivelate le realizzazioni approssimanti interrotte da localizzati cali di energia (18,3%). Un altro 5,8% è rappresentato da realizzazioni monovibranti di durata particolarmente significativa (da 33 a 50 ms) durante la quale le caratteristiche acustiche di energia denunciano una tenuta simile a quella di un'occlusiva sonora e sono confermate da un discreto scoppio (che anche all'ascolto lascia pensare a brevi /d/ alveolari o postalveolari). Realizzazioni velari, uvulari e faringali (uvularizzate o faringalizzate), approssimanti o costrittive, compaiono a Genova (1), Parma (5), Torino (4), Milano (1) e Cagliari (1) per un complessivo 10% (l'unica chiara costrittiva uvulare è di Parma, mentre sono risultate più comuni le monovibranti o approssimanti alveolari uvularizzate; cfr. Canepari, 1999). Una vibratile (talvolta lateralizzata) è infine dominante nei dati di Venezia (per un residuo 5% complessivo sui dati nazionali), rendendo i *frigoriferi* di questa località gli unici del corpus la cui provenienza geografica sia facilmente riconoscibile. L'analisi del campione raccolto a Bolzano conferma solo in parte le osservazioni generali ricavabili dai dati *CLIPS*. Infatti, sebbene presenti, le realizzazioni monovibranti apico-alveolari non sono significativamente dominanti. Al contrario esse rappresentano solo una delle possibili rese e si inseriscono in un più articolato quadro di realizzazioni caratterizzato da un'elevata variabilità che comprende, in ordine decrescente di frequenza, sia articolazioni uvulari (approssimanti e monovibranti), sia altre articolazioni apicali (approssimanti) e/o postalveolari.

Per queste rese, come per quelle dei materiali *CLIPS*, vista la rilevanza dei contributi delle altre formanti (F3 e F4) piuttosto che descrivere il suono cercando di determinare il *locus*, abbiamo osservato le caratteristiche temporali della transizione, in particolare la maggiore o minore rapidità del movimento acustico. Questo è definito attraverso la misurazione dei valori delle formanti delle vocali precedente e seguente nei punti stazionari

e di transizione. I $_F$ così ricavati, rapportati ai $_T$ e integrati con le informazioni relative alla concavità o convessità delle curve, permettono infatti di avanzare una descrizione più completa delle transizioni formantiche. Accanto alla descrizione delle caratteristiche spettro-temporali si osservano la presenza e la consistenza numerica di eventuali *burst*, rumori o frizioni legati alle diverse strategie di articolazione di /r/.

RIFERIMENTI BIBLIOGRAFICI

- Canepari, L. (1986). Italiano standard e pronunce regionali. Padova: CLEUP.
- Canepari, L. (1999). MaPI. Manuale di Pronuncia Italiana. Bologna: Zanichelli.
- Delattre, P. (1971). "Pharyngeal features in the consonants of Arabic, German, Spanish, French and American English". *Phonetica*, 54, 93-108.
- Docherty, G. & Foulkes, P. (2001). "Variability in /r/ production. Instrumental perspectives". In R. van Hout & H. van de Velde (eds.) (2001), 173-184.
- Engstrand, O., Frid, J. & Lindblom, B. (2007) "A perceptual bridge between coronal and dorsal /r/". In P. Beddor, M. Ohala & M.-J. Solé (eds.), *Experimental Approaches to Phonology*. Oxford: Oxford University Press, 175-191.
- Espy-Wilson, C.Y., Narayanan, S., Boyce, S.E. & Alwan, A. (1997). "Acoustic Modelling of American English /r/". *Proc. of Eurospeech '97* (Rhodes, 1997), 393-396.
- Felloni, M.C. (2006). "Un'indagine sociofonetica a Parma: la realizzazione del fonema /r/ nell'italiano regionale", Tesi di Laurea Specialistica in Linguistica Teorica e Applicata, Facoltà di Lettere e Filosofia, Univ. degli Studi di Pavia (rel. Gianguido Manzelli), inedita.
- Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.
- Meyer-Eppler, W. (1959). "Zur Spektralstruktur der /r/-Allophone des Deutschen". *Akustica*, 9, 246-250.
- Recasens, D. (1991). "On the production characteristics of apicoalveolar taps and trills". *Journal of Phonetics*, 19, 267-280.
- Romano, A. (2002). A contribution to the study of phonetic variation of /r/ in French and Italian linguistic domains. Poster pres. at the 2nd Int. Workshop on the Sociolinguistic, Phonetic and Phonological Characteristics of /r/ (Université Libre de Bruxelles, 5-7 Dec. 2002), preprint 62 pp. (articolo in c. di p. in H. Van de Velde, R. van Hout & D. Demolin, eds.).
- Schiller, N. (1988) "The phonetic variation of German /r/". In M. Butt & N. Fuhrhop (eds.) *Variation und Stabilität in der Wortstruktur*. Hildesheim: Olms, 261-287.
- Solé, M.J. (1999). "Production requirements of Apical Trills and Assimilatory Behavior". *Proc. of the XIVth International Congress of Phonetic Sciences* (San Francisco, 1999), 487-489.
- Sorianello, P. (2003). "Aspetti coarticolatori nel parlato di Siena". In G. Marotta (a cura di), *La coarticolazione* Pisa: ETS, 101-110.
- Vaggies, K., Ferrero, F.E., Magno Caldognetto, E. & Lavagnoli, C. (1978). "Some Acoustic Characteristics of Italian Consonants". *Journal of Italian Linguistics*, 3: 69-85.
- Van de Velde, H. & Van Hout, R. (eds.) (2001). "'r-atics, Sociolinguistics, Phonetic and Phonological characteristics of /r/'", *Etudes et Travaux*, 4.
- Vietti, A. & Spreafico, L. (2008). "Phonetic variation of /r/ in a language contact context: The case of South Tyrol Italian". Poster presented at *Laboratory Phonology 11 - Phonetic detail in the lexicon* (Wellington, 30 June - 2 July 2008).
- Wiese, R. (2001). "The unity and variation of German /r/". In R. van Hout & H. van de Velde (eds.) (2001), 11-26.

PRESENTAZIONE ED ESEMPLIFICAZIONE DI “PHON”, UN PROGRAMMA PER LA CODIFICA E L’ANALISI AUTOMATICA DEGLI ASPETTI SEGMENTALI DEL PARLATO

°Claudio Zmarich, °Maria Pia Bardozzetti, °Caterina Pisciotto, *Serena Bonifacio

°Istituto di Scienze e Tecnologie della Cognizione (ISTC), C.N.R., Sede di Padova; *IRCCS Istituto per l’Infanzia “Burlo Garofolo”, Trieste

claudio.zmarich@pd.istc.cnr.it, mariapia.bardozzetti@libero.it, caterina.pisciotta@pd.istc.cnr.it, logopedia@burlo.trieste.it

SOMMARIO

Il programma freeware Phon (© 2006-2008 The Phon Project) costituisce il passo più recente e decisivo verso l’informatizzazione della ricerca sull’acquisizione fonologica. Esso nasce all’interno della comunità CHILDES (<http://childes.psy.cmu.edu/>), quando molti dei ricercatori interessati agli aspetti fonologici di tipo segmentale dell’acquisizione linguistica e interessati a verificare l’universalità delle loro ipotesi teoriche, avendo bisogno di disporre di database interlinguistici, si trovarono limitati dalle funzionalità dei programmi fino ad allora sviluppati, CHAT (per la codifica) e CLAN (per l’analisi), che erano esclusivamente riservati agli aspetti conversazionali, sintattici, morfologici e lessicali.

Dopo una serie di tentativi, passati attraverso il tentativo di adattare CHAT e CLAN, e il ricorso a software proprietari, si decise di creare un programma ex-novo. Come è riportato sul sito di Phon, questo “è un programma software che facilita straordinariamente un gran numero di funzioni richieste per le analisi dello sviluppo fonologico. Per es., Phon supporta la connessione ai dati multimediali, la segmentazione delle unità, la trascrizione in doppio cieco, l’etichettatura automatica dei dati, e la comparazione sistematica tra il target (modello) e le forme fonologiche (effettivamente prodotte). Tutte queste funzioni sono accessibili attraverso un’interfaccia grafica user-friendly. I database codificati e analizzati in Phon possono anche essere interrogati utilizzando una interfaccia di ricerca potente. Questo programma lavora sia su Mac OS X sia su Windows, ed è interamente compatibile con il formato CHILDES. Phon è disponibile gratuitamente come software open-source. Esso soddisfa bisogni specifici principalmente correlati allo studio dello sviluppo fonologico di L1 (incluso il babbling), l’acquisizione di L2, e i disturbi di linguaggio. In futuro, Phon faciliterà lo scambio di dati tra i ricercatori, anche attraverso la costruzione di un database condiviso chiamato “PhonBank”, un’altra iniziativa nuova del consorzio CHILDES per supportare i bisogni metodologici ed empirici della ricerca in tutte le aree dello sviluppo fonologico”.

Per esemplificare le sue potenzialità, lo abbiamo applicato ai dati di registrazioni audio di 12 bambini, 4 per ciascun gruppo di età: 18-23; 24-29; 30-36 mesi.

I bambini, di sviluppo fisico, cognitivo e linguistico riportato normale, sono stati registrati in due asili nido di Trieste, mentre rispondevano a un test prototipico per la valutazione delle capacità articolatorie nei bambini di età compresa tra i 18 e i 36 mesi (Bonifacio & Zmarich, cfr. scheda ministeriale per progetto di ricerca RC02007, Resp. E. Zocconi, Burlo-Garofolo). Questo test si propone la stesura dell’inventario fonetico, basato sulla produzione verbale del bambino (stimolata ma non ripetuta), per capire se un fono o un gruppo consonantico risultano acquisiti in sede iniziale o non iniziale in parole che possono avere forme anche diverse dal target adulto (cfr. Zmarich & Bonifacio, 2005). Un tale strumento, usabile anche da Psicologi dello sviluppo e Linguisti, è attualmente carente in Italia, poiché il test PFLI (Bortolini, 1995) è dichiaratamente non applicabile al di sotto dei 24 mesi, ma, come le logopediste sanno bene, difficilmente applicabile in un contesto clinico anche ai bambini di età superiore.

A ciascun bambino è stata presentata la sezione del test relativa alla sua fascia d’età. Il test consiste in denominazioni e produzioni spontanee stimolate dalla manipolazione di oggetti-giocattolo rappresentanti alcuni item lessicali del Primo Vocabolario del Bambino (PVB) o questionario macArthur (Caselli e Casadio, 1995), scelti in base a un criterio sia fonetico (che tutti i foni della lingua italiana siano attestati in almeno 2 parole per ciascuna posizione lessicale) che semantico/frequenziale (le parole devono essere tra le più prodotte dai bambini, in base alle ricerche di Caselli Pasqualetti e Stefanini (2007). Tutta la produzione verbale ottenuta su presentazione degli item è stata audioregistrata e successivamente trascritta in simboli fonetici IPA e codificata al PC con Phon, versione 1.3R500.

La codifica è consistita nella creazione di un record per ogni parola realizzata dal bambino, codificata in caratteri alfabetici come glossa, codificata direttamente in simboli IPA (SilDoulos IPA93) secondo la pronuncia del bambino e secondo la pronuncia adulta. Le forme delle due pronunce ricevono un allineamento automatico dei segmenti nei tipi sillabici costituenti la parola, che si basa sulle regole di sillabificazione dell’italiano implementate su nostra comunicazione e richiesta da Y. Rose, uno dei creatori di Phon. Ogni record è associabile facilmente e in modo permanente alla selezione del segnale acustico che lo riguarda, in modo tale che in qualsiasi momento è possibile verificare la trascrizione ascoltando l’audio corrispondente.

Oltre a prevedere numerose funzioni di interrogazione, Phon consente diversi tipi di analisi di uso più comune nello studio dell'acquisizione fonologica. Qui verranno esemplificati quelli relativi alle statistiche di frequenza ripartite per posizione del fono rispetto alla parola e alla sillaba, e all'analisi di alcuni fra i cosiddetti processi fonologici (cfr. per es. Ingram, 1981), che semplificano la struttura e il sistema fonologico delle parole adulte, e che grazie all'allineamento sillabico trovano una facile applicazione analitica.

BIBLIOGRAFIA

- Bortolini, U. (1995), *PFLI Prove per la valutazione fonologica del linguaggio infantile*. Padova: Edit Master Srl.
- Caselli, M.C. & Casadio, P. (1995), *Il primo vocabolario del bambino*, Milano: Franco Angeli.
- Caselli M. C., Pasqualetti P. & Stefanini S. (2007), *Parole e frasi nel "Primo vocabolario del bambino"*, Milano: Franco Angeli.
- Ingram, D. (1981), *Procedures for the phonological analysis of children's language*, Baltimore: University Park Press.
- Phon*: <http://phon.ling.mun.ca/phontrac>
- Zmarich, C., Bonifacio, S. (2005), *Phonetic inventories in Italian children aged 18-27 months: a longitudinal study*, in *Proceedings of INTERSPEECH'2005-EUROSPEECH*, Lisboa, September 4-8, 757-760.

Elenco dei partecipanti

(N = 74)

Abete, Giovanni
Albanese, Maria
Albertini, Martina
Alfano, Iolanda
Avesani, Cinzia
Balloi, Barbara
Barry, William
Bernardasci, Camilla
Beyeler, Francesca
Bonaccini, Emanuela
Bravi, Paolo
Bruni, Alessandro
Calamai, Silvia
Canalis, Stefano
Cangemi, Francesco
Cirillo, Roberto
Coro Gianpaolo
Cosi, Piero
Cutugno, Francesco
De Dominicis, Amedeo
Dellwo, Volker
Falcone, Mauro
Faraoni, Vincenzo
Filipponio, Lorenzo
Fischer, Andreas
Galatà, Vincenzo
Gamal, Dalia
Garrapa, Luigina
Gobbato, Paolo
Grimaldi, Mirko
Heim, Sarah
Herrmann, Fritz
Hirsch, Fabrice
Kaufmann, Nuria
Keller, Eric
Llanaj, Endritt
Lee, Cristopher
Lio, Rosita
Lopez, Gianluca
Loporcaro, Michele
Ludusan, Bogdan
Mairano, Paolo
Margherita, Ilaria Libera
Marotta, Giovanna
Mastantuoni, Eduardo
Maturi, Pietro
Meyer, Martin
Miotti, Renzo
Nocchi, Nadia
Orletti, Franca
Paciaroni, Tania
Picuccio, Lucia
Renzelli, Mariadina
Retaro, Valentina
Ricci, Irene
Romano, Antonio
Romito, Luciano
Salza, Pier Luigi
Savy, Renata
Schirru, Carlo
Schmid, Stephan
Schwab, Sandra
Schwarzenbach, Michael
Sisinni, Bianca
Soldo, Serena
Spreafico, Lorenzo
Studer, Dieter
Talarico, Antonella
Tarasi, Andrea
Totino, Carmen
Uguzzoni, Arianna
Vayra, Mario
Vietti, Alessandro
Zmarich, Claudio

Comitato scientifico

Cinzia Avesani, CNR-ISTC, Padova
Pier Marco Bertinetto, Scuola Normale Superiore, Pisa
Silvia Calamai, Università degli Studi di Siena
Piero Cosi, CNR-ISTC, Padova
Francesco Cutugno, Università Federico II, Napoli
Amedeo De Dominicis, Università della Tuscia, Viterbo
Mauro Falcone, Fondazione Ugo Bordoni, Roma
Barbara Gili-Fivela, Università degli Studi di Lecce
Michele Loporcaro, Università di Zurigo
Giovanna Marotta, Università degli Studi di Pisa
Pietro Maturi, Università Federico II, Napoli
Maurizio Omologo, FBK-IRST, Trento
Andrea Paoloni, Fondazione Ugo Bordoni, Roma
Antonio Romano, Università di Torino
Luciano Romito, Università della Calabria, Cosenza
Pier Luigi Salza, Loquenod S.p.A., Torino
Renata Savy, Università degli Studi di Salerno
Carlo Schirru, Università degli Studi di Sassari
Stephan Schmid, Università di Zurigo
Mario Vayra, Università di Bologna
Claudio Zmarich, CNR-ISTC, Padova

Comitato organizzatore (UZH)

Organizzazione generale

Stephan Schmid

Comitato locale

Vincenzo Faraoni

Lorenzo Filipponio

Michele Loporcaro

Nadia Nocchi

Susanne Oberholzer

Tania Paciaroni

Dieter Studer

Staff

Martina Albertini

Camilla Bernardasci

Francesca Beyeler

Francesco Cangemi

Sarah Heim

Fritz Herrmann

Endritt Llanaj

Luana Massaro

Lucia Picuccio

Michael Schwarzenbach

Sponsor

Hochschulstiftung der Universität Zürich

Zürcher Universitätsverein (ZUNIV)

Phonetisches Laboratorium der Universität Zürich

Phonogrammarchiv der Universität Zürich

Harman/Becker Automotive Systems GmbH

Förderverein *Amici del Liceo Artistico* (ALA)