Università degli Studi di Padova

DISLL
Dipartimento di Studi Linguistici e Letterari

LCL
Language and Communication Lab

S2G
from sound to gesture

ABSTRACT BOOKLET
21-23 May 2014

Sala delle Edicole
Piazza Capitaniato - Padua, Italy

Code-blending and co-speech gesture in bimodal bilinguals

Karen Emmorey

Bimodal bilinguals fluent in English and American Sign Language (ASL) often produce code-blends – the simultaneous production of a sign and a word – rather than code-switches between speaking and signing. Psycholinguistic studies reveal that code-blend production does not incur a processing cost (unlike code-switching) and that code-blending facilitates comprehension of both languages. Interestingly, bimodal bilinguals also produce ASL signs and linguistic facial expressions when speaking to non-signers, which suggests a connection between code-blends and co-speech gesture. Further, learning ASL can change the rate of co-speech gesture and the form of gestures (e.g., increasing the number of handshape types). These findings have implications for the nature of lexical selection mechanisms, bilingual language control, and the multimodal nature of communication.

Why gestures are not (only) a compensatory device – evidence from language learners

Marianne Gullberg
Lund University

It is often assumed that gestures are essentially compensatory in nature and help speakers convey information they have difficulties expressing, facilitate lexical retrieval, or help speakers to solve problems. Gestures are thus seen as both communicatively and cognitively compensatory. This view is especially common in research focused on "less competent" language users such as a child and adult language learners, or atypical populations. These assumptions can also be found in theories about the relationship between speech and gesture. I challenge this compensatory view of gestures by discussing three specific assumptions: 1) gestures replace speech in cases of trouble (disfluency), 2) gestures replace vocabulary, 3) gestures express meaning not found in speech. By looking at disfluencies and bimodal information structure in child and adult learner data, I will show that gestures are co-ordinated with fluent speech, not with disfluencies; that when gestures are recruited as problem-solvers, different problems have different gestural solutions; and that children and adults generally express similar information bimodally. Based on these observations, I argue for a more nuanced view of the speech-gesture relationship in production, especially in language development.

Research on bodily communication and the multimodality of social influence

Isabella Poggi[1], Francesca D'Errico[2]

[1]Università Roma Tre, [2]Università Telematica UniNettuno

The study of bodily communication of the Hearing in everyday life, rhetorical discourse, political communication and public speaking has seen its dawn ever since the first works by Cicero and Quintilian, and through Darwin's investigation on emotion expression has reached major developments in Ekman's, Efron's, Morris' and Kendon's work in the twentieth century. But since the first insights in deaf communication by William Stokoe, research on how our body communicates has received an impressive impulse from Sign Language research. Sign Language scholars have shown how the communication systems used by the deaf are languages in their own right, and how phonological, morphological, syntactic categories, previously used by linguists to analyze vocal languages, can be applied to sign languages as well, even though, at the same time, subtle and very interesting differences in their functioning can be found out, primarily due to the visual as opposed to the acoustic medium.

Yet, many of the insights gained by Sign Language research can be generalized and transferred to research on the hearing's multimodal communication. For example, the concept of phonology has been applied, just as to signs of sign languages, to the symbolic gestures of the hearing as well: in the Italian Gestionary (Poggi, 2003; 2007), all gestures can be analyzed as a combination of values in the parameters of handshape, location, orientation and movement. Further, not only a cherology – phonology of gestures – but also an optology and an aptology – phonology of communicative gaze and of communicative touch – have been proposed. For both gaze and touch a set of parameters have been singled out to analyze all possible signals in those communicative systems.

Moreover, research on the parameters of body communication systems has gone even more in depth: within the parameter of movement – that cuts across various systems: gestures, touch, head movements – an important sub-parameter is the so-called "expressivity" of gestures (Wallbott & Scherer, 1998; Hartmann, Mancini & Pelachaud, 2006; Pelachaud & Poggi, 2008): the movement amplitude, velocity, fluidity, and repetition.

Again: recent research contends that even in gestures and gaze can we find something we might call "morphemes": the handshapes in Kendon's gesture families (Kendon, 2004), but also, say, the degree of opening of eyelids in some items of gaze (Poggi, D'Errico & Spagnolo, 2009) bear particular

meanings: a match between form and meaning in which particular body features are physiologically linked to specific contents in such a way as to be considered “embodied morphemes”.

This work proposes an agenda for research in body communication of the hearings.

A first challenge is to find out the phonology of more body communication systems, such as, for example, head movements and postures. Second, to go more in depth in finding embodied morphemes. Third, for those communication systems in which there is a codified stable relationship between signals and meanings, to discover these correspondences, in such a way as to write down “lexicons” of gesture, gaze, touch, head movements, posture.

At the same time, though, this work aims at demonstrating how the systematic research on the lexicons, phonologies and morphologies of different communicative modalities can go hand in hand with research on specific domains of communicative interaction. In our presentation we will show how the study of body lexicons can usefully intertwine with the study on how people try to influence each other through multimodal communication. This research domain, that we call “Multimodal social influence” aims at studying the communicative behaviors in various modalities exploited by people to raise their own power and to lower the others’ power in political communication. In particular, the communicative behaviors used to display dominance and charisma, to cast discredit and ridicule on political opponents, to make parodies of them, will be overviewed, to give a sample of the richness and multifunctionality of multimodal communication.

Prosody and gesture as encoders of pragmatic meaning and social action

Pilar Prieto, ICREA-UPF

Since the publication of McNeill's (1992) groundbreaking book *Hand and mind: what gestures reveal about thought*, it has become increasingly clear that gestures are tightly intertwined with speech in time and semantic function. Yet the specific ways in which speech-gesture coordination works are far from well understood. And little is known about the role of prosody in this gesture-speech integration system. In this talk I will present results from several experiments carried out by our research group within the audiovisual prosody perspective which deal with the joint role of prosody and gesture in the expression of different types of communicative functions.

First, evidence will be shown that gestural movements (a) are temporally bound by prosodic constituents; and (b) the most prominent part of gestures coordinates with prosodic prominences in speech (e.g., Loehr 2012, Esteve-Gibert & Prieto 2013, a.o.). This synchronization between prosodic and gestural movements on the one hand and prosodic and gestural prominence patterns on the other reveals that both are fundamental features of the management of information structure and phrase grouping (and even turn-taking) across languages.

Second, a range of examples will illustrate how intonational and gestural meanings share a joint management of a set of pragmatic meanings: (a) speech act management; (b) positioning/epistemic management; (c) evaluative/presupposition frame management; (d) affective management; and (e) politeness/social management. It will also be shown how crosslinguistically these functions, which are related to the illocutionary force of utterances and added to their propositional content, constraining their interpretation in various levels, can be encoded morphosyntactically or through a variety of discourse particles. Following the perspective put forward by Cognitive Pragmatics, Conversational Analysis, and Interactional Grammar, it will be claimed that gestural and prosodic conventionalized signals, as crucial elements of the human faculty of language, are fundamental underlying features of the construction of social action (Bergmann et al. 2012, Heritage 2012, Escandell-Vidal 2012, Sperber & Wilson 1986/95, a.o.).

References.

Esteve-Gibert, N. & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and gesture movements. *Journal of Speech, Language, and Hearing Research* 56(3): 850-864.

Escandell-Vidal, V. (2012). Prosodia y pragmática. *Studies in Hispanic and Lusophone Linguistics* 4/1: 1-14.

Heritage, J. (2012). Epistemics in Action: Action Formation and Territories of Knowledge. *Research on Language and Social Interaction*, 45(1), 1–29.

Loehr, D. (2012) Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology* 3(1): 71-89.

McNeill, D. (1992). *Hand and Mind: what gestures reveal about thought.* Chicago, London: The Chicago University Press.

Bergmann, P., Brenning, J., Pfeiffer, M., & Reber, E. (2012). Prosody and Embodiment in Interactional Grammar. Berlin/Boston: Walter de Gruyter.

Sperber, D. & Wilson, D. 1986/95. Relevance: Communication and Cognition. Oxford: Blackwell; Cambridge, Mass.: Harvard University Press. Second edition 1995.

The linguistic organization of the body in Sign Language

Wendy Sandler, University of Haifa

The talk is about the relation between sign language and the body that conveys it. Some take the view that sign languages are just like spoken languages, distinguished only trivially by the medium of production, while others hold that sign languages are derived directly from natural gestures. The work presented here suggests that we are confronted with this puzzling dichotomy because we have often been looking in the wrong places in our quest to understand the linguistic properties of sign language and what it has to tell us about language in general.

I begin by isolating gestures of different parts of the body that are designated to manifest grammatical structure in established sign languages. Turning to a young sign language in a Bedouin village, I will show that the body begins as a nonsegmented whole, with only the hands designated to create symbolic images. Across signers of four age groups in this preliminary study, we will see, not a magical and sudden appearance of grammatical structure, but instead a gradual activation of different components of the body, each corresponding to different components of language. I present the hypothesis that what I call "The Grammar of the Body" underlies combinatoriality in language, and has implications for the study of gesture as well.

Handshape in Italian sign language: perception vs. theoretical models

Valentina Aristodemo[1,2], Carlo Geraci[1]

[1]CNRS Paris, [2]Università Ca' Foscari

1 Goals. Handshape is one of the four essential components in the phonological description of a sign, along with movement, orientation and place of articulation. In this paper we experimentally investigate: 1) the perception of articulatory and phonological complexity of handshapes in Italian Sign Language (LIS) and 2) how proficient (three) theoretical models are in predicting such complexity.

2. Theoretical models. We considered three theoretical frameworks that model complexity: the "BASCO 15" model, and Brentari (1998) and Ann (2006) models.

- Battison (1978) identified the B, A, S, C, O, 1, 5 handshapes as unmarked, because those identify the natural class of handshapes serving as a non-dominant hand in two-handed asymmetrical signs. The entire set of handshapes of a SL can be classified into unmarked (therefore simple) and marked (e.g. complex) handshapes.
- Brentari (1998) model is based on "dependency phonology" and "optimality theory". She suggests that handshape complexity (more generally the complexity of a handshape) depends on the number of nodes required to describe its phonological/phonetic structure
- Ann (2006) used physiological criteria (e.g. muscles involved in the production) and divides handshapes into three categories: easy, difficult, impossible.

3. Method and materials. **Stimuli**: 32 pictures containing a reference handshape (Y) and a target handshape have been presented to two groups of subjects. **Subjects:** A group of 44 deaf signer subjects (26 women), mean age 37; and a control group of 56 hearing non-signers subjects (32 women), mean age 38. Both groups were recruited via Internet (Vlog and Facebook).

Task: Subjects had to mark the degree of complexity of the target stimulus was indicated on the bar below the handshape. The complexity of the handshape "Y" served as baseline, Fig. 1.

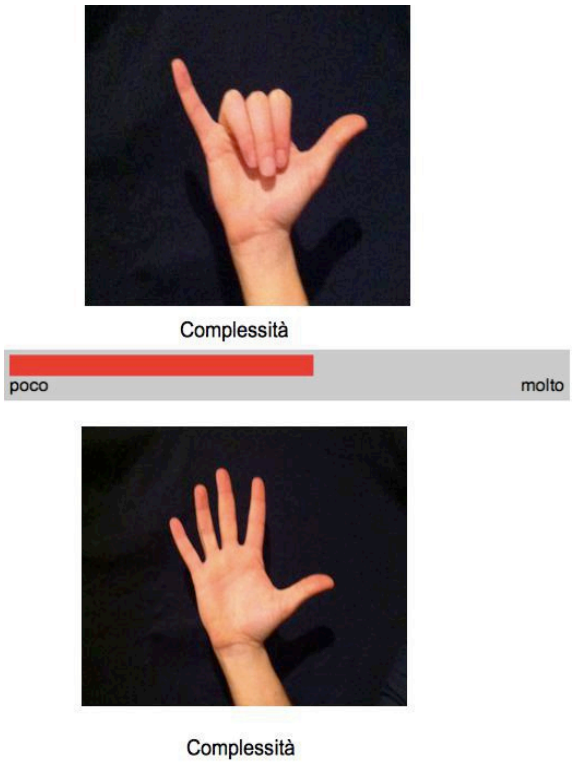


Figure 1: On the top there is the baseline handshape ("Y"). On the bottom, an example of target handshape ("5").

The task for the deaf group was to indicate the level of complexity of the handshape shown on the bottom of the picture.

The task for the hearing group was to indicate the level of complexity of the handshape shown on the bottom of the picture.

Results. Spearman correlation test reported in Table 1 shows that the best correlation between subjects' answers and the theoretical models is that predicted by Brentari (1998). However, further analysis based on mixed models (Baayen 2008) revealed a significant interaction between the complexity as predicted by the model Brentari (1998) and subjects' status (deaf vs. hearing). The interaction shows that the Brentari's model better predicts hearing subject answers than deaf subject

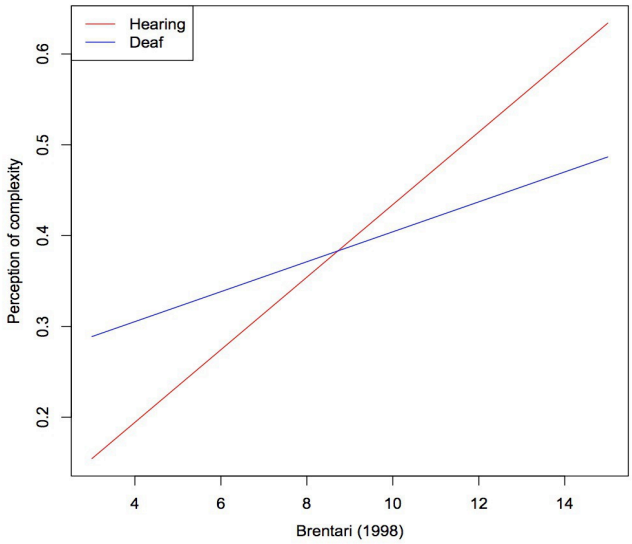
answers (fig.2).

Additional analysis that we will discuss during the presentation focused on the effect of [± curved fingers] feature showing that the hearing group perceived as significantly more complex configurations with a [+ curved fingers] feature, while no effect is found in the deaf group. This indicates that the hearing subjects have a greater sensitivity to articulatory aspects of the handshape than deaf subjects. Similarly, the analysis of the traits related to the selected fingers shows differences between deaf and hearing participants. The deaf group shows a gradient effect in the relationship between the number of fingers involved in the handshape and the complexity of the handshape itself, while deaf subjects' answers are more categorical (on effects of categoriality in the processing of the handshape see also Emmorey, McCullough & Brentari, 2003).

Table 1: Correlation between answers theoretical models complexity

Phonogical model	**Corr.**
Battison (1978)	0.149
Brentari (1998)	0.306
Ann (2006)	-0.170

Figure 2: Interaction Subjects' status* Brentari's model and



Discussion. The experimental results show that theoretical models of sign language phonology are not able to predict the complexity of the handshape parameter as it is perceived by signers. The fact that the best theoretical model is better predicts hearing subjects' answers indicates that it is built more on phonetic-articulatory features rather than phonological features. This is confirmed by the fact that the perception of complexity in the hearing group is sensitive to the [± curved fingers], while the deaf group is not sensitive to this feature. The feature [± curved fingers] is not contrastive in LIS (it only generates allophonic variation) and therefore it is not perceived as a source of complexity by the signers. Therefore we have evidence of a phonological representation in the signers' mind which eliminates some phonetic distinctions. As for the effect of selected fingers, the

difference between deaf and hearing group shows that even in this case deaf participants not use purely articulatory criteria but phonological criteria. Hearing participants are not able to access to the phonological representations of selected fingers in sign language so, they detect even small differences in articulation as bearers of complexity.

4 Conclusions. This paper shows the existence of a phonological level of representation in signers' mind mediating the perception of complexity. This level reduces the effect purely articulatory distinction. Psycholinguistic tests aimed at evaluating linguistic aspects (e.g. language competence) or cognitive aspects (e.g., short-term memory) of SL users. These tests manipulate stimuli simplicity/complexity based on a theoretical definition of complexity, assuming by default their predictive validity. This work demonstrates that 1) the assumption of validity cannot be accepted, and 2) greater attention to the selection of stimuli must be paid when phonological complexity is involved. This research is of vital importance for its contribution to both linguistic and psycholinguistics reflection and in a more general way for cognitive science applied to SL.

A Cross-Linguistic Study on Audiovisual Prosody

Ran Bi

School of Foreign Studies, Jiangsu Normal University, Shanghai

A vast majority of prosody research in the past few decades has focused on the auditory modality. However, human communication is inherently multimodal, it involves both verbal (e.g. speaking texts, intonation) and non-verbal (e.g. facial expressions) cues. With the emergence of McGurk effect (McGurk & MacDonald, 1976) and the FLMP (Fuzzy Logical Model of Perception) theory (Massaro, 1989), speech perception is no longer regarded as a purely auditory process, but a bimodal process involving both auditory and visual cues.

Previous studies on auditory prosody suggested that the prosodic features contribute to the conveyance of information focus and language information such as lexicon and syntax. Recently, in some studies on information processing, visual cues were found to play an important role in language comprehension. The important role of visual cues on the process of conceptualization is emphasized in cognitive linguistics and psychology, i.e. human beings initially form images and concepts through the perception of visual cues. The visual cues in speech communication usually involve facial expressions and gestures. The concept of "audiovisual prosody" is quite different from the traditional concept of "prosody", which is the integration of both auditory (prosodic features) and visual cues (i.e. facial expressions) in speech production and perception.

However, the number of studies on audiovisual prosody is very limited and studies on audiovisual prosody of L1 Chinese and L2 English learners have not been explored. Therefore, the present study aims to reveal the effects of different modalities (audio-only, vision-only, audiovisual) in information processing in terms of prosodic prominence and boundary tones across two typologically different languages (i.e. Chinese and English).

Looking for prosodic patterns in LSF and coverbal gestures. What about (de)synchronization?

Marion Blondel[1], Dominique Boutet[1] & Sébastien Delacroix[2]
[1]SFL, CNRS-Paris8;
[2]LAM, Institut National de Podologie, Paris

The way information is organised in discourse has to do with the relative salience the speaker/signer uses in encoding. This is true for both modalities: oral-aural and visual-gestural. Recently, motion capture has provided us with the objective measures of movement that we previously lacked. Our main objective in this presentation is to identify, in LSF (French sign language) interactions, the prosodic patterns involved in salience or retrieval of information, and to draw a comparison between these signed patterns and their coverbal gesture counterpart. We will concentrate on foregrounded elements with a contrastive focus (i) and on 'offset' elements in parentheses (ii), since these are, from a discourse perspective, opposing processes.

(i) 'not on the chair, on **the table'**

(ii) 'he gives her a journal, then another, **(maybe he's trying to chat with her, I don't know),** and the woman takes the journal, and so on…'

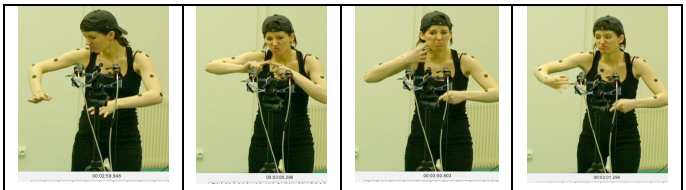
After Boyes-Braem (1999), Nespor & Sandler (1999), Wilbur & Martinez (2002), Tanaka & van der Hulst (2004) and van der Kooij et al. (2006), we examine 'intonation' in sign languages through the movement of the hands, torso and head considering their suprasegmental dimensions.
From the sign linguistics literature, Blondel (2003) compiles prosodic cues for stress (including focalization) and illustrates them for LSF: when a lexical item is put in focus, the size and the duration of the hands' movement can be increased, the location of the hands can be raised, manual movements can be associated with body and forward head movements.
Blondel and Le Gac (2007) observe three means to break with the prosodic environment and to provoke a 'parenthetical effect': a break in the distribution of sign location and the signer's body

positions, reduced amplitude of signs, and a break in rhythmic pattern for hands and body.

Signers use several manual and non-manual joints simultaneously, each joint follows its own pattern though remains constrained in distribution with the other joints. We hypothesize synchronisation and desynchronisation between manual and non-manual beats is the keystone of signed prosodic structure.

Our data were collected in a pilot study comparing the grammars of three sign languages and comparing these spatial and temporal structures with the systematic properties of coverbal gestures in hearing speakers. The video data are enriched through motion capture data. Productions of three participants per language have been recorded by a digital camera, a mocap system (Eagle) and an eye tracker (FaceLab). After watching voiceless clips, each participant was asked a series of pre-recorded questions in their language and answered an interlocutor in the same language.

We first extracted the utterances corresponding to what could be interpreted as parentheticals (fig.1) and contrastive focus respectively. Secondly, we made an qualitative inventory of the formal characteristics of these items.



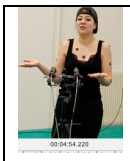
on these little white things	the white thing, you know		the round things

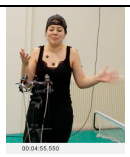
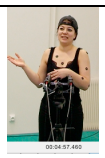
Figure 1 : LSFParticipant1

In parallel, we collected LSF data in the XYZ global coordinate system and transformed them in the autonomous coordinate system of each segment according to a biomechanical model (Wu et Al 2005). Segmental and joint kinematics was calculated to allow extraction of rhythmic patterns that appear through changes in direction or speed.

Accordingly, synchronised or unsynchronised stages between joints can be detected. Associated postures (movement or location of the trunk, backwards head movements) can be measured in an *absolute* way (geocentric frame of reference according to the lab) as in a *relative* way (intrinsic frame of reference according to each segment) in Levinson's terminology (2003). Through this frame switch, we investigate the question of the relativity of joint locations during the parenthetic or focus stages.

Vocal and gestural data of three hearing subjects have been gathered under the same conditions. Thus they allow us to investigate to what extent these patterns are specific to sign languages. As an illustration, figure2 shows a vocal parenthetic construction associated with a gestural pattern.



the man is leaving	maybe he has been called, I don't know	and while he's leaving...

Figure 2 : FrenchParisParticipant3

The two questions addressed in our exploratory study are: What are the prosodic units of sign language? What is the relationship between speakers' intonation, prosody and gestures? We intend to show that, on the one hand, the issue of interaction between non-manual and manual parameters in sign languages contributes to the fine-grained analysis of the role of prosody in conveying meaning in sign languages, and on the other hand that similar analysis can be used to address this issue in the multimodal production of hearing speakers.

References.

Blondel, M. (2003). Saillance linguistique dans une langue à modalité gestuelle", *in* J. François & A. Lacheret (eds), *Fonctions et moyens d'expression de la focalisation à travers les langues*. *Mémoires de la Société de Linguistique de Paris*, *Tome XIII*, Leuwen, Peeters, pp. 187-203.

Blondel, M. & D. Le Gac (2007). Entre parenthèses... y a-t-il une **intonation** en LSF ? *Silexicales*, 5, pp 1-16.

Boyes-Braem, P. (1999). Rhythmic temporal patterns in the signing of deaf early and late learners of German Swiss Sign Language. *Language and Speech* 42.2&3, pp. 177-208.

van der Kooij, E. Crasborn, O & W. Emmerik (2006). Explaining prosodic body leans in Sign Language of the Netherlands: pragmatics required. *Journal of Pragmatics* 38, pp. 1598-1614.

Levinson, Stephen C. *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge University Press, 2003.

Nespor, M. & W. Sandler (1999). Prosody in Israeli Sign Language. *Language and Speech* 42 (2-3), 143-176.

Tanaka, S. & van der Hulst, H. (2004). *Speed of hand movement: a quantitative study*. Poster at the TISLR 8, Barcelona, September 30-October 2.

Wilbur, R. B. & A. *Martínez* (2002). *Physical correlates* of prosodic structure in American Sign Language, *Chicago Linguistic Society*, 38, pp.693-704.

Wu, Ge, Frans C.T. van der Helm, H.E.J. (DirkJan) Veeger, Mohsen Makhsous, Peter Van Roy, Carolyn Anglin, Jochem Nagels, et al. « ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion—Part II: shoulder, elbow, wrist and hand ». *Journal of Biomechanics* 38, n° 5 (mai 2005): 981-992. doi:10.1016/j.jbiomech.2004.05.042.

Gestures and speech interactions in L2 phonetic integration activities with verbo-tonal method

Magali Boureux
Department of General Psychology, University of Padua.

Research emphasizes a close relationship between speech and spontaneous gesture [10][12]. The aim of this ongoing study is to investigate the typology and the role of gestures in support of the L2 phonetics' teaching and learning.

Foreign accents are cause of poor understanding or misunderstanding in the listener and may also generate speakers' inhibitions. On the contrary, a good pronunciation guarantees a fluent oral communication and allows to convey efficiently pragmatic meanings and communicative intents. Due to its complexity, the subject of L2 phonetics is rarely touched upon in L2 classes: teachers only introduce some isolated segmental or intonative characteristics through explicit contrastive or articulatory exercises.

However, the verbo-tonal method (VTM) of phonetic integration [7] allows to take into account the language as a complex macro-system composed by micro systems, which work in interaction [14][8]. Among these micro-systems there is the verbal one that contains sub-systems like lexicon, morpho-syntax and phonetics (which combines phonology and prosody). Another micro-system is the non-verbal one, which includes gestures, mimogestures and proxemics.

VTM stresses that production deficits are influenced by L1, both at segmental and suprasegmental levels [15][13][3]. The categorization of the L1 verbal properties [1][9][11] generates a "phonological deafness" [5][6] that influences the speech perception of a L2 and, therefore, the characteristics of the oral production in a non-native system [7][16][3].

In fact, MVT exploits the interaction between phonology, prosody and non verbal features to globally support the perception and discrimination of L2 verbal characteristics, and then to facilitate their integration and production. The teacher corrects the phonetic errors produced in L2 by proposing a speech model that fosters the acoustic L2 specific features [2][4].

In this context, gestures and body tenseness give a visual and proprioceptive perception of phonology and prosody that are considered for their acoustic and articulatory characteristics. Therefore, fundamental frequency is considered in terms of *height*, varying from low to high; frequency spectrum in terms of *timbre* as *color*, from dark to clear; amplitude, in terms of *intensity*, from loud to weak and

duration, from long to short. For didactics, the height and color variations are interpreted by hand movements that can follow the melodic contour for intonation, show a low position in the space to help the perception of the loudness and darkness of the sound, or a high position for putting in evidence higher and clearer characteristics of the sounds. The modifications of the quantity (intensity and duration) of the verbal features can be accompanied by syllabic scansion to elicit the perception of the rhythm or the vocal phonemes color. A global body contraction *vs*. release is used to contextualize the used gestures.

Our research will provide, on the one hand, the analysis of the quality and quantity of gestures that are used by experts when reinforcing L2 phonological and prosodic integration during teaching activities. On the other hand, it will allow evaluating the effects of gestures on learners' productions.

References

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

[2] Boureux, M. 2005. L'attività di ripetizione nella progressione fonetica in lingua straniera: il caso di adulti italiani. in: Cosi P. (ed.) *Proc. 1st AISV 2004 "Misura dei parametri"*. Padova: EDk Editore srl. - ICA, 525-538.

[3] Boureux, M., 2012. French prosody of Italian speakers: characteristics and didactic inputs. in Busà, M.G, and Stella A. (eds.), *Methodological Perspectives on Second Language Prosody. Papers from ML2P 2012*, Padova: CLEUP, 42-46.

[4] Boureux, M., Batinti, A. 2004. La prosodia. Aspetti teorici e metodologici nell'apprendimento-insegnamento delle lingue straniere, *Proc. 14th GFS.* Roma: Esagrafica, 233-238.

[5] Dupoux, E., Peperkamp, S. 2002. Fossil markers of language development: phonological deafnesses in adult speech processing. In: Laks, B., Durand J. (Eds), *Phonetics, Phonology, and Cognition.* Oxford : Oxford University Press, 168-190.

[6] Dupoux, E., Pallier, C., Sebastian, N., Mehler, J. 1997. A Destressing "Deafness" in French? *Journal of Memory an Language* 36, 406-421.

[7] Guberina, P., 1965. La Méthode audio-visuelle structuro-globale, *Revue de Phonétique Appliquée* 1, 35-64.

[8] Intravaia, P. 2000. Formation des professeurs de langue en phonétique corrective. Paris: Didier Erudition; Mons: CIPA.

[9] Iverson, P., Kuhl, P.K. 1995. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America* 97 (1), 553–562.

[10] Kendon, A. 1972. Some relationships between body motion and speech: an analysis of an example. in Siegman A. and Pope P. (eds.), *Studies in Dyadic Communication*. Elmsford, NY: Pergamon, 177-210.

[11] Kuhl, P. K. 1991. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 50, 93–107.

[12] McNeill, D. 1992. *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press.

[13] Polivanov, E. 1974. The subjective nature of the perceptions of language sounds. In: Polivanov, E., *Selected Works. Articles on General Linguistics (compiled by A. Leont'ev)*, The Hague: Mouton, 223-237.

[14] Renard, R. 1979. La méthode Verbo-Tonale de correction phonétique. Mons: Didier-Bruxelles.

[15] Sapir, E. 1921. *Language*. New York: Harcourt Brace Jovanovich.

[16] Trubetzkoy, N.S. 1969. *Principles of phonology*. Berkeley and Los Angeles: University of California Press.

Visual and vocal cues in signaling voice qualities

Zuleica Camargo[1,2], Thamires Rodrigues Freitas[1], Sandra Madureira[1,2]

[1]Integrated Acoustic Analysis and Cognition Laboratory-LIAAC.

[2]Graduate Applied Linguistics and Language Studies Program.

Pontifical Catholic University of São Paulo (São Paulo, Brazil)

zcamargo@pucsp.br

Introduction: The phonetic investigation of voice quality has been receiving increasing interest in speech research due to the fact that voice quality has linguistic, paralinguistic and extra-linguistic functions. However, classifying voice qualities is a difficult task and relying on a phonetic description of voice quality as provided by Laver (1980) and by the perceptually based Vocal Profile Analysis Scheme (VPAS) developed by Laver et al (2007) as a tool to evaluate voice qualities have been found useful in studies previously reported. According to Laver's phonetic model the basic analytical unit is the setting, that is described as long-term articulatory or phonatory tendencies of the vocal apparatus. All settings are described in relation to a neutral setting that corresponds to the vocal apparatus in a rest-like position. The relation between the voice quality settings and the speech segments in Laver's model is explained by means of two principles: the compatibility and the susceptibility principles. The compatibility principle states that there are movements which are compatible and others which are not (lip rounding and lip spreading, for example). The susceptibility principle proposes that depending on the inherent nature of segments, there will be less or more susceptibility of the speech segments to the long-term muscular settings. Due to this principle the analysis of key-speech segments to evaluate voice quality settings is mandatory. In this context, this study aimed at comparing visual and vocal cues in the evaluation of voice qualities settings as described in the VPAS.

Methods: 38 samples were selected from a phonetic voice quality database. The corpus design comprised 03 random repetitions of 04 Brazilian Portuguese (BP) key-speech sentences and semi-spontaneous speech samples. These recordings comprised a variety of voice quality settings from the vocal tract (lips, jaw, tongue tip and body, pharynx, velopharynx and larynx height), laryngeal/phonatory (vocal folds modes of vibration) and muscular tension mechanisms (vocal tract and laryngeal hyper and hypofunction). The audio samples were evaluated perceptually by two expertise judges in the use of VPAS. The video samples were edited and decoupaged by the softwares GIMP 2.6, PITIVI 0.13.4 and KINO 1.1.1, generating 30-56 frames per second from the images. To

analyse the images a visual profile was applied. The images were analyzed in terms of the recurring gestures in successive frames, taking into account key-speech segments. For each key-speech segment, 03 frames were analyzed. The results were described and discussed in relation to the details detected by the visual analysis and by the auditory judgments of the voice quality settings. Singularities related to specific key-segments were also discussed. To facilitate the descriptions of the long-term settings, a frame (from the same speaker) with the vocal apparatus in a steady position (without producing speech) was considered and also a frame from another speaker producing the same sound with a neutral setting, that is, without changing the inherent characteristics of the sound. The project was approved by the ethics committee (number 101/11).

Results and Discussion: The vocal tract voice quality settings that were influenced by visual perception were those related to the lips (spread, rounded, labiodentalization and limited range), jaw (closed, open, protruded, limited and extended range), tongue-tip (advanced), tongue-body (raised, lowered and advanced), pharynx (constriction) and laryngeal height (raised). In the muscular tension domain, we could find the influence of laryngeal hyperfunction and both vocal tract hypofunction and hyperfunction settings. The hyperfunction muscular tension settings were combined to non-neutral laryngeal phonatory settings, i.e., those produced with irregular vocal folds vibrations (whisper, breathy voice, harsh voice and creaky voice settings). For samples which displayed articulatory movements and phonatory maneuvers which couldn´t be observed directly by visual cues, such as pharyngeal and laryngeal (phonatory and tension) settings, recurrent adjustments of extrinsic muscles were observed, reinforcing both the segmental susceptibility and the physiological compatibility principles of the phonetic description of voice qualities. The findings related to muscular hyperfunction activity are consistent with the presence of some aperiodicity in the voice signal which characterize non-neutral phonatory settings. From the total amount of speakers with laryngeal hyperfunction settings auditorily detected, about 92% showed extrinsic laryngeal muscle activity, 73% showed laryngeal skeleton contour clearly defined and 67 % orofacial hyperfuncion activity. These data reinforce postulates of the phonetic model of voice quality, in which the mutual physiology leads to some degree of compatibility between settings and susceptibility mostly depends on the degree of shared muscular anatomy (Laver, 1994). For the visual detection of vocal quality settings, the following key-speech segments were found to be relevant: rounded [ɔ], [o] and [u] and unrounded vowels [a] for spread lips; unrounded vowels [a], [ɛ], [e] and [i] for rounded lips; bilabial consonants [p], [b] and [m] for labiodentalization; vowels [i], [o] [u] for open jaw; vowels [a], [ɛ] and [o] for closed jaw; vowels [o] and [u] and

consonants [f] [s] for protruded jaw; vowels [a] and [u] for limited and extended range of lips and jaw; consonants [t], [d], [s], and [z] for advanced tongue tip; vowels [i] and [u] for pharyngeal constriction; all vowels for raised larynx, phonatory and laryngeal hyperfunction. Some other interesting findings were related to the detection of greater degrees of the presence of settings when judging with images (especially for lips and jaw) and the importance of key-speech segments to detect the effects of long-term supralaryngeal gestures (and some implication to phonatory and muscular tension settings). The findings reinforce the claim for an integrated approach of voice quality settings, based on perceptual, physiological and acoustic descriptions in speech data.

Conclusion: The findings provide evidence in favor of the relevance of visual cues in identifying the following vocal quality settings: lips, jaw, tongue-tip and body, laryngeal height, laryngeal hyperfunction, vocal tract hypofunction and hyperfunction and non-neutral phonatory voice quality settings. They also reinforce the importance of the use of key-speech segments in identifying voice quality settings, since the principle of susceptibility turned out to be useful to identify segments affected by the settings aforementioned. The impact of using visual cues in identifying voice quality settings is also depicted.

Keywords: Voice quality; Phonetics, Expressivity, Psycholinguistics

The Emergence of Phonological Structure in Central Taurus Sign Language

Naomi K. Caselli, Rabia Ergin, Ray Jackendoff, Ariel M. Cohen-Goldberg
Tufts University (Medford/Somerville, MA, US)

Central Taurus Sign Language (CTSL) is a village sign language that emerged as a result of recessive deafness in two villages in the Central Taurus Mountain region of south-central Turkey. Approximately 15 deaf people from the second and third generations currently live in one village and 13 in the other (3-4% and 0.5-0.6% of the population, respectively). Due to cultural, geographical and financial circumstances, the language has emerged in isolation. A native Turkish Sign Language (TID) signer verified that CTSL bears only minimal influence from TID.

Like all emerging sign languages, CTSL provides insight into the human capacity for language in the absence of a linguistic model. CTSL is comparable to another well-studied emerging sign language, Al-Sayyid Bedouin Sign Language (ABSL). Both village sign languages are about three generations old but ABSL has five times as many deaf users (Sandler, Meir, Padden, & Aronoff, 2005). These simultaneously emerging languages present an opportunity to make comparisons and modest generalizations about language emergence.

Sandler, Aronoff, Meir, & Padden (2011) have argued that ABSL lacks many of the hallmarks of a phonological grammatical system. They report violations of several phonological constraints. Unlike established sign languages, they found only two groupings of selected fingers (all fingers, or only the index finger). We present a parallel description of the emergence of a phonological system of CTSL.

Selected Fingers

Six deaf CTSL signers completed a picture-naming task of common fruits and vegetables. A native ASL user coded all tokens for selected finger groupings. Selected fingers were defined as either the fingers that were moving or the fingers that were in the foreground (Brentari, 1998). To establish a reference, 286 ASL lexical signs of various semantic categories were elicited from native deaf ASL user in a picture-naming task and coded in a similar fashion.

Like ABSL, the majority of CTSL signs select all fingers or the index finger. In contrast, ASL selected finger groupings were more evenly distributed across 11 groupings of selected fingers (see Figure 1). Unlike ASL, which requires that non-selected fingers are either fully flexed or extended (Brentari, 1998), in CTSL non-selected fingers were frequently partially extended (see Figure 2).

Phonological Constraints

Like ABSL, we found several violations of phonological constraints. If both hands in a sign move, then specifications of the hands must be identical (Symmetry Condition; Battison, 1978). We identified several violations of the Symmetry Condition. Three signers produced CUCUMBER with two different handshapes (B and O), yet both hands moved (see Figure 3).[1] According to the Dominance Condition, if the two handshapes of a two-handed sign are different, the non-dominant hand must have an unmarked handshape and not move (Battison, 1978). One signer produced CORN by keeping the dominant hand still while moving the non-dominant hand. Several non-symmetrical two-handed signs used unconventional non-dominant handshapes (F).

Conclusions

CTSL appears to adhere less strictly to phonological constraints and to have fewer phonological contrasts than established sign languages. The selected finger groupings and distributions are nearly identical to ABSL. Though it is possible that these emerging languages independently arrived at the same distinction by chance, the fact that these are the two maximally distinct groupings leads us to believe otherwise. By analogy, spoken languages tend to maximize dissimilarity between the vowels used in a given language (e.g., /i/, /a/, /u/; Lindblom, 1986). As a language emerges, perhaps over time the phonological space is divided into distinct categories that are sufficiently contrastive. If CTSL and ABSL are in the initial stages of a phonological systematization, it makes sense that the sample space is first divided into two maximally different units. Interestingly, the position of non-selected fingers (neither fully flexed nor extended) does not maximize perceptual contrast as has been argued for ASL (Brentari, 1998).

One explanation for the relative dearth of contrasts is that languages crowd the perceptual space when necessary to make distinctions among many lexical items (Jackendoff, 2002). The fact that there are few minimal pairs even in established sign languages (van der Kooij, 2002) suggests that the perceptual space may be larger in the signed than spoken modality. Perhaps the perceptual space is vast enough or the lexicons small enough that there is no need for more contrasts. We speculate that the number of contrasts may grow alongside the CTSL lexicon.

[1] The restrictions of the Symmetry Condition as Battison put it are somewhat different in classifier constructions of

Though established sign languages have a large inventory of selected finger groupings and generally obey a set of phonological constraints, we find a smaller inventory and loose adherence to phonological constraints. We take this as evidence that the phonology of CTSL has not yet become fully systematized. One interpretation is that CTSL is in the process of developing systematicity. Alternatively, compliance could simply reflect motoric limitations rather than linguistic demands. These data, especially paired with those from ABSL, suggest that these factors emerge in phonological systems and are not hard constraints on manual communication. That is, phonological systematicity may be driven by external factors (e.g., a sufficiently large lexicon, enough generations of users) rather than automatically granted to all sign systems.

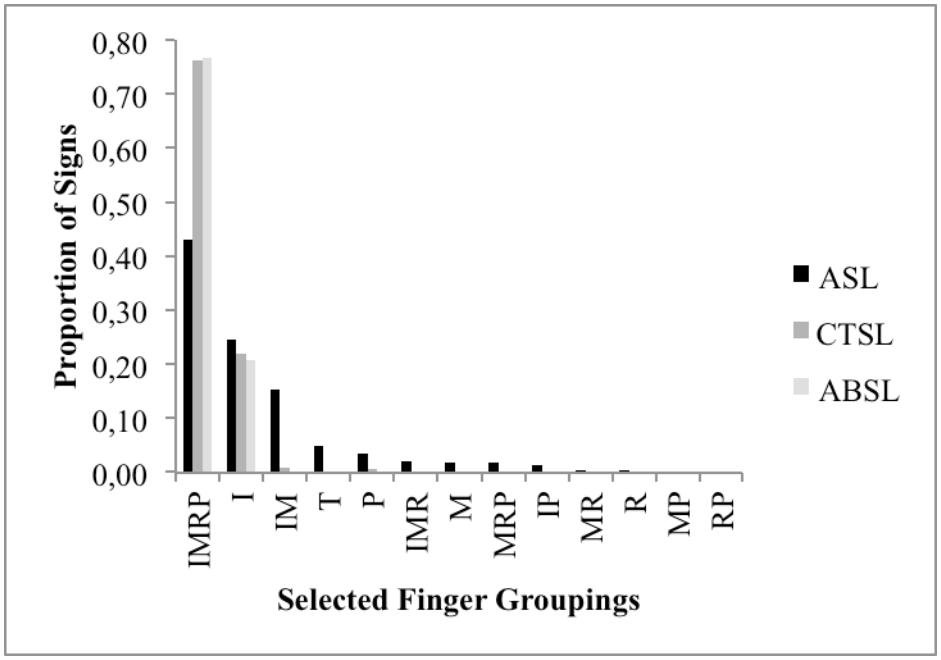


Figure 1. Distribution of finger groupings. I = index finger, M = middle finger; R = ring finger, P = pinky finger. Also note that 2% of ABSL signs were reported to use "other" groupings (Sandler et al., 2011), and thus are excluded from this figure.

Figure 2. Example of nonselected fingers (all fingers except the index) that are neither fully flexed nor fully extended.

Figure 3. Example of a violation of the Symmetry Condition in the sign CUCUMBER.

References.

Aronoff, M., Meir, I., Padden, C., & Sandler, W. (2003). Classifier constructions and morphology in two sign languages. In K. Emmorey (Ed.), *Perspectives on classifier constructions in sign languages* (pp. 53-84). Mahwah, NJ: Erlbaum Associates

Battison, R. (1978). *Lexical Borrowing in American Sign Language.* Silver Springs, MD. Linstok

Brentari, D. (1998). *A prosodic model of sign language phonology*. MIT Press.

Eccarius, P., & Brentari, D. (2007). Symmetry and dominance: A cross-linguistic study of signs and classifier constructions. *Lingua*, *117*(7), 1169-1201.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.

Lindblom, B. (1986). Phonetic universals in vowel systems. *Experimental phonology*, 13-44.

Sandler, W., Aronoff, M., Meir, I., & Padden, C. (2011). The gradual emergence of phonological form in a new language. *Natural language & linguistic theory*,*29*(2), 503-543.

Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(7), 2661-2665.

van der Kooij, Els. PhD dissertation. University of Leiden; 2002. Phonological categories in Sign Language of the Netherlands: The role of phonetic implementation and iconicity. Published by LOT.

Lexical access in sign perception: A computational model

Naomi K. Caselli & Ariel M. Cohen-Goldberg

Tufts University (Medford/Somerville, MA, US)

Psycholinguistic theories have predominantly been built upon data from spoken language, which leaves open the question: How many of the conclusions truly reflect language-general principles as opposed to modality-specific ones? We take a step toward answering this question in the domain of lexical access in recognition by asking whether a single cognitive architecture might explain diverse behavioral patterns in signed and spoken language.

Within the psycholinguistic framework, the comprehension of a single word ultimately involves mapping a physical signal onto its meaning and the production of a single word involves mapping meaning to a physical signal. Multiple stages of processing have been posited to take place in between these two endpoints, most generally the identification (or in production, the preparation) of sub-lexical and lexical units (e.g., McClelland & Ellman, 1986; Dell, 1986). According to a number of accounts, signed and spoken languages, like all languages, might have similarly organized semantic systems (e.g., Jackendoff, 2012). At the same time, their most peripheral elements clearly differ: signed languages utilize manual and facial articulators and are perceived through the visual system while spoken languages are produced with the oral articulators and are perceived through the auditory system. We ask whether the core mechanisms of sub-lexical and lexical processing might be the same despite differences in the most peripheral aspects of sign/word recognition.

Research on lexical access has revealed both parallels and discrepancies between sign and spoken word perception. In spoken word recognition, one of the most well-documented findings is that the number of words that are phonologically related to a target (its neighborhood density) inhibits recognition of the target word (Dufour & Peereman, 2003; Goldinger, Luce & Pisoni, 1989). Phonological neighborhood density also plays a role in sign processing. However, in a study of Spanish Sign Language (LSE), Carreiras et al. (2008) found that signs with many handshape neighbors ('dense handshape neighborhoods') are easier to identify in a lexical decision task than signs with few handshape neighbors. Meanwhile, signs with dense location neighborhoods are harder to identify than signs with sparse location neighborhoods.

How might we account for the apparent unique role of neighborhood type in sign language? One possibility is to assume that there are different computational principles at work in signed and spoken

language, leading to fundamental differences in the way words and signs are activated during language processing (e.g., Baus et al., 2008). The fact that it matters in sign language whether a neighbor shares its location or its handshape with the target suggests that there are sign language-specific retrieval mechanisms since there is no exact corollary of these parameters in spoken language. However, another possibility is that spoken and signed languages make use of the same core mechanisms to access the mental lexicon and it is a handful of relatively peripheral differences between modalities that accounts for the differences in the way neighbors affect processing.

Looking deeper into the effects of neighborhood density, there is a pattern of reversals in spoken and written language is not unlike the pattern of reversals in sign perception. In spoken production neighborhood density is facilitatory (Mirman, Kittredge, & Dell, 2010; Vitevitch, 1997, 2002), while in spoken perception neighborhood density is inhibitory (Dufour & Peereman, 2003; Goldinger et al., 1989). In visual word recognition neighborhood density is facilitatory (Andrews, 1992), except for high frequency words in which case neighborhood density is inhibitory (e.g., Grainger, O'Regan, Jacobs & Segui, 1989; Davis, Perea, & Acha, 2009). The effect of neighborhood density depends on modality even in spoken and written word processing.

Chen and Mirman (2012) presented a computational model of word processing that unified opposite effects of neighborhood density in speech production, perception, and written word recognition. They posit that lexical neighbors thus send both facilitatory *and* inhibitory activation to other lexical items. It is a balance of facilitation and inhibition that determines the net contribution of neighbors, and the balance can be tipped depending on how strongly the neighbors are activated. Strongly activated neighbors exert net inhibition on the target, and weakly activated neighbors exert net facilitation on the target. Chen and Mirman's theory of lexical access accounts for the pattern of reversals observed in spoken (and written) language with a single core lexical access mechanism, varying only the most peripheral elements across modality (e.g., the sequence of activation of sub-lexical units in speech perception and word recognition).

We present a computational simulation of neighborhood effects in sign perception that imports principles from Chen and Mirman's model. We test three reasons that location neighbors might be strongly activated (and thus exert net inhibition on targets) and handshape neighbors might be weakly activated (and thus exert net facilitation on targets). Location is identified earlier in perception than handshape (Emmorey & Corina, 1990; Grosjean, 1981), in the behavioral data on average locations were more frequent in the language than handshape, and on average there were more location

neighbors than handshape neighbors (Carreiras et al., 2008). We show that if a model containing these core principles is elaborated to incorporate relatively minor facts about either 1) the time course of sign perception or 2) the frequency of sub-lexical units, it produces data that match the experimental findings from sign languages. Interestingly, we were not able obtain the observed pattern of results when the number of lexical neighbors was similarly varied.

Our success in modeling the effects of location and handshape provides evidence that there may be universal principles governing the way the mental lexicon is accessed. Even though location and handshape are elements that are unique to sign languages, it appears that their influence on recognition can be modeled using the same principles that have been used to explain lexical access across tasks in spoken and written language. We wish to note that our results do not rule out the possibility that there are sign language-specific factors that influence lexical processing. They do, however, indicate that such factors are not necessary to account for the empirical data on neighborhood effects.

References.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 234. doi: 10.1037/0278-7393.18.2.234

Baus, C., Gutiérrez-Sigut, E., Quer, J., & Carreiras, M. (2008). Lexical access in Catalan Signed Language (LSC) production. *Cognition*, *108*(3), 856-865. doi:10.1016/j.cognition.2008.05.012

Davis, C. J., Perea, M., & Acha, J. (2009). Re(de)fining the orthographic neighborhood: The role of addition and deletion neighbors in lexical decision and reading. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1550–1570. doi: 10.1037/a0014253

Dell, G. S. (1986). A spreading activation theory of retrieval in language production. *Psychological Review, 93*, 283-321. doi: 10.1037/0033-295X.93.3.283

Dufour, S., & Peereman, R. (2003). Inhibitory priming effects in auditory word recognition: When the target's competitors conflict with the prime word. *Cognition*, *88*(3), B33-B44. http://dx.doi.org/10.1016%2FS00100277%2803%2900046-5

Carreiras, M., Gutiérrez-Sigut, E., Baquero, S., & Corina, D. (2008). Lexical processing in Spanish Sign Language (LSE). *Journal of Memory and Language*, *58*(1), 100-122. doi: 10.1016/j.jml.2007.05.004

Chen, Q., & Mirman, D. (2012). Competition and Cooperation Among Similar Representations: Toward a Unified Account of Facilitative and Inhibitory Effects of Lexical Neighbors. *Psychological Review*, *119*(2), 417. http://dx.doi.org/10.1037%2Fa0027175

Emmorey, K., & Corina, D. (1990). Lexical recognition in sign language: Effects of phonetic structure and morphology. *Perceptual and Motor Skills*, *71*(3f), 1227-1252. http://dx.doi.org/10.2466%2FPMS.71.7.1227-1252

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, *28*(5), 501-518. http://dx.doi.org/10.1016%2F0749596X%2889%2990009-0

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception & Psychophysics, 45,* 189–195. doi: 10.3758/BF03210696

Grosjean, F. (1981). Sign and Word Recognition: A First Comparison. *Sign Language Studies*, *32*, 195-220. http://dx.doi.org/10.1353%2Fsls.1982.0003

Jackendoff, R. (2012). *User's Guide to Thought and Meaning*. New York: Oxford University Press.

Mirman, D., Kittredge, A. K., & Dell, G. S. (2010). Effects of near and distant phonological neighbors on picture naming. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1447-1452). Austin, TX: Cognitive Science Society.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86.
http://dx.doi.org/10.1016%2F00100285%2886%2990015-0

Vitevitch, M. S. (1997). The neighborhood characteristics of malapropisms. *Language and Speech*, *40*(3), 211-228. http://dx.doi.org/10.1121%2F1.415242

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *28*(4), 735.

Spoken and gesture expression of Manner-incidental and Manner-causal events in English and Italian

Federica Cavicchio[1], Sotaro Kita[2]
[1]CIMEC, Università di Trento, [2]University of Warwick

From the seminal work of Kita and Özyürek (2003), many studies have investigated manner and path in the verbalization of motion events and the co-produced manual gestures. Fewer studies though have investigated motion events with respect to the relationship between Manner and Path, i.e. whether Manner and Path are causally related or not (Kita et al., 2007; Brown et al., 2005). In this study we focus on the effects that manner inherent events (manner of movement causes path) vs. manner incidental events (manner of movement and path bear no causal relationship) have on verbal and gesture expressions in English and Italian. Following Talmy's (1985) typology, English is a satellite framed language. When Manner and path are realised with one verb and a satellite, manner and path are conflated (e.g. *roll down*). On the other hand, Italian is considered a verb framed language, that is, manner and path are realized by two verbs. Nevertheless, satellite framed constructions such as "*rotola giù/su*" (rolls down/up) can be used by Italian native speakers.

Following Kita and Özyürek (2003), it is expected that satellite framed verbal constructions will be accompanied by conflate gestures, combining the information about manner and the path in one movement. As regards 2 clauses verbal constructions, they will be accompanied by 2 gestures, one expressing manner and one expressing path. To test this we used Özyürek et al. (2001) stimuli. In these short clips, two characters (Tomato and Triangle) perform actions involving manner and path (e.g rolling down a slope). Half of the stimuli depicted target motion events in which the manner was inherent to path. That is, manner causes or contributes to change of location. The remaining half depicted target motion events in which manner was incidental to path. Özyürek et al. stimuli can be described with a satellite framed verb (1 clause, e.g. *rolls up*) or a verb framed construction (2 clauses, e.g. ascends *or* goes up as it rolls).

40 participants (20 native Italian speakers, 20 English) saw the 10 videos. They were asked to describe each of them to a language matched listener.

To investigate the effect of verb type, a mixed effect model with verb type (1 clause vs 2 clauses) as dependent variable and fixed effects for language (English vs Italian), gestures (conflate vs 2 manner

and path), item (manner-inherent vs manner-incidental), and random intercepts for subjects and items and random slopes for item was run.

The first model showed that overall, to describe the stimuli Italian native speakers chose 2 clauses verbs over one clause verbs (Est.=3.9, SE=0.7, p<.001).Italian and English speakers are more likely to produce satellite framed verbs with manner-inherent events (Est.=1.8, SE=0.6, p<.002).

To investigate the effect of gesture type, two mixed effect logistic regression models, one with conflate gestures as dependent variable and one with manner and path gestures as dependent variable were run. The models had fixed effects for language (English vs Italian), structural tightness (1 clause vs 2 clauses), item (manner-inherent vs manner-incidental), and the same random structures: random intercepts for subjects and items and random slopes for item.

As regards the second model, speakers were more likely to use conflated gestures when they produced 1-clause expressions (Est.=3.4, SE=0.6, p<.001) in both languages. Language has no effect on conflate gestures (Est.=0.3, SE= 1.1, p=.7). Conflate gestures are more likely to be used with stimuli in which manner is incidental (not causally related) to path (Est.=0.9, S.E.= 0.4, p<.04).

As regards the third model, a reverse pattern of results was found. Speakers were more likely to use manner and path gestures when they produced a 2-clause verbal expressions (Est.=3.6, S.E=0.5, p<.0001). Language had not effect on manner and path gestures (Est.=-0.4, S.E=-0.5, p=.4). Manner and path gestures were more likely to be produced with stimuli in which manner is inherent to path (manner causes or contributes to change of location; Est.=0.7, S.E=0.3, p<.03).

As in Kita et al. (2007) for English, the event-type manipulation successfully elicited both one-clause and two-clause descriptions of manner and path in Italian speakers, with one-clause descriptions more common for manner-inherent events. More crucially, in both Italian and English, one-clause descriptions elicited conflated gestures, even when the effect of event-type is controlled for in the analysis. The information packaging in speech is mirrored by the information packaging in gesture. We thus extended Kita et al.'s finding for English to a verb-framed language, Italian. These findings support the hypotheses that representational gestures arise from the interface between spatio/motoric events and language and that conceptual message representations and syntactic representations are generated interactively during speaking (Kita & Özyürek, 2003).When a language is flexible enough to allow for a satellite framed constructions (as Italian), speakers will choose a tighter verbal construction to accompany manner-inherent movements. Our results support the idea that conceptual message

representations and syntactic representations are generated interactively (Kita & Özyürek, 2003; Viglioccо & Kita, 2006).

References.

Brown, A., Özyürek, A., Allen, S., Kita, S., Ishizuka, T., Furman, R. (2005). Does event structure influence children's motion event expressions? *Online Proceedings Supplement of the 29th Annual Boston University*. http://www.bu.edu/bucld/proceedings/supplement/

Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., Ishizuka, T. (2007). Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. Language and Cognitive Processes, 22 (8), pp. 1212-1236.

Kita, S., Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and language*, 48 (1), pp.16-32.

Ozyürek, A., Kita, S., Allen, S. (2001). *Tomato Man movies: Stimulus kit designed to elicit manner, path and causal constructions in motion events with regard to speech and gestures*. Max Planck Institute for Psycholinguistics Language and Cognition group. Nijmegen, The Netherlands.

Talmy, L. (1985). *Lexicalization patterns: semantic structure in lexical forms*. In: Shopen, T. (Ed.), Language Typology and Syntactic Description 3: Grammatical Categories and the Lexicon. Cambridge University Press, Cambridge, pp. 57–149.

Viglioccо, G., Kita, S. (2006). Language-specific properties of the lexicon: Implications for learning and processing. *Language and Cognitive Processes*, 21 (7-8), pp. 790-816.

Alignment between Gesture and Intonation in Narrative Production (in progress)

Hua Chen[1], Han Zuo[1], Yan Gu[2]

[1]Department of Applied Foreign Language Studies, Nanjing Univesity, People's Republic of China

[2]Tilburg Center for Cognition and Communication, Tilburg University, the Netherlands

The alignment between gestural and intonational structures is cited as premier evidence to the hypothesis that gesture and speech originate from the same conceptualisation. However, the pertinent research has been largely limited to first languages (Kendon, 1983; Loehr, 2004; McClave, 1991; McNeill, 1992), sparing second language context. The current study depicts the alignment between gesture and intonation in terms of prominence, phrasing, and tone in first and second English narrative production. Both native speakers of English and Chinese learners of English were asked to participate in a story recall task to elicit data. Overall, the pilot study found that gesture prominence (stroke) aligns with intonation prominence (pitch accent) in L1 English production with stroke preceding pitch accent for 285 milliseconds on average, while in L2 English production, stroke precedes pitch accent for an average of 461 milliseconds. It also found that in L1 production, gestural phrases tend to align with intonational phrases, whereas in L2 production, more cases of gestural phrases crossing the boundaries of intonational phrases were detected. Similar to Loehr's study (2004; 2007), we did not discover any paralleled rising and falling movements between gesture and pitch as claimed by Bolinger (1983). These preliminary findings suggest that interactive pattern between gesture and intonation in L2 production is more complicated. With L2 speakers encountering more lexical and expressive difficulties during speech, dysfluencies arise with a higher frequency. The time needed for speech processing is greatly increased, hence, an enlarged onset gap between stroke and pitch accent. Examination of L2 dysfluent and fluent speech is necessary for a better understanding of gesture-intonation alignment.

Language-specificity in the McGurk effect?: what mismatches in audio-visual speech perception reveal about inventory constraints

Faith Chiu[1], Shanti Ulfsbjorninn[2]

[1]University College London, [2]SOAS – University of London

This paper explores the relationship between spoken-language phonology, in particular the language's phonemic inventory, and non-verbal speech perception. We will consider language-specific variation in performance of the McGurk effect in Modern Standard Arabic against Standard Southern British English speakers (as controls). We investigate the influence of phonology on the perception of audio-visual speech cues. Using novel experimental results, we will answer the following questions focusing particularly on (4).

(1) Is there modality (in)dependence in phonology?

(2) Do phonological primitives in the audio modality affect perception in another modality?

(3) How are audio and visual cues integrated, is such integration affected by any language-specific phonological constraints?

(4) Is the inventory of phonemes, the set of what are considered contrastive phonemes, a factor in audio-visual mismatches? (What if the McGurk effect should be expected to target sounds which are not part of the phonemic inventory?

Speech is inherently a multimodal phenomenon (Rosenblum, 2008). Where there is incongruence in visual and auditory speech information, the McGurk effect demonstrates a visual influence to clear auditory information (McGurk & MacDonald, 1976). English-speaking subjects presented with a discrepant visual dubbed over an auditory syllable can report hearing syllables a fusion of both perception channels (/da/ or /ða/ for visual /ga/ dubbed over audio /ba/). Regarding language-specificity, the McGurk effect has been found in English, German, Italian and Spanish but to a lesser extent in Japanese and Mandarin Chinese (Bovo, Ciorba, Prosser & Martini, 2009, and references therein). Some have credited this to extra-linguistic, cultural factors (*i.e.* rudeness in face-reading in Japan), but we attempt to consider a more language-based analysis.

The reported experiment presents mismatched audio-visual stimuli to native Arabic monolingual speakers. Following McGurk & MacDonald (1976), the stimuli contain a token set: /ba/, /da/ and /ga/, both matched and mismatched in its audio and visual presentation read aloud by a speaker in a video. The video may show a mouthing of the token /ba/ either presented with a matching /ba/ audio, or a

mismatched /ɡa/ audio (or any combinations thereof). Modifying McGurk & MacDonald's initial tokens, we also add a voiceless unaspirated series, /pa/ /ta/ /ka/ as well as the voiceless and voiced uvular stops /qa/ and /ɢa/. Subjects will be asked to report what they hear, hence there is no limitation in options in terms of voicing and place of articulation. These reports of the heard sound will be recorded, analysed acoustically using Praat before categorisation and transcription independently then cross-checked across both authors (trained in IPA).

The stops of Modern Standard Arabic are plain unaspirated and voiced, /t k q/ and /b d/. There are gaps within this phonemic inventory where there is no voiceless bilabial stop /p/, no voiced velar stop /ɡ/ and no voiced uvular stop /ɢ/. In the experimental block where subjects are presented with tokens matched in audio and visual modalities, we record an accuracy coefficient first to understand how well Arabic speakers can perceive /pa/ /ɡa/ and /ɢa/. After recording the standard level of (in)accuracy, subjects are presented with mismatched tokens in all conditions mixed randomly with matched tokens. We would investigate if Arabic speakers, when presented with a mismatched sequence where one of the phonemes is missing in his/her native inventory (*e.g.* visual /ka/ with audio /pa/, or visual /ga/ with audio /ba/), will similarly derive the predicted /ta/ or /da/ as per other language-speakers, or switch its voicing category where visual influence is dominant (*e.g.* reporting /ta/ instead in a visual-dominant mismatched visual /ga/ with audio /ba/ pairing, since there is only [k] in Arabic). Furthermore, we consider if jaw openness is a scalar property and if it can affect the output of the McGurk effect. *e.g.* Does visual /qa/ with audio /pa/ produce /ka/ or /ta/?

This is important to inform us as to the status of the phoneme in phonological representation, both in spoken language phonology and (audio-)visual speech perception. The implications of this experiment crystallises earlier attempts in phonological theory to discredit the phoneme as a valid category. Modern phonological frameworks may not discuss the status of the phoneme explicitly, but by their rules/generalisations make reference only to sub-segmental features; the phoneme is a *de-facto* 'dodo'. These theories split the phoneme into features or elements and leave no space for the grammar to refer to phonemes as objects (Chomsky & Halle, 1968; Anderson & Jones, 1974; Harris & Lindsey, 1995).

The impact of event structure on emerging gestural communication systems

Peer Christensen[1] and Kristian Tylén[1,2]

[1]Center for Semiotics, Aarhus University, Denmark

[2]The Interacting Minds Center, Aarhus University, Denmark

Linguistic structure emerges and evolves in response to a multitude of cognitive, social and environmental pressures (Tylén, Fusaroli, Bundgaard, & Østergaard, 2013). As a medium for communication, nonverbal gesture lends itself as a particularly interesting window to conceptualization processes underlying word order due to its spontaneous and non-conventionalized nature. Previous studies using gesture elicitation tasks have indicated that people of any linguistic background will use only one specific gesture order (SOV – subject, object, verb) when asked to describe transitive events using only gesture (Goldin-Meadow, So, Ozyurek, & Mylander, 2008; Langus & Nespor, 2010). Drawing on the existing literature on emerging sign languages, home sign and language typology and change, these findings are presented as evidence for innate biases in the conceptualization and representation of events, thus transcending acquired linguistic structure.

We offer a competing explanation proposing that gestural representations are influenced by the event structure of the referent situations themselves. Whereas the previous studies have only considered what we call *object manipulation events*, in which agents perform actions that involve and affect objects, we add a novel category to the event stimuli called *object construction events*. We make the simple observation that in the former event type, actions are contingent upon the presence of the objects to be manipulated. In such cases, objects must logically precede the performed actions. Conversely, object construction events present a special case in which the object is dependent on the action being performed. In other words, in object construction events, the action logically precedes the object.

We will present data from three experiments in which pairs of participants engaged in a *referential game* (Fay, Garrod, & Swoboda, 2010), jointly matching stimulus pictures using only gesture as a means for communication. This paradigm, which was developed in the field of experimental semiotics (see e.g. Galantucci & Garrod, 2011 for a review), offers various methodological and ecological advantages in comparison with the standard elicitation task. In particular, the new paradigm engages participants in actual bi-directional communication thus creating a need for mutual comprehension of a co-developed and shared gestural communication system. The first experiment used a blocked design

so that participants produced gestures to exclusively describe one of the two event types before moving on to a second condition featuring a new event type. In the second experiment, the event types were balanced and randomly mixed. This experiment was devised specifically to test additional hypotheses concerning the possible effects of social-interactional alignment within pairs. In the third experiment, we altered the relative frequency of the distribution of the two event types, so that half of the pairs predominantly communicated stimuli featuring object manipulation events, while the other half communicated stimuli with a greater frequency of object construction events. This latter experiment was conducted to investigate whether further pressure on the gestural communication systems could potentially motivate a process of conventionalization of a single order for communicating about both event types.

Our results indicate that event structure is the main factor motivating structure in gestural communication systems emerging in a lab setting. Indeed, the event structure inherent in the referent situations serves to predict the gestural 'word order' found in the novel communication systems. Consistent with the previous studies, object manipulation events (equivalent to transitive events, e.g. 'a ballerina throwing a sweater') motivated SOV structure. However, object construction events (e.g. 'a ballerina painting a sweater') motivated SVO, which cannot be explained by the proposal that nonverbal gesture strings are governed by a cognitive bias favouring SOV structure. We argue that these representational orders reflect the two event types through *structural* or *diagrammatic iconicity,* meaning that the order in which individual signs are produced resembles the inherent structure of the stimulus events (see e.g. Sandler & Lillo-Martin, 2006:496). In other words, the structure of the emerging communication system is shaped by the experiential structure of the particular phenomena being communicated. In addition, we discuss how social communicative pressures and the frequency distribution of event types may work to stabilize (and possibly conventionalize) a particular gesture order.

References.

Fay, N., Garrod, S., & Swoboda, N. (2010). The interactive evolution of human communicative systems. *Cognitive Science*, 34, 351-386.

Galantucci, B., Garrod, S. (2011) Experimental semiotics: a review. *Frontiers in Human Neuroscience, 5*(11): 1-15.

Goldin-Meadow, S., So, W. C., Ozyurek, A., & Mylander, C. (2008). The natural order of events: how speakers of different languages represent events nonverbally. *PNAS*, 105, 9163-9168.

Sandler, W. & Lillo-Martin, D. (2006) *Sign Language and Linguistic Universals.* Cambridge, UK: Cambridge University Press.

Tylén, K., Fusaroli, R., Bundgaard, P., & Østergaard, S. (2013). Making sense together: a dynamical account of linguistic meaning making. *Semiotica*, 194, 39-62.

Voice and sense in a subjective assessment of television broadcast simultaneous interpreting

Gregorio De Gregoris, University of Trieste

The present PhD research project is a corpus-based interpreting study aiming to (i) detect how much perception of vocal non-verbal parameters may influence perception of verbal ones in subjective assessment of TV broadcast simultaneous interpreting (SI); (i) what is the relationship between holistic (or gestaltic) perception and analytic perception of evaluation parameters, if any. Subjective evaluation is made through a questionnaire-based survey which respondents have to answer after watching and listening to video excerpts of TV broadcast SI. These are taken from English-Italian and English-Spanish interpretations of the 2012 US presidential debate, a sub-corpus of the *CorIT*, the Television Interpreting Corpus developed at the University of Trieste. Data collected from the web-based questionnaire survey will be compared with those of transcripts, since it is not possible to carry out an acoustic analysis of excerpts of interpreted speech; in fact, it is not possible to detach the interpreter's voice from the one of the speaker, because they are merged in a single mono track. A questionnaire for subjective assessment of TV broadcast interpreting has been designed for a pilot study. Evaluation parameters have been ordered in such a way to help respondents to move from sound or phonic perception to sense construction. The questionnaire has been devised and will be administered – through a web-based survey software platform – to professional interpreters, interpreting students, musicians and actors. The questionnaire includes 3 video excerpts (1 min each) and 3 sets of the same questions for each video excerpt; plus a final block of questions on personal data. Since the corpus described above has not yet been completely transcribed, video excerpts for the pilot study have been selected from the Italian interpretations of 2008 US Presidential Debates (Obama vs. McCain). In order to better study perception of vocal non-verbal features in relation with the verbal aspects of speech, an experimental variable has been introduced. One video excerpt has been artificially manipulated by replacing the interpreter's audio track with the recording of a professional dubbing speaker's performance (voice-over technique), i.e. an interpretation of the original interpreter's text, read by an actor-speaker. The speaker is also a famous dubber in Rome's post-production studios: his voice and his speaking style (*phonostyle*) is quite appreciated by radio and TV professionals. The aim of such a manipulation is to create an audio-visual sample where the vocal features could be detached from the sense conveyed.

The results of the pilot study will pave the way to the definitive study, i.e. a web-based questionnaire survey with excerpts from the 2012 US Presidential Debates (Obama vs. Romney), and administered to the same subjects as the pilot study, plus TV professionals and ordinary TV viewers. The theoretical approach of this study takes into account Michel Chion's notion of "audiovision" (1990), which consists of a unique, global perception, and not the mere result of "sound+image". According to Chion, when we perceive something that is audiovisual, the sound projects itself onto the image and the other way round, creating an "illusion", a "suggestion", a kind of magic. Therefore, "audiovision" involves a "trans-sensory perception, i.e. a perception that belongs to no one particular sense, but which may travel via one sensory channel or another […] Everything involving rhythm may serve as an example" (Chion 2012:13). From the linguistic point of view, this study is based on Albano Leoni's proposal of speech perception called 'phonic facet of words' ("*volto fonico delle parole*"), where "the linguistic unit of perception and processing is a phonological word or a word group or any other significant unit grasped in its essence in discourse" (Albano Leoni 2009: 165). As the author recognizes, the notion of "phonic facet of words" is not new, in fact it was first introduced by the German psychologist and linguist Karl Bühler (1931, 1983; in Albano Leoni 2009: 166); nonetheless, it was not taken into account by phonology at that time (Albano Leoni 2009: 94). Since such a model does not allow segmentation, there is no distinction between segmental and suprasegmental features, or linguistic and paralinguistic aspects. The features of the "phonic facet of words" are "voice", "syllable" and "prosody" (Albano Leoni 2009: 183); other relevant aspects of this model are "sense" and "context". Albano Leoni's "phonic facet of words" is not so far from Meschonnic's theory of "rhythm in language", where he considers rhythm as the "form" of discourse, drawing on the original definition of rhythm proposed by Benveniste (1966: 327-335), i.e. rhythm < gr. *ῥυθμός* < *ῥειν* ("to flow"), where *ῥυθμός* means "la forme dans l'instant qu'elle est assumée par ce qui est mouvant, mobile fluide, la forme de ce qui n'a pas consistence organique" (Benveniste 1966: 333). As Meschonnic states (1982: 70), from Benveniste on, rhythm may no longer be considered secondary to form, since it means "organisation (disposition, configuration) d'un ensemble". Rhythm is the form of a language, the way discourse is organized. Meschonnic identifies the "discourse" with the "rhythm", considering the latter as an "ensemble synthétique" (1982: 216) of the elements that make up the discourse. This approach is in line with Fónagy's "vive voix" (1983), expressing the speaker's instinct, that is a pre-verbal feature; therefore, "vive voix" means 'voice of the body', 'vocal gesture', being the basic, original component of oral communication.

The role of gesture and prosody in children's multimodal pathway into negation

Christelle Dodane[1], Aliyah Morgenstern[2], Pauline Beaupoil[2], Alessandra Del Ré[3], Dominique Boutet[4]

[1]Praxiling UMR CNRS 5267 CNRS/Montepellier 3, [2]Paris 3, Sorbonne Nouvelle

[3]UNESP, Araquara, [4]Structures formelles du langage UMR CNRS/Paris 8

Through constant exposure to adult input in dialogue, children's language gradually develops into rich linguistic constructions that contain multiple cross-modal elements subtly combined for coherent communicative functions. Prosody and gesture in particular both facilitate children's entry into language. Children not only demonstrate sensitivity to the intonation and rhythm of their native tongue from birth (Mehler et al., 1978; Jusczyck, 1998), but they also use the prosodic level to structure their first productions, especially after 9-10 months. Similarly, gestures play a crucial role in children's entry into symbolic communication. Balog and Brentari (2008) have shown that since the single-word period, children coordinate their verbal and nonverbal behavior, which makes their meaning more comprehensible, just like adults do (Bolinger, 1983; Cruttenden, 1997). Combining prosody and gesture allows children to overcome difficulties in mastering the phonological system, and to enter syntax thanks to early multimodal constructions.

The expression of negation is a privileged locus to study multimodal combinations of gesture with prosody, syntax, semantics and pragmatics. There is a cross-modal continuity in the expression of negation speech acts, which are first expressed by gestures, then by speech (Bates, Camaioni and Volterra, 1976; Bates et al., 1979). From the end of their first year on, children can express negation with headshakes, index waves or palms-up gestures. Prosody and gestures are also combined to express refusals, protests or epistemic negations long before the emergence of the first verbal negation markers (around six months later). It is therefore crucial to analyze gestures and prosody with an integrative approach.

The goal of this study is two-fold. We determine to what extent children combine body movements, symbolic gestures and prosody to express their communicative intentions more efficiently when they express negation, and we analyze the respective weight of each modality in language development.

We analyzed the longitudinal recordings of a monolingual French girl recorded monthly for one hour between the ages of 1;02 and 2;09 (MLU 1.1 to 4.3) in spontaneous interaction with her parents (*Paris Corpus*, Morgenstern, 2009; Morgenstern and Parisse, 2012). Syntactic development was determined by calculating the MLU (Brown, 1973) and lexicon size, by counting the number of

different words produced during each session (Vihman, 1985; Vihman & Miller, 1988).

Our study focuses on the 96 multimodal productions containing the word "non" (no) produced in isolation and on strings of reduplicated "non"s. Three types of analyses were conducted. First we coded prosodic properties (direction of the intonation contour, accent range, register, duration, intensity), using the software PRAAT (Dodane and Massini-Cagliari, 2010). Second, we coded nonverbal behavior (hand gestures, joint attention expressed through eye gaze and checking behavior, body movement and facial expressions), using the ELAN software. Third, we compared the prosodic and gestural analyses to look for directional and temporal synchronization patterns.

At the prosodic level, results showed that the first vocal "no" emerged around 14 months in a reduplicated form and was exaggerated at the prosodic level. Between 14 and 21 months, it was mainly realized with rising intonation contours and increased syllabic duration. Between 22 and 25 months, it was mainly produced with rise-fall intonation contours and finally, from 26 to 28 months, with flat or falling intonation contours and reduced syllabic duration. Such an evolution seems to reflect a better control in the expression of negation as of 25 months.

At the non-vocal level, body movements were most often produced in coordination with verbal production and their direction was mostly synchronized with the direction of intonation contours (rising contours with rising gestures) between 14 and 19 months. The more the child expressed protests against adults, the more she exaggerated both her prosody (higher accent range, register, intensity and duration) and her body movements. After 19 months, she used mostly upper-body gestures and movements (head, chest) with a majority of forward and backward or oscillating movements in close parallel with her prosodic contours. As her mastery of speech developed, she gradually stopped exaggerating her prosody and resorted less to non-verbal behavior.

Gestures, body movements and prosody provide powerful resources that the child integrates to make her multimodal entry into language. If children use each modality (vocal and visual) more and more skillfully thanks to adults' scaffolding in everyday life interactions, both modalities actually develop together. This study therefore gives us insights on how children become experts in face-to-face social interaction, which is multimodal in nature.

References.

Bolinger, D. (1983). Intonation and gesture. American Speech, Vol. 58, 2, pp 156-174.

Balog, H. and Brentari, D. (2008). The relationship between early gestures and intonation. First

Language, 28, pp. 141-163.

Bates, E., Camaioni, L. and Volterra, V. (1976). Sensorimotor performative. In: Bates, E. (Ed.). Language and Context: the Acquisition of Pragmatics. New York: Academic Press.

Bates, E. (1979). Emergence of symbols in language and action: similarities and differences. Paper and Reports on Child Language Development, Stanford, n.17, p.106-118.

Brown, R. (1973). A first Language. Cambridge: Harvard University Press, 437 p. Cruttenden, A. (1997). Intonation (2nd ed.). Cambridge: Cambridge University Press.

Dodane, C.; Massini-Cagliari, G. (2010). La prosodie dans l'acquisition de la négation : étude de cas d'une enfant monolingue française. In ALFA, Revista de Linguística, v.54, 2, 335-360.

Jusczyk, P. W. Dividing and conquering linguistic input (1998). In: Gruber, M. C.; Olson, K.; Wysocki, T. (Ed.). CLS 34. The Panels. v. 2, Chicago, p. 293-310.

Mehler, J., Bertoncini, J. and Barriere (1978). M. Infant recognition of mother's voice. Perception, Bristol, n.7, p.491-497.

Morgenstern, A. (2009). L'enfant dans la langue. Presses de la Sorbonne Nouvelle, Paris Morgenstern, A & Parisse, C. (2012). The Paris Corpus. French Language Studies, 22(1), 7-12, Cambridge University Press.

Vihman, M. (1985). Language differentiation by the bilingual infant. Journal of Child Language, 12, 297–324.

Vihman, M. & Miller, R. (1988). Words and babble at the threshold of language acquisition. In M. D. Smith & J. L. Locke (Eds), The emergent lexicon: The child's development of linguistic vocabulary (pp. 151–183). San Diego, CA: Academic Press, Inc.

The Functional Grounding of Phonetic Variation Patterns in Spontaneous German

Pepe Droste
University of Münster, University of Paderborn

The phonetic realizations of phonological categories are characterized by systematical variation. Amongst the challenges to the research fields of contemporary phonetics and phonology is the identification of relevant factors that condition phonetic realization and phonological categorization, enlightening the complex relation between cognitive representation and phonetic implementation. Despite the obvious relevance for human communication, the phonetics of spontaneous speech is still situated on the periphery of the phonological and phonetic research agenda (e.g., Kohler 2000, 2001; Johnson 2004; Ernestus & Warner 2011). For a long time, at the most, lab speech served as the empirical base of phonetic and phonological theory, leading to the 'phonetics and phonology of the optimal code'. However, in spontaneous speech the phonetic realizations may involve deviations from the optimal code. Compared to the culturally construed canonical citation forms whole syllables are deleted and/or phones are changed. For instance, in the case of the German lexical item haben 'have' investigators of spontaneous speech usually do not find the canonical citation form [haːbən] proposed by common pronunciation dictionaries, but probably pronunciation variants like [haːbm̩, haːm̩, haːm, ham] or maybe even [ha] (see Lanwer forthcoming). There is little evidence that phenomena of reduction hinder interactional communication. Rather, phenomena of the speakers' complexity groove seems to fit functional purposes limited by structural conditions (e.g., Lindblom 1990; see Jurafksy et al. 2001, 2002; Plug 2005; 2011; Hay & Bresnan 2006; Gahl 2008; Drager 2010, 2011).

This paper reports on word-specific phonetic variation observed in German political talk show data. Phonetic variability here is primarily defined in terms of syllabic complexity, as phenomena of reduction and elaboration can be projected onto the articulatory dynamics in global opening and closing gestures, constituting fundamental phonetic structures of speech communication. In-depth analysis was conducted on tokens of the verb *haben* 'have' from the participants' speech. In mixed-effects models trends are modeled related to factors of speaking rate, phrasal accent, lexical subcategory, word order, and speaker identity potentially influencing phonetic complexity.

The results provide evidence that the phonetic realizations of *haben* 'to have' vary gradiently in terms of their different functions. The results show that factors like speaking rate and speaker identity poorly

determine processes of reduction and elaboration. Rather, the realizations vary depending on a combination of a word's grammatical function, its prosodic and morphosyntactic features.

The results hint saliently at the functional grounding and the gradient nature of linguistic knowledge in several aspects. (1) The results negate a monastic isolation of linguistic levels, showing a close interaction of the segmental and suprasegmental level as well as the phonological and morphosyntactic level. (2) Cognitively, the results challenge the view of a single phonological representation for polyfunctional words showing partial homophony. The current findings lend support to exemplar-based models of language which conceptualize the mental lexicon as storing phonetically rich detailed, multidimensionally categorized items (Johnson 1997; Pierrehumbert 2001, 2002; Bybee 2001).

The study demonstrates both richness and gradience of linguistic knowledge that is fundamentally functional motivated. In social life this richness and gradience provides an substantial resource for the construction of social identity and interactional purposes via social-indexical knowledge (Foulkes & Docherty 2006; Foulkes 2010; Docherty & Foulkes 2014). Ranging from linguistic core domains to more discourse analytic fields, the study of variation in spontaneous speech eminently contributes to the enterprise of understanding how human communication works.

Gestures and second language teaching of Medical University students: A study

Bistra Dudeva

Medical University, Sofia

The study is based on the essential relationship of speech and gestures in oral communication, on Kendon's (1982) points and McNeill's (1992) classification of gestures. McKneill's characteristics of gesture and its relationship to speech, to linguistic properties, to conventions and to semiosis are discussed to emphasize the different manners of structuring meaning through gesticulation and speech. The gesture expresses meaning synthetically and thus, connects as a whole to the meaning of the speech utterance.

The significance of language and gestures for the practice of the medical profession is revealed. First, the medical profession is predominantly practiced through verbal and most often face- to- face communication, accompanied by gestures. Second, the patient's rights to information require the medical professional to communicate respectfully with the patient. Third, patients are psychologically and emotionally vulnerable and a discrepancy between the meanings of verbal expression and gestures would decrease compliance with the doctor's treatment and advice.

In the native language there occur complex interactions between gestures and speech. They also manifest themselves in second language (L2) learning and teaching. Teaching L2 employs different approaches to reveal the linguistic interpretation of an utterance in L2 and the information of the message transmitted. However, in general little attention is dedicated to revealing or acquiring the natural gestures of L2 in the classroom.

The concept of 'gesture transfer' from L1 to L2 has already been studied by researchers such as Maria Grazia Busà and Luca Rognoni. However, this transfer is especially of interest for the medical profession. Foreign language teaching at the medical university is to manifest gestures significance in oral communication.

Aim of the study

The aim of the present study is to investigate the medical university students' ability to acquire and accompany the use of L2 phrases with the appropriate gestures (G2), characteristic of the culture where L2 communication is considered native, that is, culture 2.

Materials and methods

The main methods used in the study are statistical methods for processing the data about the participants and the frequency of occurrence of different combinations between the verbal phrases and gestures. The comparative method is used to compare the students' verbal skills and gestures with those of the original interlocutors in the dialogues, presented in the films, used for teaching purposes.

The study is carried out with 42 medical university students, studying L 2 to use it for professional purposes. The majority of the students are at higher intermediate level of L2. Forty of the participants are 18 or 19 years old and only two are 20 years old.

English is the L2 language in the study. We use a high frequency everyday native language (L1) phrase and its corresponding L2 phrase. In natural speech they are accompanied by well-established but distinctly different gestures- gesture 1 (G1) in culture 1 and gesture 2 (G2) in culture 2.

The study has followed several steps:

1. Medical university students watch short films of dialogues in L1 and L2 between medical professionals and patients. Students are asked to pay attention to the target phrase and the gesture, connected to it.
2. The L2 teacher discusses with the students the target phrase in L1 and L2. The teacher directs the students' attention to the combinations L1 phrase+ G1 and L2 phrase+ G2. The students are made aware of the expectations of L2 native speakers to witness L2 phrase used with G2 when communication is carried out in the patients' native L2.
3. The students are instructed to participate in a similar dialogue in L2 and try to accompany L2 phrase with G2.
4. The students make the dialogues and the combinations of phrases and gestures are recorded.
5. The results are analysed and discussed with the students.

Results and discussion

Of the 42 respondents 22 students used L2 phrase with G 1 gesture. A high number, 18 students, used L2 phrase without any gesture. They were rather stiff while doing the dialogue and were obviously conscious of their body movements and put a check on them for fear they would use G1 instead of G2. Only 2 students used L2 phrase with G2. Both of them were proficient in L2 and one had lived in a country where L2 is the native language.

The results show the predominant use of the combination L2 phrase +G1. Both L1 and G1 are connected to meaning and culture but the study shows L1 phrase is much more easily replaced by L2

phrase in speech. G1 was much more difficult to substitute by G2 than L1 phrase by L2 one. This is evidence of the strong rigidity of the combination of target phrase's meaning + G1.

The stability of G1 suggests several conclusions. Though in natural communication G1 is used along with L1 phrase and G2 – with L2 phrase, we might switch from L1 to L2 phrase but gestures, which express meaning synthetically, bear a stronger connection to meaning, behaviour and culture and are more difficult to substitute.

As the meaning does not change when we switch from L1 to L2, so the habit of a certain gesture remains connected to that meaning. Obviously, the change in the gesture requires a more radical change in the whole system of behaviour.

Conclusions

The study demonstrates the very limited ability of medical university students' to accompany the use of L2 with the appropriate gestures, characteristic of the L2 communication in culture 2. It also demonstrates the significance of gestures in second language teaching and acquisition. Teachers of L2 should aim not only to make students acquire skills in easy switching from L1 to L 2 but also train students to accompany it by a switch in gestures and behaviour. The switch from L1 to L2, connected to the G1-G2 switch, will ensure students are ready for efficient communication with L2 patients. Thus, medical students will acquire the necessary verbal and non-verbal skills for practicing their profession with L2 patients.

Do we need a stress tier in CVCV ?

Guillaume Enguehard

Paris 7 LLF (Paris, France)

0. In this paper, I deal with the relationship between stress and the segments. My purpose is to show that a stress tier is unnecessary in a CVCV approach.

1. Following Giavazzi (2013), stress involves two kinds of segmental operations: **i.** quantitative (vowel lengthening, post- and pretonic lengthenings); and **ii.** qualitative (preservation of vowel contrast, aspiration, glottalization).

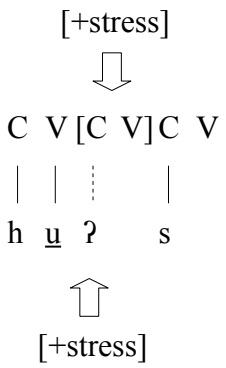
I join the CVCV framework introduced in Lowenstamm (1996). Following Larsen (1994), stress inserts a [CV] unit directly to the left (1a) or to the right (1b) of the stressed nucleus (underlined).

(1)a. Urubu-Kaapor: kaʔʔa /kaʔa/ *forest* b. Italian: faato /fato/ *fate*

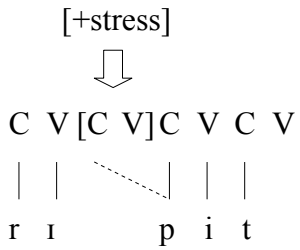
This representation accounts for the quantitative operations involved by stress, i.e. the vowel lengthening and the pre-/post-tonic lengthenings are due to the insertion of a [CV] unit.

On the other hand, it is less likely to account for the qualitative operations involved by stress. Larsen (1994) assumes that the post-stress glottalization of danish (the 'stød') is due to a floating segment inserted by stress in addition to the [CV] unit brought along (2a). Ségéral & Scheer (2001) suppose that the aspiration of voiceless plosives before stress in English is a phonetical realization of a lengthening process (2b).

(2)a. Danish: hu^{ʔ}s /hus/ *house*



b. English: ɹɪp^{h}iit /ripit/ *repeat*



2. The important fact is that the CVCV framework accounts for the segmental operations involved by stress with an inserted [CV] unit. However, this insertion *per se* is not explained. Consequently, CVCV does not account for the relationship between stress and the segments. It accounts only for the shape of this relationship.

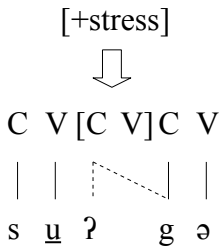
I argue that stress does not insert a [CV] unit. It results from a [CV] unit which has underlyingly nothing to do with stress. Indeed, CVCV implies other empty [CV] units: the left-edge (Lowenstamm, 1999) and the templatic unit (Lowenstamm, 2003). In what follows, I argue that both can be analyzed as the underlying representation of segmental operations related to stress in surface. I analyse the cases of Livonian and Old Norse.

3. Livonian shows a consonant alternation that applies only with consonants located after stress. This alternation is called consonant gradation (Viitso, 2007). The strong grade shows more consonantal material (3a), and the weak grade shows less consonantal material (3b).

(3)	a. Strong grade	b. Weak grade	Gloss
	su^{ʔ}**gg**ə PartSg	su**g**uu-d $^{\text{NomPl}}$	*relative*
	ta**pp**ə $^{\text{Inf}}$	ta**p**aa-b $^{\text{1pSgPres}}$	*to kill*

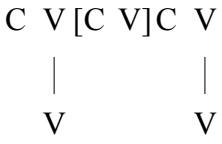
Following Kiparsky (2006), I assume that the lengthening observed in strong grade is due to stress. In CVCV, a [CV] unit inserted by stress to the right of the stressed nucleus is involved (4).

(4)



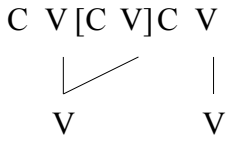
The point is that the strong grade occurs if and only if the nuclei of the word are non-branching (Viitso, 2007) (5a). If one of the nuclei is branching, the grade of the word is weak (5b-c). Note that only one nucleus can branch at a time.

(5)a. *Strong grade*



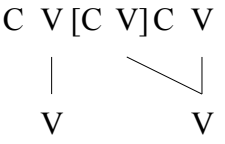
e.g. lu̲ʔggə ^{Inf} 'to read'

b. *Weak grade*



e.g. puugə ^{Inf} 'to blow'

c. *Weak grade*



e.g. ka̲laa ^{NomSg} 'fish'

Consequently, the overall length of the word remains stable. The [CV] related to stress turns out to be a templatic unit.

4. I now turn to the representation of stress in Old Norse. Stress in Old Norse is initial and fixed (Gordon, 1927). In this language, we find the voiceless fricatives f, þ and h in initial position only, and the voiced fricatives v, ð elsewhere.

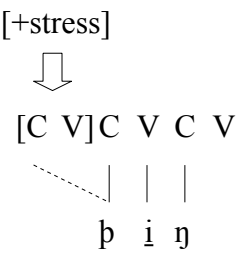
However, we observe some counter-examples. First, the Proto-Germanic root hi- gave the definite article **hinn** in some dialects (especially in Continental Old Norse), and **inn**/**enn** in other dialects. Note that the vowel reduction in the form **enn** implies that it is unstressed. There is no form ***henn**. Consequently, the preservation of the Proto-Germanic initial h is attested only before stress.

Second, the fricative marked with < þ > was voiced in Modern Continental Scandinavian in words like **þú** > **du**/**dʉ** *you*, **þat** > **dæ**(**t**) *that*, **þar** > **dær** *there*, **fyrir því** > **fu**:**ɖi** *because*, etc... All these words belong to minor categories, and they are unstressed in Norwegian (see the vowel reduction). Compare the preceding evolution of þ with its evolution in a stressed word like **þi̲ng** > **ti̲ŋ**. It is not voiced in the latter.

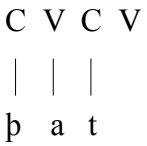
My analysis is that the preservation of Proto-Germanic voiceless fricatives was not conditioned by the initial context, but by stress. The latter inserts a [CV] unit on the left of the stressed nucleus (6a).

Following the proposition of Scheer (2000) for Proto-Germanic, I assume that fricatives resist voicing when they propagate. Compare the stressed word in (6a) with the unstressed word in (6b).

(6)a. ON. þing > tiŋ



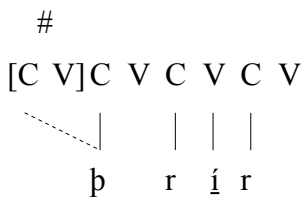
b. ON þat > dæ(t)



Hence, the conclusion. Given that the [CV] unit provided by stress in Old Norse is fixed, initial and left-inserted, it matches with the exponent of the left-edge: an empty [CV] unit inserted at the beginning of a word (Lowenstamm, 1999).

Furthermore, only the left-edge can account for a word like **þrír** *three*. Indeed, Larsen (1994) states that stress inserts a [CV] unit in the direct vicinity of the stressed nucleus. In the case of **þrír**, the voicelessness of þ is due to a preceding [CV] unit. However, this [CV] unit cannot be inserted by stress, because it precedes an empty nucleus (7). Consequently, this initial [CV] unit is a left-edge which conditions the presence of stress. Indeed, the left-edge-free words are unstressed (see þat in 6b).

(7)



5. In conclusion, I have argued that the stress tier is not necessary in a CVCV approach. Consequently, we do not need to assume that stress involves the insertion of a [CV] unit. Such an operation remains unexplained and unnecessary. Stress does not insert, but it results from a unit of the timing tier inserted by morphology: a templatic unit or a left-edge.

References.

Giavazzi, M., (2013). 'Towards a Prominence-based theory of prosodic conditioning'. *Atelier de Phonologie*, october, 16: Paris.

Gordon, E. V., (1927) [2nd edition révised by Taylor, A. R., (1956)]. *An Introduction to Old Norse*. Oxford: Oxford University Press.

Kiparsky, P., (2006). 'Livonian Stød'. Standford Universsity: [ms].

Larsen, U. B., (1994). *Some aspects of Vowel Length and Stød in modern Danish*. Mémoire de DEA [ms]: University Paris VII.

Lowenstamm, J., (1996). 'CV as the Only Syllable Type'. in Jacques Durand & Bernard Laks (eds.). *Current Trends in Phonology Models and Methods*. European Studies Research Institute, University of Salford. 419-442.

Lowenstamm, J., (1999). 'The beginning of the word'. in Rennison J. & Kühnhammer K. *Syllables ?!*. Holland Academic Graphics.

Lowenstamm, J., (2003). 'À propos des gabarits'. *Recherches Linguistiques de Vincennes, 32*. Ségéral, P. & Scheer, T., (2001). 'Abstractness in Phonology: the Case of virtual Géminates'. in: Dziubalska-Kołaczyk, Katarzyna (ed.), *Constraints and Preferences*. Berlin & New York: Mouton de Gruyter. pp. 311-337.

Viitso, T.-R., (2007). 'Livonian Gradation: Types and Genesis'. *Linguistica Uralica, 43-1*. Tartu & Tallinn: Estonian Academy Publishers: 45-62.

Gestural and prosodic functions of mouth patterns in Italian Sign Language (LIS)

Sabina Fontana
Università degli Studi di Catania

Mouth patterns co-occur with sign language utterances together with other non manual components (Boyes Braem & Sutton-Spence, 2001). Two types of mouth patterns have been described in different sign languages: Mouthings that appear to be linked to spoken languages and mouth gestures that involve different configurations of the jaw, lips, cheeks and tongue and sometimes requires some air movement or «noise». It has been shown that mouthing occur more frequently in many sign languages than mouth gestures and both of them co-occur in synchrony with the manual signs. However, whereas mouth gestures are typically considered a part of sign language, mouthings are viewed as a sociolinguistic phenomenon (Hohenberger & Happ, 2001). Although mouthings can be viewed as remnants of an oralist education, not only are they very persistent in signing, but they also seem to be considered part of the language by native signers. For this reason, it was also hypothesized that mouth patterns could convey additional meaning related to signing in the way co-speech gesture do (Pizzuto, 2003; Fontana, 2008).

Methods & Analyses: A videoclip of Laurel and Hardy was shown to elicit "mouthings" and "mouth gestures" in the signing of 20 participants: 10 adult native signers & 10 adult late-learner signers of Italian Sign Language (LIS). Deaf participants signed the video to a native Deaf signer. They were then shown the same video signed without mouthpatterns with the aim of testing its acceptability. Data were annonated in two tiers: one with Italian glosses for the manual sign and the second for mouthpatterns.

Results: There was a significant difference between the LIS native signers and late-learner signers groups for both kinds of mouth patterns: the occurrence of mouth gestures is higher in the native signer data; occurrences of mouthings seem to depend on individual style whereas their quality appears to be influenced by the age of exposure to LIS. Late signers' mouthings are more articulated and complete produced more like a fully formed spoken language word in comparison to the forms produced by the native signers group. When late learners' 'full word' mouthings are produced in sychronization with the manual sign, the overall speed of signing becomes slower. Mouthings discreteness then affects synchrony with signing which becomes slower. Co-produced mouthing phenomena appear to be

frequent (Boyes Braem, 2001; Fontana & Fabbretti, 2000) and are highlighted by the system of annotation used.

Results show that mouth patterns are important features of LIS and are considered necessary by the two groups of signers. Complex functions not only are conveyed by the two kind of mouth actions but are also linked to neuro-motor constraints. Mouthings 'stretched' over several manual components seem to have a prosodic function as well as a semantic function. Although mouthings appear to be often redundant, they function to ensure cohesion and to support the understanding of the signed message by highlighting the key information with a much greater economy and much more rapidly than the signs alone can manage. Mouthings having this function seem to be similar to co-speech gestures. Gestures seem to be shaped and act in a complementary manner in relation to the dominant linguistic modality: if language exploits the vocal channel, gesture is prevalently manual, whereas if it is conveyed in a visual gestural mode, the gesture could be oral.

References.

Boyes Braem, Penny e Sutton-Spence, Rachel (ed.), *The Hands are the Head of the Mouth- The Mouth as Articulator in Sign Language*, Hamburg,Signum Verlag, 2001.

Boyes Braem, Penny, «Function of the mouthings in the signing of Deaf early and late learners of Swiss German Sign Language (DSGS)», in Boyes Braem, Penny e Sutton-Spence, Rachel (ed.), *The Hands are the Head of the Mouth- The Mouth as Articulator in Sign Language*, 99-132, Hamburg, Signum Verlag, 2001.

Fontana, Sabina, «Mouth Actions as Gestures in Sign Language», in Kendon, Adam e Russo Cardona, Tommaso (a cura di), *Dimensions of Gestures: Special Issue of Gesture 8:1*, 104-123, John Benjamins Publishing Company, Amsterdam, 2008.

Fontana, Sabina e Fabbretti, Daniela, «Classificazione e Analisi delle forme labiali della LIS in storie elicitate», in Bagnara, Caterina, Chiappini, Giampaolo, Conte, Maria Pia e Ott, Michela (ed.), *Viaggio nella città invisibile*, 103-111, Pisa, Edizioni Del Cerro, 2000.

Hohenberger, Annette e Happ, Daniela, «The linguistic primacy of signsand mouth gestures over mouthings: evidence from language production in German Sign Language (DGS)», in BoyesBraem, Penny e Sutton-Spence, Rachel (ed.), *The Hands are the Head of the mouth - The mouth as articulator in Sign Languages*, 153-190, Hamburg, SignumVerlag, 2001.

Pizzuto, Elena, «Review of *The Hands are the Head of the Mouth- The Mouth as Articulator in Sign Language*, Boyes Braem, Penny e Sutton-Spence, Rachel (ed.),Hamburg, Signum Press», in *Sign Language and Linguistics,* 6:2, 300-305, 2003.

The interaction of segmental and suprasegmental information in word reading and picture naming. Evidence from a psycholinguistics investigation.

Beatrice Giustolisi[1], Simone Sulpizio[1, 2] & Remo Job[1]

[1] Department of Psychology and Cognitive Science, University of Trento

[2] Fondazione Marica De Vincenzi onlus

The processes of picture naming and reading words aloud reach their goal when an acoustic signal interpretable as the target word is produced. In order to generate this signal, word's segments and stress pattern have to be assembled.

Psycholinguistic models of speech production assume that phonemes and stress are retrieved separately and assembled during the prosodification process (Levelt, Roelofs, & Meyer, 1999). A separate retrieval of segments and stress has also been reported in word reading (Sulpizio, Job, & Burani, 2012). Therefore, a question we may ask is if the process of segment-to-frame association (i.e., the prosodification process) in word reading parallels that in speech production. Insight on this issue would give useful information for improving models of reading, which have usually been left underspecified with regard to the prosodification process (Coltheart et al., 2001).

In three experiments we investigated: a) how the reading system assembles segments and stress (experiments 1 and 2); b) whether reading aloud and picture naming show similarities in the computation of segmental and suprasegmental information (experiment 2 and 3). We adopted a masked prime methodology with written-word primes. Prime-target pairs were composed of three-syllable Italian words varying in stress position (penultimate- and antepenultimate-stress). In experiment 1 (reading aloud), each target (/'**fekola**/ 'starch') was associated with two word-prime conditions, which shared (congruent condition) or did not share (incongruent condition) the stress pattern with the target and always shared the initial syllable (/'fegato/ 'liver' vs. /fe'nitʃe/ 'phoenix'), and a control condition (%%%%%%). In experiments 2 and 3 (reading aloud and picture naming, respectively), each target (e.g. /**mo'neta**/, 'coin') was associated with four different word primes, which shared or did not share both the first syllable and the stress pattern with the target (*e.g.* /mo'tore/, 'engine'; /'mobile/, 'piece of furniture'; /fe'rita/, 'wound'; /'paniko/, 'panic').

Results of experiment 1 show that, for both penultimate- and antepenultimate-stress targets, words sharing stress position with the primes were read faster than those in the two other conditions. Antepenultimate-stress words, however, were read more slowly in the incongruent-prime condition

(/'**fekola**/ 'starch' - /fe'nitʃe/ 'phoenix') than in the congruent prime condition (/'**fekola**/ 'starch' - /fe'gato/ 'liver'). No difference was found between targets in the incongruent and in the control condition. For penultimate-stress targets, the incongruent-prime condition showed as much facilitation as the congruent-prime condition (/**fe'nitʃe**/ 'phoenix' - /fe'rita/ 'wound' = /**fe'nitʃe**/ 'phoenix' - /'fegato/ 'liver').

The asymmetry between antepenultimate- and penultimate-stress targets obtained for the incongruent condition may be explained assuming that: a) the segment-to-frame association and the phonological-to-phonetic mapping take place rightward incrementally (*speech planning account*, Kinoshita, 2000); b) the reading system starts the planning of articulation as soon as the relevant information for the to-be-planned unit is active. Thus, for antepenultimate-stress words the interference would be stronger as it would impact on the to-be-articulated syllable while for penultimate-stress words there would be time to mitigate its impact since articulation cannot start until the information about the stressed syllable (that is the second one) might be exploited.

Considering experiment 2, only segmental congruency affected target reading: Independently of their stress, Targets preceded by segmental-congruent primes were read faster than targets preceded by segmental-incongruent primes (**/mo'neta/** 'coin' - /mo'tore/, 'engine'; /'mobile/, 'piece of furniture' << **/mo'neta/** 'coin' - /fe'rita/, 'wound'; /'paniko/, 'panic'). A similar result for segmental congruency was also found in experiment 3 (picture naming), together with a main effect of type of target stress, with penultimate-stress targets being named faster than antepenultimate-stress targets.

The absence of any stress priming effects in experiments 2 and 3 is in contrast with results of experiment 1. Such difference might be related to different strategies adopted by the participants: In experiment 1 lexical stress is the only critical information subjects may exploit and higher attention could be addressed to it, whereas in experiments 2 and 3 both segmental and suprasegmental information were manipulated and attention could have been focused on the segmental information only, which may be assumed to be activated earlier than the suprasegmental information by the orthographic prime.

Finally, the faster naming latencies for penultimate- than antepenultimate-stress pictures might be determined by distributional differences of stress pattern: penultimate stress is the most represented pattern in the Italian lexicon (according to Thornton, Iacobini, & Burani (1997) 80% of three-syllable words bear penultimate stress) and may thus applied as a default (see Levelt et al., 1999). The effect might not be visible in reading because non-lexical mechanisms might cooperate in stress assignment.

Overall, we have shown that a) readers may compute phonemes apart from stress as predicted by models of speech production; b) the process of reading and speaking are not perfectly overlapping, since segmental information may be (initially) more constraining than stress information in reading but not in speaking.

References.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A Dual Route Cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204-256.

Kinoshita, S. (2000). The left-to-right-nature of the masked onset priming effect in naming. *Psychological Bulletin and Review, 7,* 133–141.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-75.

Sulpizio, S., Job, R., & Burani, C. (2012). Priming lexical stress in reading Italian aloud. *Language & Cognitive Processes, 27,* 808-820**.**

Thornton, A. M., Iacobini, C., & Burani, C. (1997). BDVDB: *Una base di dati sul vocabolario di base della lingua italiana*. [A database for the Italian basic dictionary]. Rome, Italy: Bulzoni.

Multimodal cues to the production and perception of irony

Santiago González-Fuente[1]; MªVictoria Escandell-Vidal[2]; Pilar Prieto[3,1]

[1]Universitat Pompeu Fabra (Barcelona, Spain), [2] Universitad Nacional de Educación a Distancia (Madrid, Spain), [3]Istitució Catalana de Recerca i Estudis Avançats.

santiago.gonzalez@upf.edu; vicky@flog.uned.es; pilar.prieto@upf.edu

In this paper we present two experiments designed to explore the production and perception of ironic speech from a multimodal point of view. Researchers focused on verbal irony expression have shown that speakers of different languages use prosodic modulations (e.g., Bryant, 2010; Cheang & Pell, 2008; Scharrer et al., 2011) and gestural markers (e.g., Attardo et al. 2003, 2011; Caucci et al. 2012) to convey information not explicitly encoded on the linguistic surface, thus facilitating the ironic interpretation. However, so far there are no studies that explore auditory and visual cues to spontaneous ironic speech in a systematic and quantitative way.

In Experiment 1, eleven pairs of Catalan friends participated in an ecological production task aimed at eliciting spontaneous ironic utterances. From 3.5 hours of videotaped dialogues, a total of 33 ironic utterances and the the 33 non-ironic utterances immediately preceding them (i.e. baseline utterances) were identified and prosodic and gestural cues to irony previously reported in literature were manually labelled and analyzed in both conditions (ironic target utterance vs. baseline utterance). Results show that ironic utterances display a highest concentration of auditory and visual markers (a mean of 8.63 cues per utterance) than baseline utterances (a mean of 3.82), and confirmed some previous results in the literature for ironic prosodic markers, namely, that the only consistent prosodic marker across all ironic utterances was the slowing down speech. Regarding the visual cues, the most consistent markers of verbal irony in this corpus were the following: (a) 84% of the ironic utterances contained smiles and/or laughs (vs. 51 % of baseline utterances), (b) 49% of ironic utterances contained gaze changes (vs. 15% of baseline utterances), and (c) 70% percent of the ironic utterances were followed by what we call gestural codas (i.e., gestural markers produced after the pronunciation of the ironic sentence, and which can consist of smiles, laughs, mouth stretching, rolling eyes,...).

A perception experiment (Experiment 2) was run to specifically test the contribution of the presence vs. absence of utterance's gestural codas to the perception of verbal irony. First, in order to obtain the audiovisual materials to be used in Experiment 2, three Catalan native speakers participated in a DCT production task (Blum-Kulka, 1989) in which 4 ironic and 4 non-ironic contexts were presented to

them. Their 24 responses (4 target utterances x 2 performance conditions –'sincere' vs. 'ironic'- x 3 subjects) were videotaped and digitally edited to obtain two sets of materials: (a) the 24 videos containing the pronunciation of the target sentences in a sincere vs ironic way followed by their respective gestural codas ('With coda' condition) and (b) the same 24 videos containing only the pronunciation of the target sentences ('Without coda' condition). The resulting 48 videos were used as stimuli in Experiment 2. Then, twenty-four native Catalan speakers participated in a perception experiment in which they were randomly presented with a set of ambiguous discourse contexts followed by a set of the ironic and sincere utterances obtained in the DCT production task in two conditions, namely 'With coda', or 'Without coda'. They were asked to judge how much irony they perceived on a Likert scale from 1 'Non-ironic' to 5 'Ironic'. The results of Experiment 2 showed that in absence of contextual cues, the presence of explicit codas (be it codas that are fulfilled with facial expressions and gaze patterns) helped listeners to achieve an ironic interpretation of an ironic utterance. Eighty-eight percent of the listeners rated as ironic ('4' and '5' scores) an ironic utterance when they saw the gestural coda vs. 58% when they saw the same ironic utterance without the gestural coda. Interestingly, listeners rated sincere utterances as more sincere (i.e. as 'less ironic' –'1' and '2' scores–) when utterance coda was not present (83%) than when it was (62 %). Thus, the results of Experiment 2 show that utterance's codas contain essential visual information that contributes to the successful understanding of an ironic utterance. All in all, the two experiments reveal the importance that both prosodic and especially visual and gestural features have in the encoding and interpretation of ironic speech.

References.

Attardo, S. - Eisterhold, J. - Hay; J. - Poggi, I. (2003). Multimodal markers of irony and sarcasm. *International Journal of Humor Research* 16: 243-260.

Attardo, S., Pickering L., Baker A. (2011). Prosodic and multimodal markers of humor in conversation. *Pragmatics & Cognition*, 19(2), 224-247.

Blum-Kulka, S., House, J., & G. Kasper. (1989). Investigating cross-cultural pragmatics: An introductory overview. In Blum-Kulka, S., House, J., & G. Kasper (Eds.), *Cross-cultural pragmatics: Requests and apologies* (pp. 1-34). Norwood, NJ: Ablex, p. 13-14.

Bryant, G. A. (2010). Prosodic contrasts in ironic speech. *Discourse Processes* 47(7): 545-566.

Caucci, G. M., Kreuz, R. J., (2012). Social and paralinguistic cues to sarcasm. Humor: International Journal of Humor Research, 25, 1-22.

Cheang, H.S. & Pell, M.D. (2008). The sound of sarcasm. *Speech Communication* 50: 366-381.

Scharrer, L. - Christmann, U., - Knoll, M. (2011). Voice Modulations in German Ironic Speech. *Language and Speech* 54(4): 435-465.

Sign recognition and non-verbal communication with robots

Michael Ingleby, Fatmah Baothman, Seyed Buhari and Alaa Almarzuki

Project Funded by KASCT: Principle Investigator Fatmah Baothman

inglebymeltham@btinternet.com; fbaothman@kau.edu.sa; a.almarzoki@yahoo.com

King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia

The areas of linguistics that concern us in this review of human-robotic dialogue include traditional phonology and its interface to morphology. We are studying human-robotic communication outside the naive ambit of speech recognizing and speech synthesizing robots, and we have examples of robots able to gesture in anthropomorphic ways (aping arm- ,hand- or face-gestures of humans). We believe that the demands of the human–robotic interface (HRI) bring a useful focus to the current upsurge of interest in the 'phonology' of sign languages, which began in the USA(ASL), and has continued in France (LSF), Italy (LIS), and Britain (BSL). All these were designed outside robotics to enable dialogue with humans suffering a speech or hearing impairment.

Phonology is a matter of constraints (as Optimality Theory practitioners emphasize e.g.[Archangeli & Langendoen, 1997]) and of cognitive models underlying articulation and perception (as element phonology practitioners argue, e.g. [Kaye *et al* 1987]). Most agree, however, in centring on one interlocutor at a time, i.e. one responder to phonotactic constraint and mental model. The first thing that emerges from HRI, possibly an object lesson to sign language analysts, is that distributional constraints and models need to be shared between interlocutors, however much their linguistic capabilities may differ. A simple example of 'phonological' sharing concerns numerical information passed between divers. Diver A indicates his pressure gauge and points to diver B (inviting B to report his own air pressure). This constrains B to report with raised fingers, one finger for each *10 bars* of pressure on his gauge. If, however, A and B are ascending and need to decompress, A's tapping his wrist computer constrains B to report how many minutes his own computer indicates to complete decompression, one raised finger for each remaining *minute* of deco needed.

We argue that shared constraints are central to gesture communication and that the sharing process is a cognitive activity. When, for example, a robot is gesturing an *inform* statement ("there is a tomato behind you") it should know that its human interlocutor is ignorant of this; i.e. it needs an estimated model of second-order beliefs. A first theory of mind for robots was set out by Scasellatis [1999, 2002], and has since been reified in working robots capable of perspective-shifting and second-

order belief estimation, for example in the MIT lab robot Ripley [Mavridis 2007] and in the HRP-2 humanoid which also has geometric reasoning in its cognitive arsenal [Marin-Urias *et al., 2009*]. The simplest gestures, analogous to demonstrative pronouns and particles in natural spoken languages, are *deictic*: pointing towards an object during a dialogue. These have long been used in human-robot interaction; starting from virtual avatar Gandalf embodying psycho-social skills [Thorisson 1996], extending to robots such as the Autonomous City Explorer [Lidoris *et al*., 2009], a robot able to navigate through Munich by asking pedestrians for directions. All such robots have embodied intelligence (e.g., Pfeifer & Bongard, 2006]; and if they can lay down markers in a shared environment to enable cooperation with agents embodying different mental models, then they are dealing with the phonology and morphology of gesture-based communication, often called stigmergy [McFarland 2009].

We will present examples of other types of gestures and relate them to mental models and the constraints of dialogue, in the light of [Mitra *et al* 2007]. In sign languages there are *alphabetic* and *numerical* gestures. The former allow words from spoken language to be imported into a signing dialogue, while the latter enable stigmergic annotation using time and distance markers as in the diver dialogues above. In HRI we have experimented with speech recognition and synthesis tools to import words and phrases instead of alphabetics, and are considering range-finders and clocks for support of stigmergic signing. We are also exploring *affective* signing in HRI. This involves the expression of voluntary or feigned emotion by a human interlocutor and the detection of the interlocutor's voluntary and involuntary emotional by the robot interlocutor: e.g., facial gestures meaning 'this disgusts me' or 'I'm *happy* you stopped doing that'. It appears that infra-red thermal imaging of faces can lead robot agents into affective computing (e.g., [Khan *et al* 2009]). Furthermore we are directing our HRI research towards teaching and learning in humans, inspired by e.g., [Lund 2004] and [Rouanet 2013].

References.

K. R. Thorisson, "Communicative humanoids: a computational model of psychosocial dialogue skills," Ph.D. dissertation, Massachusetts Institute of Technology, 1996.

G. Lidoris et al"The autonomous city explorer (ace) projectmobile robot navigation in highly populated urban environments," in Robotics and Automation, 2009. ICRA'09. IEEE International Conference on. IEEE, 2009, pp. 1416–1422.

R. Pfeiffer and J. Bongard, *How the Body Shapes the Way We Think*, MIT Press 2006

D McFarland, *Guilty Robots, Happy Dogs: the question of alien minds,*Oxford University Press 2008

B. Scassellati, "Theory of mind for a humanoid robot," Autonomous Robots, vol. 12, no. 1, pp. 13–24, 2002.

L. F. Marin-Urias et al "Towards shared attention through geometric reasoning for human robot interaction," in Humanoid Robots, 2009. Humanoids 2009. 9th IEEE-RAS

N. Mavridis, "Grounded situation models for situated conversational assistants," Ph.D. dissertation, Massachusetts Institute of Technology, 2007.

M.M. Khan, R.D. Ward and M. Ingleby, "Classifying pretended and evoked facial expression of positive and negative affective states using infrared measurement of facial skin temperature," ACM Transactions on Applied Perception, vol. 6, no. 1, pp. 6:1-6:22, 2009

Rouanet, P., et al., *The Impact of Human &Robot Interfaces on the Learning of Visual Objects*. Robotics, IEEE Transactions on, 2013. **29**(2): p. 525-541.

Mitra, S. and T. Acharya, *Gesture recognition: A survey*. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 2007. **37**(3): p. 311-324.

H.H Lund et al., *Educational robotics: manipulative technologies for cognitive rehabilitation*. in *Ninth international symposium On Artificial life and robotics (AROB 9th'04), Oita, JAPAN*. 2004.

Hand Gestures in Childhood Apraxia of Speech: a natural way towards 'Alternative' Augmentative Communication

Irene Lorenzini, Florida Nicolai
Università degli Studi di Pisa

BACKGROUND:

Childhood Apraxia of Speech (CAS) is a congenital neurogenic disorder which undermines language acquisition from the very beginning. CAS children differ from typically developing children in that, in the absence of neuromuscular, cognitive and sensory deficits, they are unable to acquire the automatic motor schemes underlying speech production. Thus, they face with extreme difficulty the otherwise "instinctive" developmental task of imitating, automatising and reproducing the phonemes and phoneme sequences of the target language. As a result, their babbling is quantitatively and qualitatively poor and results in an incomplete phonological inventory mastered variably that, in turn, determines an extremely low level of lexical development and general intelligibility.

In a nutshell, speaking seems to be the hardest trial for these subjects so much so that, in the most severe cases, they turn out to be non-verbal. In this sense, CAS children resemble the congenitally deaf: in the former case what is missing is the output, while in the other it is the input, but the outcomes seem to be somehow comparable. What is more, the few available studies on the topic acknowledge that children suffering from severe forms of CAS tend to give up on their efforts to communicate verbally and spontaneously make do with gestural communication, sometimes even developing out-and-out homesign systems.

Moreover, these subjects are not fully competent in language comprehension and simply 'blocked' in speech production since, when tested on phoneme discrimination and complex grammatical structure comprehension tasks, they show above-chance performances. Such results suggest that the disorder has an impact not only on *speech*, but on *language* too and, therefore, that this population runs the risk of remaining excluded from a complex symbolic communicative dimension.

AIMS:

The crucial importance of early access to symbolic communication is a self-evident concept nowadays and the very existence of Augmentative and Alternative Communication Systems often implemented in the speech therapy of severe CAS is connected to this principle. Nevertheless, it is still possible to encourage the concomitant adoption of more ecological methods such as the exposure to Sign

Language (SL). The case of CAS seems to make a good point for this as SL is not only able to open up a truly linguistic communicative and cognitive dimension but also seems to be a *natural* alternative for this type of patient, given their spontaneous tendency towards adopting forms of gestural communication.

METHODS AND RESULTS:

This theoretical standpoint is supported by the data obtained from the analysis of the communication skills of a group of eight children between the ages of 3,8 and 7 years. The group, which showed clear but variable deficits in every aspect of language competence (phonology, lexicon, morphosyntax), demonstrating a compelling need for intervention in order to stimulate language acquisition before the maturational constraints set in, resorted widely to gesture. This was particularly evident in three non-verbal subjects with relatively higher verbal comprehension who spontaneously answered verbal and visual stimuli by resorting to a wide range of gesture types often combined in complex sequences and supported by non linguistic vocalisations. Thanks to this strategy, they were able to engage in an interaction with the interlocutor, through relevant, albeit simple, contributions.

The gesture types observed were characterised by the prevalence of symbolic gestures representing communicative routines (e.g. yes, no, hello, etc.); deictic gestures, particularly elicited by the extra-linguistic context (a story-telling activity conducted with the aid of a visual support); mime and iconic gestures used to convey more articulated complex concepts (e.g. Subject 3 made a circle connecting the forefinger with the thumb and repeated the gesture while moving his hand from the top to the bottom to represent a falling raindrop).

CONCLUSIONS:

The above phenomenon seems to be interpretable in terms of the spontaneous development of an embryonic form of gestural communication which, if appropriately stimulated, could easily become truly symbolic and linguistic, thereby releasing language acquisition from the burden of the disorder and thus creating the preconditions for verbal rehabilitation. From a theoretical point of view, the data support a view of gestures and words as equipotential communicative options - as two alternatives with the same degree of naturalness. If a 'language instinct' exists, then it exists as much for gesture as for words.

References.

American Speech-Language-Hearing Association, *2007 Technical Report on Childhood Apraxia of Speech* (http://www.asha.org/public/speech/disorders/childhoodapraxia.htm)

Ball, L. J., Bernthal, J. E., & Beukelman, D. R. (2002). Profiling communication characteristics of children with developmental apraxia of speech. *Journal of Medical Speech-Language Pathology*, vol. 10(4), pp. 221-229.

Bridgeman, E. & Snowling, M. (1988). The perception of phoneme sequence: A comparison of dyspraxic and normal children. *British Journal of Disorders of Communication*, vol. 23(3), pp. 245-252.

Davis B.L. & Velleman S.L. (2000). Differential diagnosis and treatment of Developmental Apraxia of Speech in infants and toddlers. *Infant Toddler Intervention*, vol. 10(3), pp. 177-192.

Dodd B. (2005). *Differential Diagnosis and Treatment of Children with Speech Disorders (2nd Edition)*. London: Whurr Publishers Ltd.

Gopnik M., & Crago, M. B. (1991). Familial aggregation of a developmental language disorder. *Cognition*, vol. 39, pp. 1-50.

Greenberg M.T. (1980). Mode use in deaf children: the effects of communication method and communication competence. *Applied Psycholinguistics*, vol. 1(1), pp. 65-79.

Groenen P., Maassen B., Crul T., Thoonen G. (1996). The Specific Relation Between Perception and Production Errors for Place of Articulation in Developmental Apraxia of Speech. *Journal of Speech and Hearing Research*, vol. 39(3), pp. 468-482.

Marquardt T.P., Jacks A., Davis B.L. (2004). Token-to-Token Variability in Developmental Apraxia of Speech: Three Longitudinal Case Studies. *Clinical Linguistics and Phonetics*, vol. 18(2), pp. 127-144.

Rinaldi P. & Caselli C. (2009). Lexical and Grammatical Abilities in Deaf Italian Preschoolers: The Role of Duration of Formal Language Experience. *Journal of Deaf Studies and Deaf Education*, vol. 14(1), pp. 63-75.

Sabbadini L. (2007). *La disprassia in età evolutiva: criteri di valutazione ed intervento*. Milano: Springer-Verlag ed.

Contingent teaching: Multimodal pedagogy in an elementary classroom

Steven McCafferty[1], Alessandro Rosborough[2]

[1]University of Nevada, [2]Brigham Young University

The purpose of this study was to investigate how a teacher and young English second language (L2) learners created and expanded opportunities for literacy learning during teacher-led read aloud activities through the use of speech, gesture, visuals, and objects. Specifically, we focused on how these different modalities combined to create language-learning affordances for students. The teacher in question did not adhere to a traditional pedagogy of simply verbally representing the text, that is, pre- teaching vocabulary, reading the text aloud, having the students engage in choral reading, and then asking comprehension questions. Instead, her practice proved to be highly *contingent*. In the classroom, creating language learning contingencies has to do with situated and responsive adaptations through departing from the preplanned lesson, allowing students to engage in meaningful interaction among themselves and with the teacher (van Lier, 1996).

In order to achieve contingent teaching, the teacher in these data, depended on mimetic forms of mediation, materializing concepts, lexis, and other foci of the lesson through iconic gestures, deictic pointing at illustrations in the texts, drawing, and the use of objects found in one of the stories. The use of such a multimodal approach, and especially gesture, allowed the teacher and students to robustly shape dialogic exchanges to include shared memories of events experienced at the school, discussion of the nature of genre, student hypotheses in relation to science, the introduction of relevant personal experiences, and other contingent interactional features aligned with instructional goals.

The classroom was located in a low socio-economic, inner-city school in the southwest region of the United States. The narratives the teacher presented were two stories: one about a giant carrot stuck in the ground and the needful contributions of all the strong

and weak animals to pull it out; and a second about a boy, his messy hair, trying to tidy it, and going to school only to find his hair going back to its natural shape. Three video cameras were used to gather data from different angles for the same event. A modified version of McNeill's (1992) classification scheme was used to code gesture data. Speech-gesture transcriptions were created using ELAN and Quicktime software to identify and record gesture types and functions with video reviewing up to 1/10 of a second. In addition, teacher interaction and use of the texts as well as the drawing board and objects were part of the recordings, and all multimodal interactions were coded and analyzed in relation to one another.

Findings for the confluence of modalities in the study support the view of gesture as a "material carrier" (McNeill, 2012 following Vygotsky, 1987). In L2 contexts, gesture has been found at times to be redundant with speech (Gullberg, 1998; McCafferty, 2004; and Negueruela & Lantolf, 2004) as a specialized feature of native- non-native speaker interaction, also thought to operate as an equivalent form of "teacher -" or "foreigner talk" aimed at inducing comprehension (Hudson 2012). However, both students and teacher also engaged in the use of spontaneous gestures or "gesticulations" (McNeill, 1992) during interactions as expected in native speaker conversational contexts.

In the *Big Carrot* story, the teacher attended to establishing the setting by deicticly referencing words and objects in the illustrations and entertaining student questions and comments resulting from these references through drawing as well.

Additionally, she rendered the text in a highly dramaturgical fashion, which entailed exaggerated forms of speech, particularly stress, intonation, tone, and pitch when taking the part of characters and when rendering rhythmic passages or phrases repeated throughout the text. Exaggerated speech was accompanied by exaggerated forms of gesture, particularly the use of gesture space and gesture articulation.

In the *Bed Head* story, in addition to the dramaturgical features cited above, the teacher focused on one student in particular, who like the main character in the story had difficulties with his hair. Also, the teacher performed gestures and bodily positionings to coordinate and mirror events in the text, mimetically re-enacting the story together with the use of objects for example, a handheld mirror. Furthermore, with this book she pointed at the written page, reinforcing various prosodic elements and language patterns.

Teaching in the data analyzed proved to be an embodied, cross- modal orchestration of contingent

interaction. Studying the nature of such multimodal engagements in the classroom provides insights into the role of the many semiotic differentiations that are qualitatively distinct in teacher-student and student-student interactions as leading to both language and content learning.

The dynamics of lexical access in sign production: Phonological activation of ignored pictures in Italian sign language

Eduardo Navarrete[1], Michele Miozzo[2,3], Luigi Lerose[4] and Francesca Peressotti[1]

[1]Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università di Padova, Italy

[2]Johns Hopkins University (Baltimore, MD, US)

[3]Columbia University (New York, NY, US)

[4] Mason Perkins Deafness Fund

The main purpose of the experiment we report here is to characterize the flow of activation between the layers of processing involved in Italian sign language (LIS) production. We explore the course of feed-forward activation through the conceptual, lexical, and phonological levels of representation. In models of spoken language production there are two theoretical proposals regarding this topic. The full-cascade proposal assumes that any activated representation propagates activation to other representations at subsequent levels in the system. By contrast, the discrete proposal restricts the flow of activation between levels. We have put to test the full-cascade proposal in LIS by analyzing whether distractor pictures which have to be ignored are capable of activating their respective LIS-phonological parameters. According to full-cascade models, if the distractor picture is conceptually activated, this activation may spread until the lexical and the LIS-phonological levels of the production system.

Twenty-four LIS speakers took part in the experiment. All participants had normal or corrected to normal vision and none had other sensory deficits except for deafness. They used LIS as their primary and preferred means of communication at school and in everyday life. Participants named pictures (depicted in green) while ignoring the presentation of superimposed distractor pictures (depicted in red). Thirty-two pictures (line-drawings) were used as target pictures and another set of 32 as distractor pictures. Each target picture (e.g., *hat*) appeared along with a distractor picture whose sign was LIS-phonologically related (e.g., *tree*), and along with a distractor picture whose sign was LIS-phonologically unrelated (e.g., *bell*). Furthermore, and in order to reduce the number of related trials, a set of 16 filler targets and 16 filler distractors were selected. Targets and distractors pictures in filler trials were LIS-phonologically unrelated. Target and distractor pictures in all the trials were semantically unrelated. We observed a phonological facilitation effect: naming latencies were faster when the sign of the distractor picture was LIS-phonologically related to the target picture than when it was unrelated. We run a control experiment in order to assess whether the factor behind the phonological facilitation effect was actually the sign-phonological overlap

between the target and distractor and no other uncontrolled variables (e.g., visual masking). A group of 24 native speakers of Italian students at the University of Padova conducted the same experiment as the deaf group in Italian. Crucially, there was not phonological overlap between target and distractor Italian names. The results of this control experiment showed no differences between the LIS-phonologically related and the unrelated conditions. The pattern of these results replicates in sign language the phonological facilitation effect reported in spoken languages as English (Morsella & Miozzo, 2002) and Spanish (Navarrete & Costa, 2005).

The presence of a LIS-phonological activation from semantically unrelated distractor pictures suggests that in the course of language production in LIS, whenever a conceptual representation is sufficiently activated, some activation spread to the lexical and phonological levels. These results support the notion that activation flows in a cascade manner through the sign production system.

‘Co-signed’ and ‘co-speech’ gestures: linguistic or non-linguistic issues?

Giulia Petitta, Alessio Di Rienzo, Olga Capirci

ISTC CNR, Roma

“How can we justify ‘relegating’ the forms that occur in spoken communication in the ‘gestural’ domain, outside of language proper while, at the same time, conceding the status of linguistic items to the ‘same forms’ when they occur within sign languages” (Pizzuto 2007: 292. Quotation marks in the original). Involved here is the question of where we are to draw a boundary between what is ‘linguistic’ and what is not.

Visible bodily actions (Kendon, 2004) as used by speakers have never been regarded as part of their language, and were never included in linguistic descriptions. However, co-verbal gestures have compositional structure and semantic significance.

Deaf people, despite their own culture and language (with a special focus on linguistic issues, related to the use of sign languages, see Padden & Humphries, 2005), are perfectly integrated in the surrounding hearing community, and share its culture, practices, and communication habits. Thus, signers like speakers, use ‘co-signed’ gestures in their communication (Emmorey, 1999).

This paper aims to illustrate how sign language can integrate gestural units that resemble co-verbal gestures used by hearing people. For this purpose, we will compare narratives produced by deaf signing people and spoken narratives produced by hearing people. Both hearing and deaf adults and children have been asked to watch an extract (2’47” minutes) of a wordless cartoon, taken from the series “Tom and Jerry”, and to retell (constrained narrative) the story it depicted. The participants’ narratives were videotaped for later analysis.

To analyse this corpus, we defined a common procedure to transcribe and annotate the verbal, signed and gestural data in a formal and semantic perspective. For sign language we use a Signwriting-based transcription, in order to represent the form of signs (Sutton, 1995; Antinoro Pizzuto et al., 2010).

From a semantic point of view, according to several researchers (Kendon, 2004; McNeill, 1992), we identified and coded four main functions: representational gestures that help to identify or represent concrete and abstract referents; framing and pragmatic gestures that express social attitudes, mental states, and emotions and that help perform speech acts and comment on one’s own speech as well as others’; discursive gestures that mark speech and discourse, including dicourse cohesion gestures; and interactive gestures that help to synchronize the speaker/signer ‘s behavior with the interlocutor’s behaviour in social interaction.

In sign languages, these functions are mainly fulfilled by highly iconic structures (Cuxac, 2000),

that have been often considered as 'non-linguistic', but are instead part of sign language structure. These elements cannot be indeed simply discounted as 'non-linguistic' or 'partially linguistic' "simply because it is difficult to 'assimilate' them to what are considered 'typical' spoken language structures" (Pizzuto, 2007; Antinoro Pizzuto et al., 2010). Furthermore, it is important to note that signers produce other 'gestural' elements to support the signed discourse, apart from what have been defined as 'mouth gestures' (Boyes Braem & Sutton-Spence, 2001).

From a formal point of view, we first considered whether the gestures were produced with one or two hands. In the first case, we transcribed which was involved; in the second case, we analyzed the symmetry between the two hands. Gestures were then coded according to the same parameters used to analyze sign languages: handshapes, place of articulation, hand orientation and movement (Pettenati et al., 2010; Capirci et al., 2010) as well as non-manual elements (e.g. eye-gaze direction, facial expressions, body movements, etc.).

Our results show that gestural, verbal and signed units are strictly related in a semiotic perspective, since the semantic consistency is clear. Moreover, the affinities between gesture and sign language can give scholars the possibility of investigating the similarities between the principles on which their representation of reality and the internal structure of their units are based.

Furthermore, according to Adam Kendon (2012), studying the visible actions of speakers and signers adopting an approach in "a semiotically comparative fashion", we will be able estimate the dichotomies between "sign" and 'gesture' as over-semplified, in order to cease to consider something 'linguistic' or not, and to develop an approach to 'language' as a form of action (Capirci, Volterra, 2008).

References.

Antinoro Pizzuto, E., Chiari, I., Rossini, P. (2010), Representing signed languages: theoretical, methodological and practical issues, in M. Pettorino, A. Giannini, I. Chiari, F.M. Dovetto (eds.), *Spoken Communication*, Newcastle, Cambridge: 205-40.

Boyes-Braem, P., Sutton-Spence, R. (eds.) (2001), *The hands are the head of the mouth*, Hamburg, Signum Press.

Capirci, O., Volterra, V. (2008). Gesture and Speech. The emergence and development of a strong and changing partnership. *Gesture, 8/1*, 22-44

Capirci, O., Cristilli, C., De Angelis, V., Graziano, M. (2010). Learning to use gesture in narratives: developmental trends in formal and semantic gesture competence. In G. Stam and M. Ishino (eds) *Integrating Gestures*. John Benjamins: 189-200

Cuxac, C. (2000), *La langue des signes française (LSF): les voies de l'iconocité*, Paris, Ophrys (*Faits de langues* 15-16).

Emmorey, K. (1999). Do signers gesture? In L. S. Messing & R. Campbell (Eds.), *Gesture, speech, and sign*. New York: Oxford University Press: 133–160

Kendon, A. (2004), *Gesture. Visible action as utterance*, Cambridge, Cambridge University Press.

Kendon, A. (2012), Communication presented during the conference "In cammino verso il linguaggio. Dalla deissi gestuale alle strutture di grande iconicità. In ricordo di Elena Antinoro Pizzuto", Rome, 10 jan. 2012.

McNeill, David (1992). *Hand and mind. What gestures reveal about thought*, Chicago, The University of Chicago Press.

Padden, C.A., Humphries, T.L. (2005), *Inside Deaf Culture*, Cambridge, MA, Harvard University Press.

Pettenati, P., Stefanini, S., Volterra, V. (2010). Motoric characteristics of representational gestures produced by young children in a naming task. *Journal of Child Language, 37 (4),* 887-911.

Pizzuto, E. (2007), Deixis, anaphora and person reference in signed languages, in E. Pizzuto, P. Pietrandrea, R. Simone (eds.), *Verbal and signed languages. Comparing structures, constructs and methodologies*, Berlin-New York, Mouton De Gruyter: 275–308.

Sutton, V. J. (1995), *Lessons in sign writing*, La Jolla, Deaf Action Committee for Sign Writing.

Positional effects in short-term memory of the deaf

Anna Petrova[1], Michele Miozzo[2], Simon-Fischer-Baum[3], Francesca Peressotti[1]

[1]University of Padua, [2]Columbia University (New York, NY, US), [3]Rice University (Houston, TX, US)

Many studies report that immediate serial recall of verbal items from short term memory (STM) in deaf signers is reduced compared to hearing participants (e.g. Bellugi, Klima, & Siple, 1975; Conrad, 1970; Henson, 1982; Geraci et al., 2008). However, the source of such a difference remains still not fully understood (e.g., Bavelier et al., 2006; Hall & Bavelier, 2011; Wilson & Emmorey, 2006; 2008). Given that deaf and hearing people do not differ in terms of working memory resources (e.g., Boutla et al., 2004), the lower STM span for signs with respect to speech may depend on processes related to serial encoding and/or recall (Hall & Bavelier, 2011). The question addressed in the present study is whether the same position scheme for order encoding in STM is used for signs and speech. We can explore this by identifying the scheme used to represent the position of items in a sequence in STM for signs and compare that to previous results from STM for speech.

Fischer-Baum (2010), analysing the pattern of perseveration errors, demonstrated that hearing participants coded the position of each item in STM with reference to the both edges of the sequence, i.e. with reference to the first and the last item.

In this study we presented sequences of consonants of Italian sign language (LIS) alphabet to a group of 20 deaf students of Magarotto Institute in Padua, a secondary school where the communication among students and teachers is based on LIS. Sign sequences varying in length from 4 to 7 consonants were randomly presented on a computer screen at a rate of 1 second per sign. At the end of each sequence, participants were instructed to recall the letter signs in the same order. The length of the sequences was often purposely overspan to generate errors. We analysed the perseveration error pattern using the same technique as Fischer-Baum (2010).

The results showed that, compared to hearing people, deaf participants demonstrated a reduced span, however, we found evidence for the same both-edges representation of position in STM for signed stimuli, suggesting that the same scheme is used to represent position in this task as is used in the STM for speech stimuli task.

Headshakes in Polish Sign Language (PJM): A Corpus–based Study

Pawel Rutkowski, Ania Kuder, Joanna Filipczak, Piotr Mostowski
Warsaw University

The grammatical status of non–manual signals (i.e. various configurations of the signer's face, head, and body) is one of the most debated issues in current sign language linguistics. In principle, their role in the linguistic system seems intriguingly similar to that of spoken language prosodic phenomena. Although it has been pointed out (cf. Sandler, 2012, among others) that 'non–manuals' do not in fact constitute a natural class in sign languages (as both manual and non–manual elements can play either syntactic or prosodic roles in visual–spatial communication), at least some uses of non–manuals do resemble suprasegmental patterns found in spoken languages. On the other hand, as visual signals (including non–manuals) are also present in spoken language, one might wonder whether their sign language counterparts really have become part of grammar (in other words, whether they are used more systematically than in oral communication). Headshakes constitute a good example of a non–manual element that is used extensively by both speakers and signers. It has been hypothesized to play the role of negation marker in many (if not all) sign languages. Sandler (2012) discusses its problematic status in the following way: "sometimes attributed to prosody or intonation, this element is at least sometimes a non–linguistic gesture, as it is for hearing speakers in the ambient culture." As pointed out by Quer (2012), among others, the grammatical properties of the headshake differ cross–linguistically. There are sign languages in which the headshake is sufficient to express negation (e.g. in American Sign Language), whilst in others it needs to be combined (co–articulated) with a manual negator of some sort (e.g. in Italian Sign Language). Zeshan (2006) distinguishes the two types by labeling them "non–manual dominant" and "manual dominant" languages, respectively.

The aim of this paper is to present the distribution of the negative headshake in Polish Sign Language (*polski język migowy*, hereinafter PJM), an understudied natural language used by the Deaf community in Poland. We want to discuss how the non–manual element in question is integrated in PJM texts. Our analysis is based on empirical data extracted from a corpus of PJM that is currently being compiled in Warsaw. The underlying idea of that project is to compile a collection of video clips showing Deaf people (native signers) using PJM in a variety of different contexts. The corpus is diversified geographically and the group of signers participating in the project is well balanced in terms of age, gender, as well as for social and educational background. Recording sessions always involve two signers and a Deaf moderator. Typically, the signers are asked to react to certain visual stimuli, e.g. by describing a scene, naming an object,

(re–)telling a story, or explaining something to their partner. The elicitation materials include pictures, videos, graphs, comic strips etc., with as little reference to written Polish as possible. The raw material obtained in the recording sessions is further tokenized, lemmatized, annotated, glossed and translated using the iLex software developed at the University of Hamburg (Hanke & Storz, 2008).

For the purposes of the present study, we have analyzed 6 hours of corpus footage. An in–depth examination of that sample has allowed us to produce a typology of negative constructions in PJM. Some details thereof will be discussed in this paper, in particular with respect to the phenomenon of non–manual negation. The video material that we have analyzed consisted of approx. 11 000 segmental signs. It also included more than 600 suprasegmental headshakes (defined as side–to–side movements of the signer's head). Interestingly, only 101 headshakes expressed negation. Most of them were co– articulated with a segmental sign of negative interpretation: either a verb inflected for negation (see example (1) below; note that PJM has a negative prefix) or a manual negator (belonging to one of four types attested in PJM).

headshake

(1) BEAR SAME RICH RUSSIA PLACE EAST POLAND NEG–HAVE 'There are a lot of bears in Eastern Russia, in Poland there are none.'

The findings are summarized in the following table:

Headshakes + manual negation				
Morphologically negated verbs	B–shape 'no' (two–handed)	B–shape 'no' (one–handed)	Z–shape 'no'	Palm–up 'no'
36	8	11	5	15

Only 26 headshakes were not supported by a negative segment. They either stood alone (as negative answers, 15 examples) or were co–articulated with a non–negative lexical sign (yielding a negative interpretation thereof). The following table shows what types of segmental signs those headshakes combined with:

Headshakes + non–negative signs					
n–manual only	Verbs	Nouns	Adjectives	Particles	Numerals
15	4	2	1	1	3

We also examined the distribution of those segmental negators that were not co– articulated with a suprasegmental headshake:

Manual negation without headshakes		
B–shape 'no' (two–handed)	B–shape 'no' (one–handed)	Z–shape 'no'
10	19	2

On the basis of the above data we conclude that, in Zeshan's (2006) terms, PJM belongs to the "manual dominant" type. The headshake has not been grammaticalized as a negation marker. Instead, it is optionally co–articulated with segmental negators (similarly to headshakes in oral communication).

References.

Hanke, T. & J. Storz (2008). iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In: O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood & E. Thoutenhoofd (eds.), *Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages*. Paris: ELDA, p. 64–67.

Quer, J. (2012). Negation. In: R. Pfau, M. Steinbach & B. Woll (eds.), *Sign Language: An International Handbook*. Berlin: Mouton de Gruyter, p. 316–339.

Sandler, W. (2012). Visual Prosody. In: R. Pfau, M. Steinbach & B. Woll (eds.), *Sign Language: An International Handbook*. Berlin: Mouton de Gruyter, p. 55–76.

Zeshan, U. (2006). Negative and Interrogative Constructions in Sign Languages: A Case Study in Sign Language Typology. In: Zeshan, Ulrike (ed.), *Interrogative and Negative Constructions in Sign Languages*. Nijmegen: Ishara Press, p. 28–68.

How French syllabic patterns still influence children productions in Primary School

Samantha Ruvoletto
University of Paris 8, UMR 7023: Structures Formelles du Langage

INTRODUCTION

French language is a "cursus" language which means that is organized in open syllables CV.CV (Pulgram, 1970; Bechade, 1992). This characteristic emerges in French oral language for the couple article-noun that sounds like a disyllabic structure stressed on the second syllable. This perception influences first infant productions. Indeed at 2 years of age French children produce 'fillers', some monosyllabic structures always put before the nouns that have a phonological and prosodic specification but not a morphological one as in [lətɛ] for *le chien* [le ʃjɛ̃], en. *the dog* (Bassano, 2010; Wauquier-Gravellines, 2004).

Moreover in French, syllables mask words boundaries in cases of the phonological processes of liaison (e.g. les [le], en. *the* + ours [uʁs], en. *bears* => [lezuʁs], en. *the bears*) and elision (le [lə], en. *the* + ours [uʁs], *en. bear* => [luʁs], en. *the bear*) and cause infant difficulties in speech segmentation revealed in their wrong productions (e.g. 'integration' of the liaison consonant at the beginning of the noun, *le(s)[n]éléphants* [le.ne.le.fɑ̃] for *les[z]éléphants* [le.ze.le.fɑ̃], en. *the elephants,* or the 'omission' of it *u(n)[Ø]avion* [ɛ̃Øa.vjɔ̃] for *un[n]avion* [ɛ.na.vjɔ̃], en. *a plane*). The problematic segmentations decrease in number after 5 years old and seem vanished at 6 years of age (Basset, 2007). Nonetheless the role of prosody remains important at the beginning of primary school. Recent studies concerning early literacy tell us the central role of syllabic patterns in the first stage of learning written code. French children who start to learn reading and writing not only focus their attention on the syllables but also they use them as basic units to access to mental lexicon (Chetail, 2012).

Considering the central role of syllable in French infant acquisition and at the beginning of primary school, we're going to find out if syllabic patterns still influence the phonological competence and word segmentation analyzing the production of liaison and elision in a group of French first graders.

METHOD

43 French children (average age 6; 3) are tested individually through a task of picture naming. We propose 18 couples of cards depicting the same animal or object but differing in number (one vs.

many). The interviewer names the first card producing an elision (e.g. *l'ours* [luʁs], en. *the bear*) then asks to the child to name the second card in order to produce the plural and so the liaison (e.g. *les[z]ours* [le.zuʁs], en. *the bears*) and vice-versa (interviewer's input: liaison *les[z]oreillers,* [le.zɔ.ʁɛj], en. *the pillows* → child's production: elision *l'oreiller* [lɔ.ʁɛj]*,* en. *the pillow*). A long training precedes the performing of the task and the tests items are alternated at random with fillers that don't require productions of liaison or elision (e.g. interviewer's input: *les gateaux* [le gɑ.to], en. *the cakes* → child's production: *le gateau* [lə gɑ.to]*,* en. *the cake*).

The items for the task are chosen from the lists in the French database of child lexicon MANULEX (Lété, Sprenger-Charolles & Colé, 2004).

RESULTS / DISCUSSION

The results of the task show that syllabic patterns still influence word segmentation in production after 6 years of age. Indeed first graders still product some sequences as *le[z]ours* [lə.zuʁs] for [luʁs], en. *the bear* or *la[n]oie* [la.nwa] for *l'oie* [lwa], (en. *the goose*) showing the tendency to fill with a segment ([z] or [n]) the first onsetless syllables and reproduce CV.CV structures.

Moreover for this task we record a high percentage of hiatus (e.g. *le[Ø]oreiller* [leØɔ.ʁɛj], en. *the pillow*) when it's asked to produce an elision (*l'oreiller* [lɔ.ʁɛj]). Analyzing these 'hiatus productions' through the software PRAAT, we discover that when the interviewer gives an input in liaison (e.g. *les[z]oreillers* [le.zɔ.ʁɛj]*,* en. *the pillows*) children produce systematically *le(s)[ʔ]oreillers* [lə.ʔɔ.ʁɛj], putting a glottal stop [ʔ] between article and noun. This glottal stop is the evidence of the influence of the input syllabic pattern. Thanks to their lexical and morphological competences children can segment the input couple article+noun and can find the article for the singular form *le* [lə] that they have to produce. However the hiatus and the consequent elision are avoid because the prosodic patter CV.CV of French is preserved, thanks to the glottal stop that fills the empty position at the onset of the first syllable of the noun.

All the segments put between articles and nouns ([n], [z] or [ʔ]) seem reveal a phonological position existing and belonged to the liaison consonant presented in the input. This position confirms the liaison representation of the autosegmental model which consists in a skeletal unit and in a floating segment, with respect of both the segmental and syllabic tiers (Fig. 1, Encrevé, 1988).

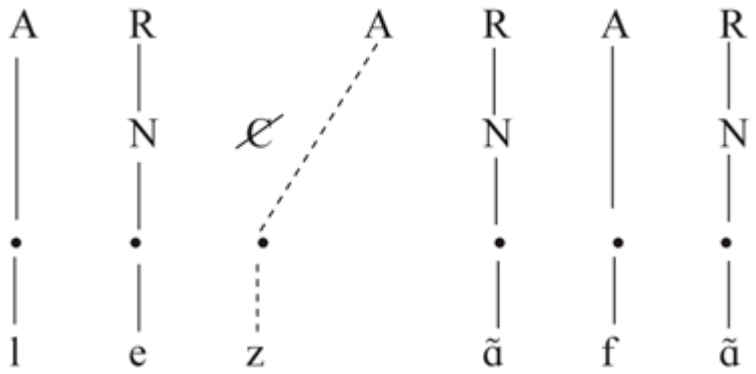


Fig. 1. Autosegmental representation of French liaison in *les enfants* [le.zɑ̃.fɑ̃], en. *the children* (Encrevé, 1988)

References.

Bassano, D. (2010). L'acquisition du déterminant nominal en français : une construction progressive et interactive de la grammaire. *Cognitextes* [En ligne], *5*, URL : http://cognitextes.revues.org/315

Basset, B. (2000). *La liaison à 3, 7 et 11 ans : description et acquisition*. Mémoire de Maîtrise Sciences du Langage, Grenoble 3

Bechade, H.-D. (1992). *Phonétique et morphologie du français moderne et contemporain*. Paris : Presses Universitaires de France

Chatail, F. (2012), La syllabe en lecture, Press Universitaire de Rennes (Rennes)

Encrevè, P. (1988). *La Liaison avec et sans enchaînement, Phonologie tridimensionnelle et usage du français*, Paris: Éditions du Seuil

Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods*, 36(1), 156-166

Pulgram, E. (1970). *Syllable, Word, Nexus, Cursus*. Paris.Mouton : The Hague

Wauquier-Gravelines, S. (2004). Les liaisons dangereuses. Psycholinguistique et phonologie : une interface complexe, in Ferrand, L. & Grainger, J. (eds.) *Psycholinguistique cognitive, Essais en l'honneur de Juan Segui*, Bruxelles : De Boeck Université, Collection Neurosciences et Cognition.

Movement, gesture and drama for langage acquisition - body and oral dynamics : *from gesture to sound*

Stéphane Soulaine
Université Catholique de l'Ouest

Keywords : gesture - body - enaction - English phonology - oral - drama

Research

My work inquires into the role that movement, gesture and drama can play in the teaching of foreign languages, and more particularly english phonology (intonation, accentuation, rhythm). The research is based on the postulate that it is the body-in-action which serves as the bedrock of the learner's relationship to the environment, the social world, and his capacity to acquire foreign rhythmical patterns.

The research which I have undertaken in the field of TEFL (teaching English as a foreign language) is based on an enactive approach which draws on two theoretical frameworks : English phonology and Laban's theory of movement. The aim of the research is to look at how a physical approach to the study of English pronunciation facilitates the learning of the rhythm of spoken English. The central hypothesis is that by tapping into the learner's founding rhythms incorporated since birth, the acquisition of English rhythm can be facilitated for the French junior high school student by:

- A physical approach to learning which uses ***gesture*** as the main support for accentuation
- A specific approach to learning ***accentuation*** focused on the phenomenon of rhythmic beats (alternating full and reduced forms), rhythmic groups, intensity and pitch.

Presentation

The paper I would like to present at the 'From sound to gesture' conference is particularly interested in the questions of rhythm, pairing *body dynamics* through movement and gestures, and *vocal dynamics* through the principles of English phonology. The presentation is thus composed of three points :

- A brief account of the methodology which was used to link gestures and oral English, as well as the process of creating the gestures which I coined *gestographs*.
- The experiments conducted in schools and also during training sessions for teacher students.
- The results of the experiments, including some videos of performances by students.

The presentation can be given in English and/or in French, as well as some Italian, if necessary. The audience may be invited to perform simple basic *gestographs* and movements to fully understand the relationship which I have established between gestures and sounds.

Facial and vocal gestures in the speech expression of emotions

Mario Augusto Souza Fontes, Sandra Madureira

São Paulo University

Introduction: This work develops an experimental study whose objective is to investigate the role of facial and vocal gestures in the expression of emotions during speech. It concerns speech expressivity and the role of visual and vocal prosody in the oral discourse.

Methods: The corpus consists of 30 sentences in Brazilian Portuguese,

10 of them with positive qualifiers such as "lindo" (beautiful), 10 with negative modifiers as "horroroso" (awkward), 10 without qualifiers. The sentences were divided into these three groups so that semantic and prosodic interactions could be discussed. The sentences were extracted from a documentary by Eduardo Coutinho film entitled "Jogo de Cena" which contains true stories, self-narratives of real-life experiences told by women who experienced them and retelling of these same stories by actresses. Valence range (from positive to negative) and 07 basic emotions (happiness, sadness, anger, fear, shame, anguish and joy) were investigated. In order to investigate the preponderance of the vocal or the facial gesture in the identification of valence (Kehrein, 2002) and these 07 basic emotions perceptual tests concerning visual and vocal cues were carried out and acoustic analysis of audio data performed. Correlations between visual and vocal prosody were investigated by means of the several methodological procedures and analytical methods. The acoustic phonetic analysis was performed by means of the ExpressionEvaluator script developed by Barbosa (2009). The ExpressionEvaluator extracts five classes of acoustic parameters and four statistical descriptors, producing 12 acoustic parameters. The five classes of acoustic parameters comprise the fundamental frequency (F0) with the extraction of the following descriptors: median, inter-quartile semi-amplitude, skewness, and 0.995 quantile; the fundamental frequency first derivative (dF0) with the descriptors mean, standard- deviation and skewness; global intensity skewness, spectral tilt (SpTt), and Long-Term Average Spectrum (LTAS) standard-deviation. The perceptual tests which involved the identification of valencefeatures and the type of emotion were answered by 30 judges. The Gtrace developed by McKeown et al (2011) was used as a tool to present the data to the judges. The judges were randomly presented with just the sound track, the image or both the image and the sound track (videos) in a three block session, with an interval of about 15 minutes between two blocks. An average session lasted for two hours. To describe the facial gestures a profile describing the visual gestures in terms of the part, its movement and directionality (upwards,

downwards, leftwards, rightwards) was used. These facial gestures were annotated by means of the ELAN developed by the Max Planck Institute of Psycholinguistics. To describe the voice quality settings and the dynamic prosodic aspects, the Vocal Profile Analysis Scheme developed by Laver et al (2007) was used. In order to correlate the acoustic and perceptual data, Explorative Multivariate Analysis procedures using R were applied. The following techniques wereused: Multiple Correspondence Analysis (MCA), Multiple Functional Analysis(MFA), Principal Component Analysis (PCA) and Factor Analysis of Mixed Data (FAMD). Cluster diagrams, dendrograms, data statistical mapping were generated and statistical significance assessed. The variables were grouped into 5 settings, being two of them qualitative and three of them quantitative. The project was approved by the Ethics Committee (Number:117.440 and CAAE number: 08319712.2.0000.5482). **Results and Discussion:** The analysis carried by the ExpressionEvaluator showed that the following variables were found to be significant ($p < .05$): the 99r quantile, the fundamental frequency first derivative (mean values) and global intensity skewness. The analysis of the vocal quality settings and the dynamic vocal aspects revealed two predominant settings: raised larynx and high pitch. The analysis of the facial expressions showed that lips, eyebrows and the forehead showed significant correlation values. The analysis correlating all the groups of variables were performed by FAMD in relation to 5 dimensions. The results were summarized in tables and graphics. **Conclusion:** The semantic positive or negative meanings of some of the sentences were modified under the influence of the vocal and or the visual prosody. Neutral sentences were judged as positive or negative based on the vocal or the visual prosody characteristics. Some samples were judged to display the same emotion, no difference between the three media: image, sound and video. However, the majority of samples showed differences in emotion evaluation depending in the media concerned. Media integrations discussed in relation to the semantic content of the sentence.

Key-words: Vocal gestures; Bodily gestures; Phonetics, Speech Expressivity, Emotions

"Mode-mixing" in the written production of deaf signers: preliminary evidences from a set of online texts

Maria Tagarelli De Monte

Istituto Statale per Sordi di Roma/Università degli Studi Roma Tre (Rome, Italy)

In this contribution I will provide an analysis of a set of texts written by Deaf people Sign Language L1 (SL_{L1}) in online chat situations. The goal of this study is to understand how code-switching and code-mixing emerges in texts produced by Deaf people who are bimodal bilingual, and the way in which this phenomena can be studied from multiple perspectives of linguistic analysis.

The written production of D/deaf people is a matter of discussion in a vast portion of the existing literature on deafness. Many D/deaf people suffer certain degrees of difficulties in reaching full literacy skills, affecting their full inclusion in the social and professional environment (Leeson, 2006). Recently, the matter has been approached from the point of view of studies in foreign language learning. The assumption is that Sign Language can be considered as an L1 in case of deaf people naturally exposed to this language modality; considering the speech rehabilitation that all deaf people attend, this specific type of language competence can be considered as bilingual bimodal, as it refers to two different language modalities - spoken and signed. (Lillo-Martin, 2010)

In the attempt to study the way in which bimodal bilingualism could influence the written production of Deaf people in online chat situations, we have conducted a preliminary study on 23 chat sessions held by users of popular websites. Fiorentino (2004) defines online written texts as an "italiano dell'uso medio" (Italian of the average use), with tendencies to orality. Personal email messages, board messages and discussion lists show characteristics such as informal registry, imitation of speech, abbreviations, emoticons and Anglicisms. Phenomena of code code-switching and code-mixing are common and typical on the Web (Orletti, 2004). Punctuation is enriched by repetition of signs such as question or exclamation marks or enriched by the use of capital words.

In the case of deaf people writing on the Internet, it's possible to observe a double layer of mode/code-mixing; the first appears between the written and spoken version of the language they are using, the second reveals the interference of structures typical of Signed Languages in the attempt to write in a correct Italian form.

Through this work I will introduce a preliminary study conducted on a set of chat sessions selected to show how the Italian written production of deaf signers seems to borrow some words and structures from Signed Languages actually acting on it as a "repairing" resource from which help can be withdrawn in case of need. In particular, we will focus on the construction of nominal

phrases and the effects that the use of name (instead of verbs) has on lexicon, syntax, phrases and punctuation. Preliminary results show that there are multiple linguistic influences on the written production of Deaf people:

- a vast influence of structures typical of spoken Italian used "as they are" also in their written text. This phenomena is typical of the written text on the Web and, thus, it involves D/deaf people as well;
- the use of structures typical of SL discourses;
- atypical structures in the construction of phrases and code-mixing with languages other than Italian or SL.

We will offer a presentation of the research introduced here as well as a discussion useful to a deeper understanding of the written production of the Deaf.

References.

Fiorentino, G. (2004), Frasi nominali nel parlato dialogico: problemi empirici e teorici. In: F. Albano Leoni, F. Cutugno, M. Pettorino, R. Savy (a cura di), *Il parlato italiano. Atti del convegno nazionale di Napoli.*, Napoli: M. D'Auria Editore, B05 [il volume è pubblicato sotto forma di CD-rom].

Leeson, L. (2006), *Signed Languages in Education in Europe - a preliminary exploration*, Strasbourg, Council of Europe.

Lillo-Martin, D., Muller De Quadros, R.,Koulidobrova, H., Pichler, D.C., (2010), Bimodal Bilingual Cross-Language Influence, in Costa, J., Castro, M. Lobo, Pratas, F. (a cura di), *Language Acquisition and Development: Proceedings of GALA 2009*, Newcastle upon Tyne: Cambridge Scholars Press, pp.264-275.

Orletti, F. (2004), *Scrittura e nuovi media. Dalle conversazioni in rete alla Web usability.*, Roma: Carocci.

Coordinating Gesture, Speech, and Diagrams in Explanations

Barbara Tversky[1,2], Seokmin Kang[3], John Black[1]

[1]Columbia Teachers College, [2]Stanford University

[3]Arizona State University

Explaining how to get somewhere, how to do something, how something works, or how something happened are common in formal and informal settings. Explanations typically harmoniously combine words, gestures, gestured worlds, and artifacts. In teaching, explanations use diagrams, real diagrams or, as in informal situations, virtual diagrams created in the air with gestures. Diagrams are especially effective for conveying structure as they map elements and relations of the system to elements and spatial relations in space. Gestures are especially effective for conveying action, as they are actions. Here we present two experiments exploring production and comprehension of gestures in explanations of complex systems. One examines the spontaneous coordination of words, gestures, and diagrams, real and virtual, in explanations of dynamic systems to experts and novices. The second examines learning the structure and dynamics of complex systems through explanations accompanied by gestures that depict structural or dynamic properties.

Producing Explanations of Dynamic Systems

Participants first studied a dynamic system, the circulatory system or the rock cycle, and then explained the system to a video camera under the assumption that the video would be viewed by either an expert or a novice. Structural properties are relatively easy to represent in diagrams but dynamic properties are harder to represent in diagrams but often more naturally represented in gesture. Thus, participants were expected to rely on diagrams for structural information and on gesture for dynamic information. Explanations to experts need to establish that the explainer has correctly mastered the information whereas explanations to novices have to instill that information in someone who knows little or nothing about the system. Thus, explanations to novices should contain more information, some of it redundant.

Methods. Participants first studied the circulatory system and the rock cycle from diagrams in order to prepare video recorded explanations either for an expert in the field or for a novice with no knowledge of the systems. They were told that experts or novices would view the videos. Each participant explained one system to an expert and one to a novice (counterbalanced). Participants then gave explanations to a video camera, free to use the diagrams provided.

Findings. Gestures and speech were coded, with high reliability established on a subset of the data. As expected, participants conveyed structural information using deictic gestures that pointed to the

relevant part on the diagram. They conveyed dynamic information using iconic gestures depicting the actions of the system. They conveyed far more dynamic information than structural information probably because the structural information was apparent in the diagrams. Although explanations to novices and to experts took the same amount of time, explanations to novices contained far more information than explanations to experts. Explanations to both novices and experts used gestures referencing the given diagram. However, explanations to novices also typically used gestures to create large virtual diagrams of the structure of the system in the air. They then explained the dynamics of the system with iconic gestures that referenced the invisible virtual diagram.

Learning from Dynamic Gestures

Because gestures are actions, they should be especially effective in conveying action, an assertion tested in the second study, which examined learning of students who watched videoed explanations of the workings of an engine. Some saw gestures that depicted parts and their structure; others saw gestures that depicted the actions of the parts. The gestures accompanied the same verbal script. The expectation was that those who watched action gestures would achieve a deeper understanding of the dynamics of the system.

Methods. Students watched explanations of the workings of an engine, accompanied by a schematic diagram of the spatial structure of the parts of the engine. For half the students, the verbal explanation was accompanied by iconic gestures that depicted the parts of the system in the appropriate spatial array. For the other half, the verbal explanation was accompanied by iconic gestures that showed the actions of the parts. The dynamic gestures were not keyed to the spatial array. Following viewing the video four times, students were given a verbal test, provided visual explanations of the workings of the engine, and made a video explaining the workings of the engine to someone not knowledgeable of the workings of the engine.

Findings. The test of structural and dynamic properties of the system could be answered solely on the basis of the verbal explanation, which was the same for both gesture conditions. Nevertheless, students who saw the dynamic gestures performed better on the questions about the actions of the engine. Compared to those who had viewed gestures depicting the parts of the engine, students who had viewed gestures depicting the action of the parts depicted more action in their visual explanations and used more action gestures in their videoed explanations.

Conclusions

In the production study, students explaining complex systems used words, gestures, diagrams, and virtual diagrams created by gestures in the air. Explainers used iconic dynamic gestures depicting the actions of the parts of the system and deictic gestures pointing to the structure of the system. They conveyed more dynamic information than static information, and more information to novices

than to experts. Explanations to experts relied primarily on the given diagram whereas explanations to novices typically used the given diagram but also created large virtual diagrams in the air, later gesturing on the to explain dynamics.

In the learning study, students watched explanations of a complex system that were accompanied by either gestures depicting the form of the parts or depicting the actions of the parts. The verbal script was the same, and sufficient for perfect performance on the test. Those who saw gestures depicting action performed better on the action questions, showed more action in their visual explanations, and used more action gestures in their videoed explanations.

Gestures are actions, and naturally used in explaining action. They enhance knowledge of action over and above sufficient descriptions in words.

Features and Modality:
Speaker/Hearer Neutrality in Signed and Spoken Language Phonologies

Shanti Ulfsbjornin
SOAS - University of London

In this presentation we argue that signed and spoken language phonologies operate with very different types of features, the former articulatory, the latter acoustic. We will argue that this difference exists because the different feature sets are speaker/hearer neutral in different modalities, a conclusion which we believe is in line with the aims of the Minimalist Program.

One aspect of generative grammar which persists from Chomsky (1965) to the modern day is the assertion that a Grammar, is neutral between speaker and hearer; as such it does not in itself "prescribe the character or functioning of a perceptual model or a model of speech production" (Chomsky 2008). At the same time, the Strong Minimalist Thesis (Chomsky 1998, 2010) claims the language faculty can be summarised as (1)

(1) Merge + Interfaces (CI (Conceptual/Intention) and SM (Sensory Motor)) = language

Furthermore, the Minimalist Program defines language as an: "optimal solution to legibility conditions" (Chomsky 1998).

We contend that these properties, neutrality between speaker and hearer and the creation of optimal solutions to legibility conditions, applied to the SM interface (phonology) has important conclusions for the nature of the features in sign language and spoken language phonologies.

The majority view in phonology is that the features of spoken language phonology are primary articulatory based and taken from the *Sound Pattern of English* (Chomsky and Halle 1968), or Articulatory Phonology (Browman and Goldstein 1989), which betrays the centrality of speech production and varying degrees of commitment to the motor theory of speech perception (Harris and Lindsey 1995). Harris and Lindsey (1995) go on to argue that it was in error to abandon the Jakobsonian insight that subsegmental primes should be on acoustic terms, that in fact, the speech signal is shared by speaker and hearer (Jakobson, Fant and Halle 1962).

The minority view in phonology, though perhaps a growing minority, is in support of acoustically driven features called elements (Anderson and Jones 1974, 1977; Anderson and Ewen 1987; Kaye et al. 1985; Charette and Kaye 1993; Ploch 1997; Scheer 1998; Kaye 2000; Botma, Kula and Nasukawa 2011; Backley 2011; Nevins 2012; van der Hulst 2014).

In order to further support the element theory of features we will present novel research showing

perception of Mandarin L2 targets by (non-l-vocalising) Southern British English speakers. The subjects are exposed to two deviant versions of l-"coda" targets (such as 'sell') one which is acoustically coherent: [sew] 'sell' and another which is not: [seɻ] 'sell'; while both deviant versions are articulatorily equally deviant, British speakers overwhelmingly decode the acoustically coherent tokens.

Having settled and hopefully demonstrated that generative grammar operates on elements rather than articulatory features, we turn to a discussion of sign language.

Sign language phonology has also been broken down into features which are articulatory in nature (Sandler 1989; van der Hulst 1995; Moren 2003; van der Hulst and van der Kooij 2006). In this case however, we will claim that articulatory features are the neutral ones between speaker and hearer as the articulators in signed languages are visual and shared by both speaker and hearer. We will present our revised set of features for sign language. Therefore, although signed language could have piggy-backed on dependency (van der Hulst 1996; 2003), headedness, and other aspects of syllabic structure (Brentari 2002), sign languages could not have piggy-backed on the features of phonology itself.

We argue therefore that signed and spoken language phonologies operate with very different features, precisely because different feature sets are neutral between speaker and hearer in different modalities, a conclusion which we believe is in line with the aims of the Minimalist Program.

What pitch range dimensions tell us.
The case of Italian and American English

Martina Urbani, Università degli Studi di Padova

Pitch range is at the same time a universal characteristic of any language and a language-specific factor that differs across speakers who have distinctive language backgrounds. As a consequence, a language may possess recurring pitch range patterns that do not apply to other languages. What is more, pitch range has a socio-cultural impact because it may be variously interpreted and perceived, based on the different expectations of a community of speakers.

Unlike other prosodic cues such as stress, rhythm and speech rate, it is not clear to what extent pitch range may contribute to the detection of accented speech produced by L2 speakers. What is evident is that the speakers of different languages seem to have distinctive characteristics of pitch range (Chen, 2009). Pitch range has been erroneously considered as a single unitary measure, while, actually, it is the result of two different dimensions: level and span. While pitch level is a sort of reference line calculated over the rises and falls within each intonation contour, pitch span is a measure of the distance between the highest and the lowest F0 value in the contour (Ladd, 1996; Gussenhoven, 2004).

In this study, differences in pitch range across American and Italian speakers are compared. Two experiments were carried out on speech material produced by some Californian American and Veneto Italian native speakers to analyze the patterns of pitch range in English as L1 and L2. The investigation of the language-specific use of F0 prompted a number of speculations and hypotheses related to the characteristics of pitch range across languages, speakers, genders, and sentence types. Coherently with the model proposed by Mennen et al. (2012), the analysis of pitch range was based on the investigation of LTD (long-term distributional) and linguistic measures. LTD measures deal with the F0 distribution within a speaker's contour (e.g. F0 minimum, F0 maximum, F0 mean, F0 median, F0 span) while linguistic measures are linked to specific targets within the contour, such as peaks and valleys (e.g. high and low landmarks) and preserve the temporal sequences of pitch contours.

Results show that American English and Italian significantly differ in the pitch range adopted by their L1 and L2 speakers. The Italian subjects of the experiment appeared to project their standard L1 pitch range onto their L2. What is more, the different phonological and phonetic conventions displayed in languages such as American English and Italian had an influence on the modulation of pitch range trends, that were perceived and interpreted depending on the socio-cultural expectations

of the linguistic communities. For example, the American speakers seemed to speak with lower F0 levels than the Italians, thus their pitch space appeared as shifted downwards. As far as pitch span is concerned, the Americans' speech was characterized by wider F0 span than the Italians' speech. Meaningful phonetic differences were also found across three sentence types (yes/no questions, wh-questions and statements) produced by Americans and Italians.

References.

Chen, A. (2009). Perception of paralinguistic intonational meaning in a second language. *Language Learning,* 59 (2), 448-451.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: CUP.

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

Mennen, I., Schaeffler, F., Docherty, G. (2012). Cross-language difference in f0 range: a comparative study of English and German. *Journal of the Acoustical Society of America*, 131 (3), 2249-2260.

The role of depictive gestures in learning how to make music together: observations from ensemble music workshops

Daniela Veronesi
Università Libera di Bolzano/Freie Universität Bozen

In recent decades, the question about the embodied nature of human communication has gained more and more centrality in the study of social interaction; within Conversation Analysis, for instance, this question has been addressed with the aim of understanding how different modalities such as talk, gaze, gesture etc. are integrated so as to form coherent courses of action and contribute to the definition of action, interaction and participation frameworks (cf. Goodwin 1981 and 2012; Goodwin & Goodwin 2004; Mondada 2007; Streeck, Goodwin & LeBaron 2011, among others).

A case in point in this respect is interaction in musical settings, which, as discussed by anthropologists, ethnomusicologists, linguists and psychologists alike (see Berliner 1994; Black 2008; Haviland 2007 and 2011; Poggi 2002 and 2011; Streeck & Henderson 2010; Weeks 1996), deeply relies on the interplay between various semiotic resources - the first among them being the body itself - drawn upon by musicians in accomplishing the joint activity of playing together, in defining their roles and identities within the group and in reaching agreement on musical actions in performance, rehearsal and educational activities.

Carried out within the framework of Conversation Analysis and research on multimodality, the study presented here aims at advancing this body of literature by examining a specific practice of ensemble music making, known as "Conduction®" and based on a codified lexicon of gestural instructions ("directives") addressed by a conductor to instrumentalists in the absence of notated music. Bearing only little similiarity with gestures employed by conductors in Western art music (cf. Boyes Bräm & Bräm 2004; Rudolf 1994) or by bandleaders of jazz ensembles, such gestural instructions are not shared across musical communities; in the past, they have thus been introduced by the creator of *Conduction*, US composer and conductor Lawrence D. "Butch" Morris, within musical workshops held internationally both with professional musicians and with music students.

Through the analysis of audio-and videorecorded examples taken from a collection of such workshops, held in Italy between 2008 and 2011, the study explores the role of gestures in the context of two recurring types of activities in these events, namely explanations and corrections. After providing a brief characterization of the practice of *Conduction* and an overview of the techniques drawn upon by the conductor so as to introduce *Conduction* gestural directives and to correct musical action, it is focussed in particular on "depictive" or "representational" gestures

(Streeck 2009; Kendon 2004) - as those, for instance, providing a visual illustration of a specific sound quality - and their interplay with co-occuring talk on the one hand, and with *Conduction* gestural directives they are put in relation to on the other.

It is thus examined how withdrawing on such depictive gestures - not part of the lexicon of Conduction but rather reminding of gestural practices in ordinary conversation - is an essential component of musical pedagogy and instructional activities; specifically, it is shown how in the context under consideration they contribute, together with talk and with their specific affordances, to the definition and clarification of the "meaning" of *Conduction* gestural directives as codified and standardized signs which are to be learnt by participantig musicians.

References .

Berliner, P. F. 1994. *Thinking in Jazz: The Infinite Art of Improvisation.* Chicago*:* The University of Chicago Press. Black, S. 2008. Creativity and Learning Jazz: The Practice of "Listening". *Mind, Culture, and Activity* 15, 279-295.

Boyes Bräm P., and T. Bräm. 2004. Expressive gestures used by classical orchestra conductors. In *The semantics and pragmatics of everyday gestures*, eds. C. Müller and R. Posner, 127-143. Berlin:Weidler.

Goodwin, C. 1981. *Conversational Organization. Interaction Between Speakers and Hearers*. New York: Academic Press.

Goodwin, C. 2012. The co-operative, transformative organization of human action and knowledge. *Journal of Pragmatics* 46(1): 8-23.

Goodwin, C. and M. Goodwin. 2004. Participation. In *A Companion to Linguistic Anthropology*, ed. A. Duranti, 222-244. Oxford: Basil Blackwell.

Haviland, J. B. 2007. Master Speakers, Master Gesturers: a string quartet master class. In *Gesture and the Dynamic Dimension of Language: Essays in honor of David McNeill*, eds. S.D. Duncan, T. Levy and J. Cassell, 147-172. Amsterdam/Philadelphia: Benjamins.

Haviland, J. B. 2011. Musical spaces. In *Embodied Interaction*. *Language and Body in the Material World,* eds. C. Goodwin, J. Streeck and C. LeBaron, 289-304. Cambridge: Cambridge University Press.

Kendon, A. 2004. *Gestures: Visible Action as Utterance*. Cambridge, Cambridge University Press.

Mondada, L. 2007. Multimodal resources for turn-taking: pointing and the emergence of possible next speakers. *Discourse Studies* 9(2), 194-225.

Poggi, I. 2002. The lexicon of the conductor's face. In *Language, Vision, and Music, eds.* P. S McKevitt, S. O' Nuallàin and C. Mulvihill, 271-284. Amsterdam, Benjamins.

Poggi, I. 2011. Music and leadership: the Choir Conductor's multimodal communication. In *Integrating Gestures. The Interdisciplinary Nature of Gestures*, eds. G. Stam and M. Ishino, 341-353. Amsterdam, Benjamins.

Rudolf, M. 1994. *The grammar of conducting: A comprehensive guide to baton technique and interpretation* (2nd edition), New York: Schirmer.

Streeck, J. 2009. *Gesturecraft. The manu-facture of meaning*. Amsterdam: Benjamin.

Streeck, J., D. Henderson. 2010. Das Handwerk des Hip-Hop: Freestyle als körperliche Praxis [The handcraft of hip- hop: Freestyle as bodily practice]. In *Gesten: Inszenierung, Aufführung und Praxis*, eds. C. Wulf and W. Fischer-Lichte, 62-92. München: Wilhelm Fink.

Streeck, J., C. Goodwin, and C. LeBaron. eds. 2011. *Embodied Interaction. Language and Body in the Material World.*

Cambridge: Cambridge University Press.

Weeks, P. A.D. 1996. Synchrony lost, synchrony regained. The achievement of musical co-ordination. *Human Studies* 19, 199-228.

Multimodal Vagueness in dreams.
A comparative study between Italian and Swedish

Laura Vincze[1], Isabella Poggi[2]
[1]Università di Macerata, [2]Università Roma 3

The notion of vagueness has been mainly investigated in philosophy (Russell 1923; Keefe 2000) with the challenge posed by the Sorite Paradox: how many sand grains make a "sorite", a heap of sand? Epistemicists see vagueness as stemming from the Speakers' knowledge limitation (Williamson 1994). In Linguistics it is seen as a typical property of language, as a lack of clear-cut boundaries between categories (take for instance Labov's (1973) continuous transition between cups and bowls), that leaves room for a permanent feasibility to creativity and innovation (Wittgenstein).

Not so many studies have been dedicated to how people acknowledge their vagueness in discourse through multimodal communication. Recently, Vincze et al. (2012) provided a cognitive definition of vagueness as a property of knowledge and consequently a property of a Speaker's communication: a lack of detail in what one knows and/or communicates about something. One may be vague either because one personally has vague knowledge (*no power to be precise*), or, although having detailed information, one does not want to reveal it to the listener (*no goal to be precise*) because possibly harmful, either for the Interlocutor (take the case of serious diagnoses), or for himself.

In Vincze et al. (2012), vagueness was contrasted to precision; the fact of having beliefs on each spe- cific aspect of a topic, but distinguished from uncertainty, since we may have a vague knowledge, a vague idea, a vague remembering of something, but still be certain of it. Vagueness was distinguished from approximation, a lack of precision concerning quantitative aspects of the topic, as opposed to vagueness that concerns qualitative aspects of it: the former has to do with measuring, the latter with de- scribing. Approximation is close to uncertainty and, like vagueness, opposed to precision, and precision is the opposite of both approximation and vagueness, but viewed from two different angles: quantity and quality.

Besides setting the conceptual differences between these phenomena, Vincze et al. (2012) looked at how we multimodally communicate meanings of vagueness and approximation during discourse, by defining "vagueness signals" as the verbal or bodily metadiscursive signals (Poggi, 2007) that convey the meaning "I am being vague". In general, metadiscursive signals reveal the Sender's goals con- cerning her/his discourse planning, i.e. what s/he considers important, what s/he affords

to skip, and what logical links s/he states among parts of her/his plan. During discourse, if we want to convey we are being less detailed or accurate in some parts of it, e.g., because those parts are not so important in the economy of the whole discourse, we may do so by words, gestures, gaze or facial expressions. These are "vagueness signals", i.e. metadiscursive signals that convey "I deliberately choose to be vague about this".

Data collection and analysis

Vincze et al. (2012) investigated vagueness signals in a videotaped corpus where Italian students were asked to tell, in front of the camera, a dream they dreamt. Capturing instances in which people report about vague concepts required eliciting situations where one does not have precise knowledge/remembrance about the specific concepts at issue: such as in dream-telling. The events in a dream are, by their nature, often confuse and vague, as one does not only mismatch entities (people, places) with one another by attaching some attributes of *x* to *y*; but sometimes, one does not have precise remembrance about some relevant attributes of these entities.

25 students in Education Sciences, 23 females and 2 males, between 20 and 30, were asked to tell a recent dream. While telling their dream to an interviewer, the participants were recorded by a digital Panasonic camcorder. Participants were seated on a chair without armrest and only their upper body (trunk, arms, hands and head) were video- recorded. When participants, during their narration, mentioned vague remembrance or vague knowledge of events happening in the dream, the interviewer would ask them to detail those aspects of the dream with the intent of "raising" (Gianturco, 2004) the performance of (possibly vague) gestures accompanying vague memo- ries.

A total of 25 fragments of dream telling were collected, consisting of approximately 5 minutes each. The verbal behaviour of participants was transcribed by taking into account the intonation unit (IU) as the basic unit of transcription. The intonation unit is a prosodic unit in natural discourse, a speech segment that falls into a single coherent intonation contour, and is sometimes separated by pauses at the beginning and the end (Chafe 1987, Du Bois et al. 1992).Transcribing the data in IUs, each IU lined up on a separate line, helps readers to more easily grasp the pauses in speech.

Before this transcription, the video-recorded data were first viewed on mute mode to avoid bias from the verbal context. When items of gestures or facial expressions possibly conveying vagueness meanings were singled out, the video fragment was reviewed on voiced mode, transcribed and later coded by two independent coders. All the body signals conveying vagueness, approximation and word search were transcribed and analyzed in an annotation scheme of multimodal communication (of the type of Poggi, 2007). For each signal we annotated: 1. concomitant verbal behaviour, 2. analysis of the signal (for a gesture, its handshape, place, orientation, and the parameters of movement, such as direction, path, tension, amplitude, fluidity, repetition); 3. possible concomitant

body behaviour, like gaze, smile, posture; 4. the meaning attributed to the signal at hand. Based on such annotation, each signal was coded as one of vagueness, approximation, word search, hesitation or hastiness, and a hypothesis was made as to the reason (no-goal or no-knowledge) for the participant to be vague, approximate or other in that context.

A correspective Swedish corpus (although smaller) was collected: 7 native Swedish speakers, participants in a summer school in Mullsjö, Jönköping County (Sweden), were videorecorded while asked to tell a vague dream in their own language. They were videotaped seated on a chair without armrest and and only their upper body (trunk, arms, hands and head) is seen in the video. The addressee of the dream telling was another student, who, though not a native speaker of Swedish, had a proficient knowledge of the language. We wanted to avoid this way the lack of feedback (or even a fake feedback) a non Swedish speaker (like the authors themselves) would give while listening to a monologue that they do not understand. Same as in the previous recording, when participants, during their narration, mentioned vague remembrance or vague knowledge of events happening in the dream, the interviewer would ask them to detail those aspects of the dream with the intent of "raising" (Gianturco, 2004) the performance of (possibly vague) gestures accompanying vague memories.

Same as the previous Italian corpus, the Swedish corpus was first viewed on the mute mode. When items of gestures or facial expressions possibly conveying vagueness meanings were singled out, the video fragment was reviewed on voiced mode, transcribed and translated into English by Swedish speakers and later coded by two independent coders. All the body signals conveying vagueness, approximation and word search, were transcribed and analyzed in an annotation scheme of multimodal communication (of the type of Poggi, 2007). Same as for the Italian corpus, for each signal we annotated: 1. concomitant verbal behaviour, 2. analysis of the signal (for a gesture, its handshape, place, orientation, and the parameters of movement, such as direction, path, tension, amplitude, fluidity, repetition); 3. possible concomitant body behaviour, like gaze, smile, posture; 4. the meaning attributed to the signal at hand. Based on such annotation, each signal was coded as one of vagueness, approximation, word search, and a hypothesis was made as to the reason (no-goal or no-knowledge) for the participant to be vague, approximate or other in that context.

Aims

The present study wants to investigate vagueness signals in two corpora of videotaped dream telling, an Italian and a Swedish one. It is an intercultural study, aimed at analysing possible similarities and differences in communicating vagueness in two different cultures. This study aims to be a continuation of another intercultural study where the cognitive and semantic nature of vagueness, unspecificity and approximation (vagueness related phenomena) were investigated by

examining some of the linguistic expressions connected with these phenomena in English, Swedish and Italian (Allwood et al. 2013).

Breaking boundaries:
A study of Italian boundary crossing situations in speech and gesture

Bjørn Wessel-Tolvig
Centre for Language Technology, University of Copenhagen
bwt@hum.ku.dk

It is long thought a linguistic constraint for Romance languages to express boundary crossing situations with verb particle constructions (VPCs) (Aske, 1989; Cadierno & Ruiz, 2006; Slobin & Hoiting, 1994). Crossing a spatial boundary is "*conceived as a change of state, and (that) state changes require an independent predicate in such languages*" (Slobin, 1997, p. 441). Thus Romance languages, and thereby Italian, need two predicates to express a figure crossing a physical boundary. Preliminary results from a multimodal corpus selection challenge this view.

Italian is traditionally seen as a verb framed language that packages path of motion in verb roots and leaves manner to be expressed in subordinate clauses - if at all. Recent studies show how Italian may deviate from this standard verb framed pattern by expressing motion events in VPCs (Folli, 2008; Iacobini & Masini, 2006), and some scholars take this point even further arguing that Italian may be experiencing a typological shift towards a more satellite framed oriented pattern (Iacobini, 2010). These ideas seem to support the notion of a split system typology for lexicalization (Talmy, 2000), and that Italian may express more manner than previously thought (Slobin, 2004). The question remains how a path-oriented language as Italian conceptualizes the crossing of boundaries.

Gestures may help shed light on this issue. Studies involving co-speech gestures and motion events have shown how gesture production vary cross-linguistically as a factor of different lexicalization patterns, namely how the semantic components of manner and path are mapped onto linguistic form. Most studies reveal how speakers of satellite framed languages often express motion in tight one clause constructions e.g. manner verb + path particle, and accordingly produce one gesture, whereas speakers of verb framed languages often separate manner and path in two clauses and consequently have the tendency to produce two separate gestures for manner and path. Recent studies on Italian gesture production reveal how Italian speakers can, and do, express motion in satellite framed constructions which also influences the way the speakers gesture (Rossini, 2005; Wessel-Tolvig, 2014). This gesture-typological variation might reveal how motion is conceptualized by speakers of different languages as "*gestures reflect linguistic conceptualization and cross-linguistic difference in such conceptualizations*" (Gullberg, 2011)

This study investigates how Italian speakers express motion in boundary crossing situations. Fourteen native Italian speakers from Rome narrated four distinct motions events consisting of a tomato-like ball jumping or rolling into or out of a small yellow house on a field (Wessel-Tolvig, 2013). Results show how Italians speakers display an array of different lexical constructions for expressing these particular motion events. These expressions were grouped into three categories: Path only, path + subordinate manner and manner verb + path particle descriptions. For gestures also three labels were given: manner only, path only and manner-path conflated gestures.

Although echoing a preference for dividing manner and path in boundary crossing situations (61.11%), and for expressing path only in motion descriptions (25.93%), thus expressing the crossing of a physical boundary with path verbs +/- manner descriptions, there is a small number of speakers who express the boundary crossing situation with manner verb + path particles (12.97%). When separating manner and path in speech there is a tendency that speakers produce two separate gestures. In 58% of the motion constructions with two clauses speakers produced two gestures e.g. one for manner and one for path, and in the path only constructions (technically only one clause) the speakers produced one gesture. Interestingly in the situations where the Italian speakers expressed a manner verb + path particle, these constructions were accompanied by one gesture mainly manner-path conflated information. This indicate that Italian can express motion in boundary crossing situations with manner verbs + path particles, and gesture patterns only seem to confirm that the motion is conceptualized as a single motion for manner and path.

Further research is needed to reveal if these "deviations" from standard typological verb framed patterns in boundary crossing situations are in fact only linguistic slips, corpus noise or an indication that the Italian language is moving towards a more satellite framed oriented pattern in event construal of motion. Gestures may help shed new light on the issue of how Italian speakers conceptualize and express motion in speech and gesture.

References.

Aske, J. (1989). *Path predicates in English and Spanish: A closer look.* Paper presented at the Proceedings of the Berkeley Linguistic Society 15.

Cadierno, T., & Ruiz, L. (2006). Motion events in Spanish l2 acquisition. *Annual Review of Cognitive Linguistics, 4*, 183-216.

Folli, R. (2008). Complex PPs in Italian. In A. Asbury, J. Dotlacil, B. Gehrke & R. Nouwen (Eds.), *Syntax and Semantics of Spatial P.* (pp. 197-221): John Benjamins Publishing Company.

Gullberg, M. (2011). Thinking, speaking and gesturing about motion in more than one language. In A. Pavlenko (Ed.), *Thinking and speaking in two languages* (pp. 143-169). Bristol: Multilingual Matters.

Iacobini, C. (2010). The number and use of manner verbs as a cue for typological change in the strategies of motion events encoding. In G. Marotta, A. Lenci, L. Meini & F. Rovai (Eds.), *Space in Language: Proceedings of the Pisa International Conference* (pp. 495-514). Pisa: Edizioni ETS.

Iacobini, C., & Masini, F. (2006). The emergence of verb-particle constructions in Italian: locative and actional meanings. *Morpholog, 16*, 155-188.

Rossini, N. (2005). *Phrasal verbs or words? Towards the analysis of gesture and prosody as indexes of lexicalisation.* Paper presented at the On-line Proceedings of the 2nd ISGS Conference "Interacting Bodies", Lyon, France.

Slobin, D. I. (1997). Mind, code, and text. In J. Nybee, J. Haiman & S. A. Thompson (Eds.), *Essays on language function and language type: Dedicated to T. Givón* (pp. 437-467). Amsterdam/Philadelphia: John Benjamins.

Slobin, D. I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Strömqvist & L. Verhoeven (Eds.), *Relating events in narrative: Typological and contextual perspectives* (pp. 219-257). Mahwah, NJ: Lawrence Erlbaum Associates.

Slobin, D. I., & Hoiting, N. (1994). *Reference to movement in spoken and signed languages: Typological considerations.* Paper presented at the Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society.

Talmy, L. (2000). *Toward a Cognitive Semantics* (Vol. II). Cambridge: The MIT Press.

Wessel-Tolvig, B. (2013). Boundary Ball: An animated stimulus designed to elicit motion with boundary crossing situations. University of Copenhagen.

Wessel-Tolvig, B. (2014). *Up, down, in & out: Following the Path of motion in Danish and Italian.* Paper presented at the 1st European Symposium on Multimodal Communication, Valletta, Malta.