

Book of Abstracts MMSYM 2019

Ambrazaitis, Frid & House , Multimodal prominence ratings: effects of screen size and audio device	2
Beukeleers, Hsu, Brône, Feyaerts & Vermeerbergen , Complex composite depictions and their semiotic diversity: Evidence from gestures and signs	4
DeAngelis & Jakobsen , The impact of virtual reality on visuospatial abilities and free recall	6
De Geldre & Oben , Multimodal predictors for conversational dominance.....	7
Ferrara , The semiotic diversity of signed language interactions and what this means for the study of multimodal communication and language theory	9
Frid, Svensson Lundmark, Ambrazaitis & House , EMA-based head movements, word accent and vowel length.....	11
Hodge, Schembri, Johnston & Cormier , Semiotic diversity in signed utterances: evidence from conversations in the British Sign Language Corpus	12
Lausberg, Reinecke, Kryger-Krämer & Skomroch , What the patient's gestures reflect about her/his concepts of symptoms and relevant relationships.....	14
Lepeut , Crossroads between Gesture, Sign, and Language.....	16
Leung, Fu & Monahan , Effect of Visual Integration of Pitch Contour in Mandarin Tone Perception.....	18
Murphy , Practical Implications of Second Modality Second Language (M2L2) Acquisition in Pedagogical Contexts: A Call for Research.....	20
Navarretta, Paggio & Wessel-Tolvig , Dialogue Acts in a First Encounter Corpus	22
Pouw & Dixon , Gesture networks: Introducing dynamic time warping and network analysis for the kinematic study of gesture ensembles	24
Rasenberg, Dingemans & Özyürek , Lexical and gestural alignment in collaborative referring	25
Rosborough & Wimmer , Multimodal Contingency in Developing Second Language Learning and Literacy	27
Rossano , Where should I begin? Calibrating the design of first actions in conversation	29
Valeiras-Jurado , When pointing becomes more than pointing: multimodal evaluation in product pitches.....	30
Van der Sluis & Redeker , The PAT Annotation Model for Multimodal Instructions	32
Von Au, Hümmer, Garling, Lausberg & Helmich , Are emotions hemispheric specialized?	35

Multimodal prominence ratings: Effects of screen size and audio device

Gilbert Ambrazaitis¹, Johan Frid² and David House³

¹Linnaeus University, Växjö, ²Lund University Humanities Lab, ³KTH (Royal Institute of Technology), Stockholm, Sweden

[gilbert.ambrazaitis@lnu.se, johan.frid@humlab.lu.se, davidh@speech.kth.se]

Keywords: audio-visual perception, crowdsourcing, web-based, inter-rater reliability, headphones

Prosodic prominence is a multimodal phenomenon involving both acoustic and kinematic dimensions. In order to study the multimodal nature of prominence, we need to collect prominence ratings based on audio-visual speech material from large groups of speakers. This is feasible by means of a web-based crowdsourcing set-up, allowing volunteers to participate using a private computer or mobile phone. However, this freedom also implies a certain reduction of experimental control due to variation in hardware used by the raters.

In this pilot study we explore potential effects of two hardware features – screen size and audio device (headphones vs. loudspeakers) – on multimodal prominence ratings. To this end, 16 brief clips from Swedish television news (218 words in total) were rated by 31 native Swedish volunteers using a web-based set-up. In our GUI, orthographic representations of the text were displayed below the video player. Each word was to be rated as either non-prominent, moderately prominent, or strongly prominent, by means of clicking on the word in question until the desired prominence level was encoded through a specific color (yellow: moderate; red: strong). Participants were free to use a mobile phone, a tablet, or a computer, and headphones or loudspeakers, and we collected information about their hardware using a questionnaire. In addition, we automatically logged the screen size of the participant's computer/phone.

We applied two different approaches to analyze the participant's rating behavior as a function of the hardware features under discussion. First, we calculated a selection of five variables from the raw prominence ratings: (i) the sum of all ratings (over all 218 words), (ii) the percentage of words rated as (moderately or strongly) prominent, (iii) among prominent words, the proportion of words rated as strongly prominent, and (iv-v) the relative prominence rating of two selected words. Effects of screen size and audio device on these variables were analyzed using linear regression models. Second, we calculated inter-rater reliability for multiple raters using Fleiss' kappa, both for all raters as a reference and for subgroups concerning audio device and screen size.

The results reveal a significant model fit for variable (iii) defined above (proportion of strong ratings; $F(5;21) = 5.332$; $p=.0022^{**}$), suggesting a significantly higher proportion of strong prominent ratings obtained with loudspeakers (34.0% of words rated as prominent on average) compared to with headphones (18.3%; $t=2.944$; $p=.0073^{**}$), as well as with medium size screens (34.2%) compared to with small screens (24.4%; $t=2.433$; $p=.0232^*$); however, the proportion of strong prominent ratings tended to be lowest with large screens (14.2% on average). Effects of screen size were also reflected in inter-rater reliability, revealing the highest κ for users with medium-sized screens ($\kappa=.566$, when ratings are recoded to a binary decision) compared to large ($\kappa=.485$) and small screens (mobile phones, $\kappa=.437$). However, inter-rater reliability was less affected by the listening condition (headphones vs. loudspeakers).

To conclude, the choice of hardware might have effects on multimodal prominence ratings, which has to be considered in crowdsourcing approaches. More detailed results will be presented at the conference.

Complex composite depictions and their semiotic diversity: Evidence from gestures and signs

Inez Beukeleers, Hui-Chieh Hsu, Geert Brône, Kurt Feyaerts & Myriam Vermeerbergen

KU Leuven

[inez.beukeleers@kuleuven.be, huichieh.hsu@kuleuven.be, geert.brone@kuleuven.be,
kurt.feyaerts@kuleuven.be, myriam.vermeerbergen@kuleuven.be]

Keywords: depiction, gesture, Flemish Sign Language, composite depiction, semiotic diversity

Recent studies have demonstrated the interlocutor utilizes and coordinates symbolic, iconic, and indexical resources in meaning communication, through descriptions, depictions, and indications, respectively (e.g. Ferrara & Hodge, 2018). In particular, Clark (2016) identifies depicting as a basic method of communication where the interlocutor expresses meaning by creating a physical analog of the depicted scene. In this talk, we explore the notion of depicting, drawing on evidence from Flemish Sign Language (VGT) narratives and gestures in face-to-face interaction in American English, offering corpus-based insights into the comparability thereof.

Within the domain of sign linguistics, researchers have argued that signers draw on fully-lexical signs, i.e. signs that are conventionalized in form and meaning and entrenched within a community (e.g. Johnston & Schembri, 2010) when describing, i.e. *telling* meaning, and on more iconic structures, i.e. depicting signs and enactment, when depicting, i.e. *showing* meaning (e.g. Brennan, 1990 for BSL; Johnston, 1996 and Johnston & Schembri, 2007 for Auslan; Vermeerbergen, 1996 for VGT). In the current study we aim to broaden the notion of depicting within our field and explore the different semiotic resources that signers can draw on when showing meaning. Based on an analysis of 10 VGT narratives from the Corpus VGT (Van Herreweghe et al., 2015), we show that signers do not only use depicting signs and enactment, but also modify fully-lexical signs, pointing signs, and/or indicating verbs in order to depict the movement and/or location of a referent. Moreover, we found that signers tend to combine these different types of semiotic signs in the creation of complex composite depictions.

As for gestures, parallel phenomena emerge from our corpus of American TV talk shows, where focus is laid on non-verbal depictions embedded in speech; that is, cases where gestures communicate meaning without temporally co-occurring speech, a much-overlooked field in gesture studies (but see Keevallik, 2010; Ladewig, *forthc.*). Incorporation of descriptions and indications is evident in tokens of such depictions. In addition, series of consecutive depictions are observed building up complex composite depictions, which consist of a base — which serves as the common thread throughout the series — and, staged on top of the base, several elaboration-depictions — which either contribute to the composite meaning of the series individually, or build and elaborate on prior elaborations in the same series. These composite depictions, though not novel to the sign linguist, align with recent findings in spoken language linguistics demonstrating gestures' recurrent patterns and potential to form complex constructions (e.g. Bressemer, 2014; Müller, 2017; Ruth-Hirrel &

Wilcox, 2018), further showing that even gestures created and assembled *ad hoc* can be combined to create larger structures with complex meanings.

In addition to comparing empirical data and identifying parallel findings from the perspectives of sign linguistics and gesture studies, the present study more importantly singles out phenomena that are familiar in one field but alien in the other, calling attention to neglected phenomena in both fields.

References:

- Brennan, M. (1990). *Word formation in British Sign Language*. Stockholm: University of Stockholm.
- Bressemer, J. (2014). Repetitions in gesture. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & J. Bressemer (Eds.), *Body – Language – Communication*, Vol. 2 (pp. 1641–1649). Berlin: De Gruyter Mouton.
- Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review*, 123(3), 324–347.
- Ferrara, L., & Hodge, G. (2018). Language as description, indication and depiction. *Frontiers in Psychology*, 9(716).
- Johnston, T. (1996). Function and medium in the forms of linguistic expression found in a sign language. In W. H. Edmonson & R. Wilbur (Eds), *International Review of Sign Linguistics*, Vol. 1 (pp. 57–94). Mahwah, NJ: Lawrence Erlbaum.
- Johnston, T., & Schembri, A. (2007). *Australian Sign Language: An Introduction to Sign Language Linguistics*. Cambridge: Cambridge University Press.
- Johnston, T., & Schembri, A. (2010). Variation, lexicalization and grammaticalization in signed languages. *Language et société*, 1(131), 19–35.
- Keevallik, L. (2010). Bodily quoting in dance correction. *Research on Language and Social Interaction*, 43(4), 401–426.
- Ladewig, S. (forthc.). *Integrating Gestures: Cognitive Grammar Multimodal*. Berlin: De Gruyter Mouton.
- Müller, C. (2017) How recurrent gestures mean: Conventionalized contexts-of-use and embodied motivation. *Gesture*, 16(2), 277–304.
- Ruth-Hirrel, L., & Wilcox, S. (2018). Speech-gesture constructions in cognitive grammar: The case of beats and points. *Cognitive Linguistics*, 29(3), 453–493.
- Van Herreweghe, M., Vermeerbergen, M., Demey, E., De Durpel, H., Nyffels, H. & Verstraete, S. (2015). *Het Corpus VGT. Een digital open acces corpus van video's en annotaties van Vlaamse Gebarentaal*, ontwikkeld aan de Universiteit Gent i.s.m. Katholieke Universiteit Leuven. Accessible at: www.corpusvgt.be.
- Vermeerbergen, M. (1996). *ROOD KOOL TIEN PERSOON IN: Morfo-syntactische Aspecten van de Vlaamse Gebarentaal*. Vrije Universiteit Brussel: Doctoral dissertation.

The impact of virtual reality on visuospatial abilities and free recall

Kaela DeAngelis & Terne Thorn Jakobsen

University of Copenhagen

[dws342@alumni.ku.dk, ktj250@alumni.ku.dk]

Keywords: Virtual reality, block design, visuospatial, free recall, oculus rift

Virtual Reality (VR) technologies using head-mounted displays are increasingly under evaluation for integration into teaching basic and complex skills and procedures over traditional methods in many different fields and industries. However, the impact of such immersive experiences on complex human cognitive processes that are important to learning must be investigated to understand if trainees can adapt their perceptions and movements to the changes virtual technologies and simulations bring without sacrificing performance. This study outlines and evaluates a methodological approach to discern differences in perceptual and short-term memory performances in virtual environments against a non-virtual environment, by adapting Samuel C. Kohs' original Block Design Task (1920), and related experiments by Fenouillet and Rozenchwajg (2015) and also Rozenchwajg and Corroyer (2002). An experiment involving 24 participants was conducted in which two groups were exposed to a block design task but in either a virtual environment or non-virtual environment. Aspects examined in the experiment were visuospatial abilities and free recall. The results indicated that there was no difference between visuospatial and free recall performances of subjects when tested in either the virtual environment or the non-virtual environment. However, testing time in the virtual environment took significantly more time. These results suggest that VR technology may be complementary, or suitable replacements, for unimodal and conventional teaching methods.

References:

- S. C. Kohs. The block-design tests. *Journal of Experimental Psychology*, 3(5):357–376, October 1920.
- F. Fenouillet and P. Rozenchwajg. Visual-spatial abilities and goal effect on strategies used to solve a block design task. *Learning And Individual Differences*, 39:158–163, April 2015.
- P. Rozenchwajg and D. Corroyer. Strategy development in a block design task. *Intelligence*, 30(1):1–25, 2002.

Multimodal predictors for conversational dominance

Kaat De Gueldre & Bert Oben

KU Leuven

[kaat.degueldre@student.kuleuven.be, bert.oben@kuleuven.be]

Keywords: dominance, multimodality, gesture, eye gaze

Research on the topic of conversational dominance has revealed that dominance correlates with a battery of behavioural features. Concerning verbal behaviour, dominant speakers typically talk more, longer, louder, introduce new content words, ask questions, elicit turns, switch conversational topics, interrupt or ignore their partner, start talking during silences, etc. (Dunbar & Abra 2010, Guerrero 2005, Nakano & Fukuhura 2012). At the non-verbal level, dominance is associated with smiling, having more relaxed or more masculine facial expressions, leaning towards or touching the partner, looking at the partner while speaking, not breaking mutual gaze etc. (Dovidio 1985, Knapp & Hall 2010, Watkins, Jones & DeBruine 2010). Overarching the verbal and non-verbal level also phenomena such as interactional alignment, i.e. copying behaviour, seem to play a role: dominant speakers align less to their non-dominant conversational partners than the other way around (Worgan & Moore 2011).

With the current study we aim to add to the existing literature by :

considering several multimodal features at the same time (allowing to weigh their relative explanatory power)

considering hand gestures and gaze behaviour as possible multimodal features

studying dominance in conversational settings without prefixed roles (i.e. without existing power relations as in teacher-student, doctor-patient or judge-lawyer interactions)

Using data from the Insight Interaction corpus (Brône & Oben 2015), we presented 2-minute video clips from a brainstorm session between two speakers to naïve raters. Those raters viewed the clips online and were asked to intuitively indicate how dominant each of the conversational partners in the clips were. Every clip was rated by at least 15 raters and the dominance score per speaker was calculated as the average of all rating values for that speaker. Subsequently, we annotated a number of multimodal features for each video clip. Those features include number of words, amount of copying behaviour, number of hand gestures, types of hand gesture, gesture size, gaze behaviour as speaker and gaze behaviour as listener.

What we found, is that most of the multimodal features correlate with conversational dominance as expected, but that the amount of speech (measured as number of words) was such a good predictor for dominance that the effects from the other factors did not significantly contribute in explaining dominance in our regression models. That result highlights the relevance of considering multiple factors at the same time and appears to suggest that, at least for conversational settings without predetermined conversational or institutional roles, the amount of speaking time offers the most efficient resources for speakers that want to assert their dominance.

References:

- Brône G., Oben B. (2015). InSight Interaction: a multimodal and multifocal dialogue corpus. *Language Resources and Evaluation* 49 (1), 195-214.
- Dovidio, J. & Ellyson, S. (1985). Patterns of Visual Dominance Behavior in Humans. In: S. Ellyson & J. Dovidio (Eds.), *Power, Dominance, and Nonverbal Behavior*, 129-149, New York: Springer- Verlag.
- Dunbar, N. & Abra, G. (2010). Observations of Dyadic Power in Interpersonal Interaction. *Communication Monographs* 77, 657-684.
- Guerrero, L. (2005). Observer ratings of nonverbal involvement and immediacy. In: V. Manusov (Ed.), *The sourcebook of nonverbal measures: Going beyond words*, 221-235, Mahwah: Lawrence Erlbaum.
- Knapp, M. & Hall, J. (2010). *Nonverbal Communication in Human Interaction*. Fort Worth: Harcourt College Publishers
- Nakano, Y. & Fukuhura, Y. (2012). Estimating Conversational Dominance in Multiparty Interaction. Proceedings of the 14th ACM international conference on Multimodal interaction, 77-84.
- Watkins, C., Jones, B. & DeBruine, L. (2010). Individual differences in dominance perception: Dominant men are less sensitive to facial cues of male dominance. *Personality and Individual Differences* 49, 967-971.
- Worgan, S. & Moore, R. (2011). Towards the detection of social dominance in dialogue. *Speech Communication* 53, 1104-1114.

The semiotic diversity of signed language interactions and what this means for the study of multimodal communication and language theory

Lindsay Ferrara

Norwegian University of Science and Technology

[lindsay.n.ferrara@ntnu.no]

Keywords: signed language, interaction, semiotic, indexicality, depiction

From the early days of research on signed languages, linguists have been challenged with the observation that signers use their whole bodies to communicate during signed language interaction. It is common for signers to combine hand movements with facial expressions, eye gaze direction, and other bodily movements to talk about the world (much like speakers). Early work attempted to fit these actions into structuralist paradigms focused on describing the paradigmatic and componential structure of propositional meanings (e.g., Klima & Bellugi, 1979). However, a more recent turn has seen researchers embracing the ‘multimodal’ view pioneered by Johnston and Schembri (1999), Liddell (2003), and Taub (2001), among others, in order to explore how such complexity manifests and what this means for our understandings of how signed languages work (and by extension, face-to-face interaction more generally). This turn has been coupled with work and advancements in other related fields such as linguistics, gesture and multimodality studies, semiotics, cognitive science, and anthropology.

In this talk, I will review research at the intersections of this new cross-disciplinary thinking, which demonstrates potential new directions for linguistic theory. By using various signed languages as evidence, I will first illustrate how the multimodal actions recruited by signers are coordinated across specific spatiotemporal contexts and combined into ‘composite utterances’ (Enfield, 2009; see e.g., Hodge & Johnston, 2014; Johnston, 2018; Ferrara, 2019). Some of these actions are highly conventionalized, while others are less conventionalized and more context-dependent. Importantly, they also vary in their semiotic mode, that is, they instantiate different types of semiotic signs that describe, depict, and index referential and contextual meanings (Peirce, 1955; see also Clark, 1996; Enfield, 2009; Ferrara & Hodge, 2018). And as integrated parts of composite utterances, these different semiotic signs are used to communicate and do various interactional, ‘linguistic’ work (e.g., Ferrara & Johnston, 2014; Hodge, Ferrara, & Anible, 2019; Puupponen, 2019).

What these studies show is that signers routinely index and depict meanings and that these actions interact with each other to shape symbolic structure. The semiotic diversity of face-to-face multimodal interaction is thus essential to understanding how (signed) languages work (Kendon, 2014). In this way then, linguistic theory cannot only focus on the symbolic and fully conventional aspects of language use but must also consider the contextual rootedness and emergent meaningfulness of indexicality and depiction (e.g., Silverstein, 1976; Dingemanse, 2015; Ferrara & Halvorsen, 2017; Ferrara & Hodge, 2018). Only then can we begin to unite the communicative practices resulting from diverse ways of being (e.g., deaf, hearing) into a global theoretical framework that facilitates continuity with other lines of scientific enquiry (Ferrara & Hodge, 2018).

References:

- Clark, Herbert H. (1996). *Using language*. Cambridge, England: Cambridge University Press.
- Dingemanse, Mark. (2015). Ideophones and reduplication: Depiction, description, and the interpretation of repeated talk in discourse. *Studies in Language*, 39(4), 946-970. doi:10.1075/sl.39.4.05din
- Enfield, Nick J. (2009). *The anatomy of meaning: Speech, gesture, and composite utterances*. Cambridge: Cambridge University Press.
- Ferrara, Lindsay. (2019). Coordinating signs and eyegaze in the depiction of directions and spatial scenes by fluent and L2 signers of Norwegian Sign Language. *Spatial Cognition and Computation: An Interdisciplinary Journal*, 9(3), 220-251. doi:https://doi.org/10.1080/13875868.2019.1572151
- Ferrara, Lindsay, & Halvorsen, Rolf Piene. (2017). Depicting and describing meanings with iconic signs in Norwegian Sign Language. *Gesture*, 16(3), 371-395. doi:https://doi.org/10.1075/gest.00001.fer
- Ferrara, Lindsay, & Hodge, Gabrielle. (2018). Language as description, indication, and depiction. *Frontiers in Psychology*, 9: 716. doi:10.3389/fpsyg.2018.00716
- Ferrara, Lindsay, & Johnston, Trevor. (2014). Elaborating who's what: A study of constructed action and clause structure in Auslan (Australian Sign Language). *Australian Journal of Linguistics*, 34(2), 193-215. doi:10.1080/07268602.2014.887405
- Hodge, Gabrielle, & Johnston, Trevor. (2014). Points, depictions, gestures and enactment: Partly lexical and non-Lexical signs as core elements of single clause-like units in Auslan (Australian Sign Language). *Australian Journal of Linguistics*, 34(2), 262-291. doi:10.1080/07268602.2014.887408
- Hodge, Gabrielle, Ferrara, Lindsay, & Anible, Benjamin. (2019). The semiotic diversity of doing reference in a deaf signed language. *Journal of Pragmatics*, 143, 33-53. doi:https://doi.org/10.1016/j.pragma.2019.01.025
- Johnston, Trevor, & Schembri, Adam. (1999). On defining Lexeme in a Signed Language. *Sign Language and Linguistics*, 2(2), 115-185. doi:10.1075/sll.2.2.03joh
- Johnston, Trevor. (2018). A corpus-based study of the role of headshaking in negation in Auslan (Australian Sign Language): Implications for signed language typology. *Linguistic Typology*, 22(2), 185-131. doi:https://doi.org/10.1515/lingty-2018-0008
- Kendon, Adam. (2014). Semiotic diversity in utterance production and the concept of 'language'. *Phil. Trans. R. Soc. B*, 369, 20130293. doi:http://dx.doi.org/10.1098/rstb.2013.0293
- Klima, Edward, & Bellugi, Ursula. (1979). *The signs of language*. Cambridge, MA: Harvard University Press.
- Liddell, Scott K. (2003). *Grammar, gesture, and meaning in American Sign Language*. New York: Cambridge University Press.
- Peirce, Charles S. (1955). *Philosophical writings of Peirce*. New York, NY: Dover Publications.
- Puupponen, Anna. (2019). Towards understanding nonmanuality: A semiotic treatment of signers' head movements. *Glossa: A journal of general linguistics*, 4(1), 39. doi:https://doi.org/10.5334/gjgl.709
- Silverstein, Michael. (1976). Shifters, linguistic categories, and cultural description. In Keith H Basso & Henry A Selby (Eds.), *Meaning in anthropology* (pp. 11-55). Albuquerque, NM: University of New Mexico Press.
- Taub, Sarah. (2001). *Language from the body: Iconicity and metaphor in American Sign Language*. New York: Cambridge University Press.

EMA-based head movements, word accent and vowel length

Johan Frid¹, Malin Svensson Lundmark², Gilbert Ambrazaitis³ and David House⁴

¹*Lund University Humanities Lab*, ²*Centre for Languages and Literature, Lund University*,
³*Linnaeus University, Växjö*, ⁴*KTH, Stockholm*

[johan.frid@humlab.lu.se, malin.svensson_lundmark@ling.lu.se,
gilbert.ambrazaitis@lnu.se, davidh@speech.kth.se]

Keywords: multimodal prosody, pitch accent, head nod, GAM, articulography

This paper describes on-going work in the field of multimodal prosody carried out by means of simultaneous recordings of speech acoustics, articulation and head movements. People naturally move their heads when they speak, and head movements have been found both to correlate strongly with the pitch and amplitude of the speaker's voices and to convey linguistic information. Here, we report on a study that explores how head movement patterns vary and co-occur with lexical pitch accents (and their acoustic correlates F0 and intensity) and vowel length. The study uses data from Swedish, where there are both two lexical pitch accents and two vowel lengths that differ phonologically.

We use EMA (Electromagnetic articulography), which allows for high sample rates, accurate synchronisation of kinematic and acoustic recordings, as well as three-dimensional movement data. Kinematic data is obtained by gluing small sensors on the speakers' articulators (tongue, lips, jaw). Head movement data is obtained by similar sensors on the nose ridge and behind the ears, which allows us to capture the angle of the tilt of the head.

Articulatory data was collected from 18 South Swedish speakers (12 female) using a Carstens AG501. Each speaker read leading questions + sentences containing a target word from a prompter (presented eight times in random order), an arrangement employed to put a contrastive focus onto the last element in the target sentence. This left the target word in a low-prominence inducing context, hence controlling for possible effects of sentence intonation.

For this study we used eight target words where pitch accent and vowel length were cross-matched so that there were two cases of each combination of word accent category and vowel length category. All words shared the similar word-initial C /m/, followed by a vowel that was either /a/ or /ɑ:/. The target words were segmented and time-normalized between 0 to 1 and the head tilt angle (sagAng) was normalized for each speaker by z-transforming the angles per speaker. Spatial movements were analysed using Generalized Additive Models, which we used to test if there were effects of segmental position (C versus V in the first syllable), word accent (1 or 2) and vowel length (short or long) on sagAng. Models were fit using the maximum likelihood (ML) estimation method.

The Chi-Square test on the ML scores indicates that a model with the word accent distinction is significantly better than a model without it ($X^2(4.00)=632.796$, $p<2e-16^{***}$). Similarly, a model with vowel length distinction is significantly better than a model without it ($X^2(4.00)=820.997$, $p<2e-16^{***}$). Finally, a model with segmental position is significantly better than a model without it ($X^2(8.00)=173.316$, $p<2e-16^{***}$).

The results indicate that head nod patterns that occur in synchronisation with the stressed syllable of spoken words differ with respect to word accent, vowel length and segmental position. This could possibly point to an effect of F0 and intensity on the head nod movements.

Semiotic diversity in signed utterances: evidence from conversations in the British Sign Language Corpus

Gabrielle Hodge¹, Adam Schembri², Trevor Johnston³ & Kearsy Cormier¹

¹*DCAL, University College London*, ²*University of Birmingham, Birmingham*, ³*Macquarie University, Sydney*

[g.hodge@ucl.ac.uk, a.schembri@bham.ac.uk, trevor.johnston@mq.edu.au, k.cormier@ucl.ac.uk]

Keywords: deaf, multimodal, semiotic, sign language, corpora

Signers and speakers coordinate a broad range of communicative actions during their face-to-face interactions, not all of which are arbitrary, conventionalised symbols (see e.g. Parmentier, 1994). Signed and spoken language use can therefore be understood as 'actioned' via three different methods of signalling: describing, indicating, and depicting (Clark, 1996; Ferrara & Hodge, 2018). Each method can be used alone or in various combinations during the creation of multimodal 'composite utterances' (Enfield, 2009). These utterances can be analysed as interactionally-driven clauses (Thompson & Couper-Kuhlen, 2005) containing a 'mixed syntax' of semiotic elements with varying degrees of conventionality (Slama-Cazacu, 1973). However, most analyses of signed languages to date have focused on intuitions from small numbers of signers, monologic retellings and/or spontaneous narratives. Unlike spoken languages (e.g. Hakulinen & Selting, 2005), little is known about the grammar of signed language conversations.

Here we consider how deaf signers draw upon these methods of signalling to co-create composite utterances during dyadic conversations. We use data from twenty pairs of deaf signers from four regions documented in the BSL Corpus (Schembri et al., 2014). Preliminary results based on 2,550 composite utterances from ten pairs of signers in two regions illustrates four types: (1) single, stand-alone propositional clauses; (2) multi-propositional utterances where one clause is hypotactically linked to an adjacent clause; (3) shared utterances, where the content is shared by two signers; and (4) pragmatically-oriented backchanneling moves. Analysis of each type reveals syntagmatic and paradigmatic regularities for describing, indicating and depicting, many of which may be explained by the moment-by-moment unfolding of discourse (Enfield, 2009). For example, two signers from Manchester co-construct a shared memory by using all three strategies to elaborate different physical qualities of the same referent. This research contributes to our understanding of the semiotic diversity of face-to-face interactions and aligns with research on the embodied nature of grammar (e.g. Keevallik, 2018).

References:

- Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620539
- Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. Cambridge: Cambridge University Press.
- Ferrara, L. & G. Hodge. (2018). 'Language as description, indication, and depiction'. *Front. Psychol.* 9:716. doi:10.3389/fpsyg.2018.00716
- Hakulinen, A., & Selting, M. (Eds.). (2005). *Syntax and Lexis in Conversation: Studies on the use of linguistic resources in talk-in-interaction*. Amsterdam: John Benjamins.
- Keevallik, L. (2018). 'What does embodied interaction tell us about grammar?' *Research on language and social interaction*, 51(1): 1 – 21. <https://doi.org/10.1080/08351813.2018.1413887>

Parmentier, R. J. (1994). *Signs in Society: Studies in Semiotic Anthropology*. Indianapolis, IN: Indiana University Press.

Schembri, A., Fenlon, J., Rentelis, R., & Cormier, K. (2014). *British Sign Language Corpus Project: A corpus of digital video data and annotations of British Sign Language 2008-2014 (Second Edition)*. London: University College London. <http://www.bsllcorpusproject.org>

Slama-Cazacu, T. (1973). *Introduction to psycholinguistics*. Berlin: Walter de Gruyter.

Thompson, S. & E. Couper-Kuhlen. (2005). 'The clause as a locus of grammar and interaction'. *Discourse Studies*, 7(4-5): 481-505. doi:10.1177/1461445605054403

What the patient's gestures reflect about her/his concepts of symptoms and relevant relationships

Hedda Lausberg¹, Katharina C. H. Reinecke¹, Monika Kryger-Krämer¹,
Harald Skomroch¹

¹*Department of Neurology, Psychosomatic Medicine and Psychiatry, German Sport University
Cologne, Germany*

[h.lausberg@dshs-koeln.de, k.reinecke@dshs-koeln.de, monikakryger86@googlemail.com,
skomroch@googlemail.com]

Keywords: co-speech gestures; psychotherapy; mental concepts; NEUROGES; health communication

In health communication, the patient's gestures reflect her/his feelings and thoughts. As an example, pictorial gestures reveal mental concepts of symptoms, significant relationships, developments, etc. Specifically in psychotherapy, the patient's gestural expression provides information that the patient is often not aware of and that might indicate unconscious conflicts. Accordingly, in these settings gesture-speech mismatches occur frequently. In order to effectively use the gestural information for diagnostics and therapy, it is necessary to explicitly recognize the gestural information.

The data corpus comprises complete psychodynamic psychotherapies (25 sessions each therapy) of out-patients with social phobia (n = 24) and eating disorders (n = 3). The first and the pre-last psychotherapy sessions were submitted to gesture analysis. In addition, if there was a substantial formal and conceptual change in a specific gestural expression, it was followed up in the course of the complete psychotherapy e. g., at the beginning of psychotherapy a patient with eating disorders displayed *self-deictics* pointing to the belly, at the end of the psychotherapy pointing to the sternum. In psychotherapy data, the frequently occurring gesture-speech mismatches constitute a specific challenge for multimodal analysis. As in general the analysis of gestures is biased by the verbal context rather than vice versa, in the present study, the psychotherapy patients' gestures were first coded independently from speech and only in a second step the relation to the verbal context was analysed. The NEUROGES system (Function category) was chosen for the analysis of the gestures as it enables an objective and reliable gesture analysis based on movement criteria, i. e., primarily independent of speech. The analysis was conducted by two independent NEUROGES-certified raters. The interrater agreement, as measured with EasyDiag, was substantial.

Gesture-speech mismatches were present and indicative of unconscious conflicts, a. o. discrepancies between the patient's self-perception and his/her perception by others. *Body-deictics* were used to localize psychosomatic problems attributed to certain parts of the body, and modifications in *self-deictics* indicated changes in the patient's self-concept. *External target deictics* were used to localize significant others in the gesture space, reflecting their psychological position, e. g. proximity, relative to the patient. Psychological developments in these relationships were accompanied by changes of the significant other's position in the gesture space. *Egocentric directions* revealed how the patient moved others relative to her/himself. *Pantomime* gestures, in which the gesturer acted as if, were associated with high emotional engagement, especially when re-enacting another person's action, e. g. a gestural identification with the aggressor. *Form presentation* and *spatial relation presentation* gestures revealed how the patient grouped relevant others, e. g. conceptualizing the family as one unit

or as consisting of opposing subgroups. Finally, *motion quality presentation* gestures revealed concepts about mental developments.

To conclude, gestures provide relevant information about the patient's concepts of their symptoms, relationships to significant others, and developments. In the context of psychotherapy, the gestural information is of specific value as it provides – as in the case of unconscious conflicts – information that the patient does not verbalize.

Crossroads between Gesture, Sign, and Language: The Palm-Up in the management of spoken and signed interaction in French-speaking Belgium.

Alysson Lepeut

University of Namur & F.R.S.-FNRS

[alysson.lepeut@unamur.be]

Keywords: sign language, gesture, multimodal corpora, social interaction, pragmatics

This paper lies at the gesture-sign interface. It aims at discussing the implications of approaching the concept of *gesture* through the contrastive multimodal analysis within and across two modalities and languages in French-speaking Belgium, namely, French Belgian Sign Language (LSFB) and spoken French (BF), of a shared gesture: the Palm-Up (PU). One particularity about PU is that, unlike content-oriented gestures, PU is “interactional in nature” (Cooperrider et al. 2018, p. 5). The current focus is on how PU participates in the management of signed interaction, and how its uses compare to those used in spoken interaction. Previous work has shown that PU’s *interactive* functions include – but are not limited to – regulating turn-taking, marking new or shared information, acknowledging and eliciting shared understanding through feedback. Yet, these interactional components in gesture and sign language conversation remain understudied. The objectives are: (i) to study PU frequency; (ii) to investigate its interactive functions and (iii) to examine if specific gaze directions (e.g. addressed or floating) co-occur with some PUs and/or functions. Annotation and analysis of approximately 3 hours of video-recorded material were conducted. The data are drawn from (1) The CorpAGEst Corpus with 4 speakers (≥ 75 y. old) recorded in semi-directed interviews at their home; (2) The LSFB Corpus including dyadic conversations of 2 pairs of signers (≥ 66 y. old) recorded in the university lab; and (3) The FRAPé Corpus composed of 2 pairs of speakers (≥ 66 y. old) collected in the university lab as well. Comparing data from (2) and (3) allowed conducting the first multimodal cross-linguistic study between LSFB and BF. One strength in this approach is that such comparison implies two facets: “commonalities of gesture and sign resulting from a shared medium of expression”, on the one hand, and “commonalities resulting from language use within and across language communities”, on the other (Müller 2018, p.15). Preliminary results suggest that PU is extremely multifunctional in nature, which echoes McKee and Wallingford’s previous study (2011). We found that all participants use PU to manage interaction, but they do it for different communicative purposes. While some interactive functions overlap, signers use PUs especially to regulate their turn-taking system and to provide feedback. Speakers, however, use it for marking new or shared information. In line with other researchers (Müller 2018), this study shows how putting gesture and sign on common ground favors commonalities between them, and how examining an interactive gesture in a signed and spoken language reinforces the argument for gesture as part of linguistic activities, and as to what it means for spoken and signed languages to be *gestural languages*.

References:

- Cooperrider, K., Abner, N., & Goldin-Meadow, S. (2018). The palm-up puzzle: Meanings and origins of a widespread form in gesture and sign. *Frontiers in Communication*, 3(23).
- McKee, R. L., & S. Wallingford. (2011). ‘So, well, whatever’: Discourse functions of palm-up in New Zealand Sign Language. *Sign Language & Linguistics*, 14(2).

Müller, C. (2018). Gesture and Sign: Cataclysmic Break or Dynamic Relations? *Frontiers in Psychology*, 9(1651).

Effect of Visual Integration of Pitch Contour in Mandarin Tone Perception

Andrea Michelle Leung, Zhanao Fu, Philip J. Monahan

University of Toronto

[andreamichelle.leung@mail.utoronto.ca, zhanao.fu@mail.utoronto.ca,
philip.monahan@utoronto.ca]

Keywords: tone sandhi, lexical tone, perception, underlying representation, surface representation

Previous studies have found that tone pitch contours (pitch gestures) improve English speakers' ability to acquire Mandarin lexical tones (Morett & Chang, 2015), and that the articulation of Mandarin tone 3 sandhi demands more phonological encoding (Zhang, Xia, & Peng, 2015). However, there is a knowledge gap between the visual integration of pitch contours and auditory modalities, and what effect it would have on the perception of Mandarin tone sandhi.

This study focuses on how visual perception of tone contours would affect the audio perception of Mandarin tone 3 sandhi. In Mandarin Chinese, Tone 3 Sandhi (also called Third-Tone Sandhi) occurs when there are two consecutive Tone 3 syllables (i.e. both underlyingly "T₃T₃"), the first Tone 3 syllable would be pronounced as Tone 2 or Tone 2-like (i.e. surface representation of "T₂T₃") (Duanmu, 2007). In this study, native Mandarin speakers are recruited to perform an audio-visual matching task of Mandarin tone contours. They are asked to judge whether the visual representation of tone contours displayed on screen matches the auditory presentation of the word. The audio stimuli consist of 90 two-character words selected from Cai and Brysbaert's (2010) SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. Among these the list of words, 30 is allocated in the sandhi case (i.e. underlyingly "T₃T₃" words), 30 with non-sandhi case (i.e. "T₂T₃" on both the underlying and surface representation), and 30 words as the controlled condition (i.e. words in "T₁T₃" or "T₄T₃"). The tone contours displayed are either congruent or incongruent with the tone sandhi cases. Currently, two experiments are conducted, with the first experiment with instructions focused on words and the second experiment with instructions focused on phonetics, other aspects of the experiments were kept constant.

The pattern of results are almost identical irrespective of the change in instructions, and the combined results based on 57 participants are as follow: In terms of response accuracy, the average accuracy for the sandhi condition with the surface representation of tone contour displayed (21.81%) is significantly lower than the sandhi condition with the underlying representation of tone contour displayed (63.80%), whereas all non-sandhi cases and controlled conditions have over 90% of accuracy. The lower accuracy in tonal matching in the sandhi case shows that Mandarin speakers are more inclined to match the audio with the visual contours based on the underlying representation of the tones. This could be due to the lexical awareness of the conflicting underlying representation of tone sandhi reducing the accuracy in audio perception of its surface representation, at least in parts. This finding could shed light on the effect of multimodal integration on tone sandhi perception.

References:

Morett, L. M., & Chang, L. (2015). Emphasising sound and meaning: Pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience*, 30(3), 347-353.

Zhang, C., Xia, Q., & Peng, G. (2015). Mandarin third tone sandhi requires more effortful phonological encoding in speech production: Evidence from an ERP study. *Journal of Neurolinguistics*, 33, 149-162.

Duanmu, S. (2007). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Cai, Q., & Brysbaert M. (2010). SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. *PLoS ONE* 5(6): e10729. <http://doi:10.1371/journal.pone.0010729>

Practical Implications of Second Modality Second Language (M2L2) Acquisition in Pedagogical Contexts: A Call for Research

Jillian Murphy

Department of Linguistics, Northeastern Illinois University, Illinois, USA

[jmmurphy2@neiu.edu]

Keywords: American Sign Language, second language acquisition, second modality acquisition, bimodal bilingualism, language pedagogy

The phenomenon of acquiring a second modality in the context of second language acquisition is one that is severely understudied. Yet, the implications of studying the impact of second modality second language (M2L2) acquisition could be wide ranging. For example, very few studies have been conducted to explore what specific pedagogical needs related to M2L2 acquisition should be considered in sign language instruction. The study of M2 acquisition adds not only to the field of L2 acquisition but also to traditional theoretical fields in linguistics such as semantics and phonology. Results from recent M2L2 acquisition studies have revealed new information about the role of working memory in language acquisition, some of which are discussed in this presentation. These findings are of interest not only from a cognitive standpoint but also in terms of modeling how language is produced, in other words, depicting how an utterance gets from the lexicon to its surface form. Several questions arise concerning the study of M2L2 acquisition. For example, is learning a second modality during L2 acquisition more challenging than learning a second language with the same modality that one is accustomed to? Does M2 acquisition impact the speed with which a second language is acquired? Since determining when or even if a second language can be deemed “acquired” is problematic, a more appropriate question might be how does learning a second modality affect proficiency in a second language? What practical implications does learning a second modality during second language acquisition have for the instruction of signed languages? In other words, what pedagogical factors should be considered, and what measures might need to be implemented to account for any challenges associated with sign language instruction? This presentation serves both as an overview of some of the recent M2L2 acquisition research studies and a call for continued research in this area, particularly with regard to how M2L2 acquisition might be better accounted for in pedagogical contexts. Future studies in M2L2 acquisition research should seek to understand the challenges that both learners and instructors face in the teaching of sign languages, with consideration given specifically to modality. Understanding the practical needs of both learners and instructors from a pedagogical standpoint will help to properly set expectations and perhaps reduce misunderstanding in the L2 acquisition process. Furthermore, examining modality provides second language acquisition (SLA) researchers with another benchmark they can use to evaluate, compare and contextualize their findings. As such, the field of second language acquisition research will only benefit from and expand as a result of these types of studies.

References:

Bel, Aurora, Marta Ortells, and Gary Morgan. "Reference Control in the Narratives of Adult Sign Language Learners." *The International Journal of Bilingualism*, vol. 19, no. 5, 2015, pp. 608-624. ProQuest, <https://neiulibrary.idm.oclc.org/login?url=https://search-proquest-com.neiulibrary.idm.oclc.org/docview/1721570410?accountid=28190>.

- Frederiksen, Anne T., and Rachel I. Mayberry. "Reference Tracking in Early Stages of Different Modality L2 Acquisition: Limited Over-Explicitness in Novice ASL Signers' Referring Expressions." *Second Language Research*, vol. 35, no. 2, 2019, pp. 253-283. ProQuest, doi:<http://dx.doi.org/10.1177/0267658317750220>.
- Hilger, Allison I., et al. "Second Language Acquisition Across Modalities: Production Variability in Adult L2 Learners of American Sign Language." *Second Language Research*, vol. 31, no. 3, 2015, pp. 375-388. ProQuest, doi:<http://dx.doi.org/10.1177/0267658315570648>.
- Kemp, Mike. "Why is Learning American Sign Language a Challenge?" *American Annals of the Deaf*, vol. 143, no. 3, 1998, pp. 255-9. ProQuest, doi:<http://dx.doi.org/10.1353/aad.2012.0157>.
- Martinez, David, and Jenny L. Singleton. "Predicting Sign Learning in Hearing Adults: The Role of Perceptual-Motor (and Phonological?) Processes." *Applied Psycholinguistics*, vol. 39, no. 5, 2018, pp. 905-931. ProQuest, doi:<http://dx.doi.org/10.1017/S0142716418000048>.
- Perniss, Pamela. "Why We Should Study Multimodal Language." *Frontiers in Psychology*, 2018, Vol. 9, <https://www.frontiersin.org/article/10.3389/fpsyg.2018.01109>. Accessed 23 April 2019.
- Quinto-Pozos, David. "Teaching American Sign Language to Hearing Adult Learners." *Annual Review of Applied Linguistics*, vol. 31, 2011, pp. 137-158. ProQuest, doi:<http://dx.doi.org/10.1017/S0267190511000195>.
- Williams, Joshua T., Isabelle Darcy, and Sharlene D. Newman. "The Beneficial Role of L1 Spoken Language Skills on Initial L2 Sign Language Learning." *Studies in Second Language Acquisition*, vol. 39, no. 4, 2017, pp. 833-850. ProQuest, doi:<http://dx.doi.org/10.1017/S0272263116000322>.
- Williams, Joshua, Isabelle Darcy, and Sharlene Newman. "Second Language Working Memory Deficits and Plasticity in Hearing Bimodal Learners of Sign Language." *Psychology of Language and Communication*, vol. 19, no. 2, 2015, pp. 128-148. ProQuest, doi:<http://dx.doi.org/10.1515/plc-2015-0008>.
- WILLIAMS, JOSHUA, and SHARLENE NEWMAN. "Interlanguage Dynamics and Lexical Networks in Nonnative L2 Signers of ASL: Cross-Modal Rhyme Priming." *Bilingualism*, vol. 19, no. 3, 2016, pp. 453-470. ProQuest, doi:<http://dx.doi.org/10.1017/S136672891500019X>.

Dialogue Acts in a First Encounter Corpus

Costanza Navarretta¹, Patrizia Paggio^{1,2}, Bjørn Wessel-Tolvig¹

¹University of Copenhagen, ²University of Malta

[costanza@hum.ku.dk, paggio@hum.ku.dk, bwt@hum.ku.dk]

Keywords: dialogue act, multimodal corpora, annotation, first encounters, agreement

This abstract concerns the annotation and analysis of dialogue acts in a naturally occurring corpus of first encounters. Dialogue acts describe the functional actions that utterances have in conversations, and they are essential for understanding the dynamics and semantics of dialogues. Many dialogue acts classifications have been proposed, e.g. HCRC MapTask, Verbmobil, and DAMSL, and they have been combined in a standard, the ISO 24617 (Bunt et al. 2010, 2017).

We describe the annotation process and provide a first analysis of the dialogue acts identified in the corpus. Differing from many corpora annotated with dialogue acts, our corpus is not task-oriented, and therefore we had to add a few labels to cover specific aspects of the encounters.

The annotated corpus comprises 12 video-recorded Danish dyadic conversations between six male and six female participants who meet each other for the first time. The corpus is about one hour long and was collected and annotated at the University of Copenhagen in the Nordic project NOMCO (Paggio and Navarretta 2017). The annotations comprise the transcriptions of the conversations, annotations of the form and functions of head movements, facial expressions, and body postures, and their relation to speech. Recently, we have annotated dialogue acts.

The dialogue act specifications originate from the ISO 24617-8. First a coder annotated the dialogue acts in one conversation. The annotations were then checked by two other coders. Disagreements and problematic cases were discussed, and an agreed upon version was produced. Additional dialogue act labels were added if an appropriate label was not found in the ISO specifications. Based on the common understanding of the dialogue labels achieved through discussions of the first trial conversation, all the remaining dialogues were annotated. Currently, all the annotations are being checked, therefore the following analysis addresses the first annotation version. In the symposium presentation, a complete analysis referring to the checked data will be provided.

3285 dialogue acts were annotated falling under 31 dialogue act types. Figure 1 shows the most frequently used dialogue act labels, i.e. the labels that occurred at least 10 times in the corpus.

The most frequent dialogue act is INFORM, followed by ALLOFEEDBACKGIVE, INFORM-ANSWER, AGREEMENT, and CONFIRM. This reflects the conversation type, in which the participants introduce themselves and exchange a lot of information about themselves in a short time. The frequency figures also indicate that many common dialogue acts are feedback-related. Also this reflects the conversation type and the physical settings in which the participants stand in front of each other, are kind and show constant interest for what their interlocutor says.

At present, we are analyzing the dialogue acts in relation to the functions of the gestures which co-occur with them. This analysis will also be presented at the symposium.

References

Paggio, P. and Navarretta, C. (2017). The Danish NOMCO corpus: multimodal interaction in first acquaintance conversations. In *Language Resources and Evaluation*. Vol.51, Issue 2, pp. 463-494.

Bunt, H., Alexandersson, J., Carletta, J., Choe, J. W., Fang, A., Hasida, K., Lee, K., Petukhova, V., Popescu-Belis, A., Romary, L., Soria, C., & Traum, D. (2010). Towards an ISO standard for dialogue act annotation. In *Proceedings of the 7th international conference on language resources and evaluation (LREC 2010)*. Paris: ELRA. Google Scholar

Bunt, H., Petukhova, V., & Fang, A. (2017). Revisiting the ISO standard for dialogue act annotation. In *Proceedings of joint ISO-ACL workshop on interoperable semantic annotation (ISA-13)*, Montpellier, France (pp. 37–50).

Figures:

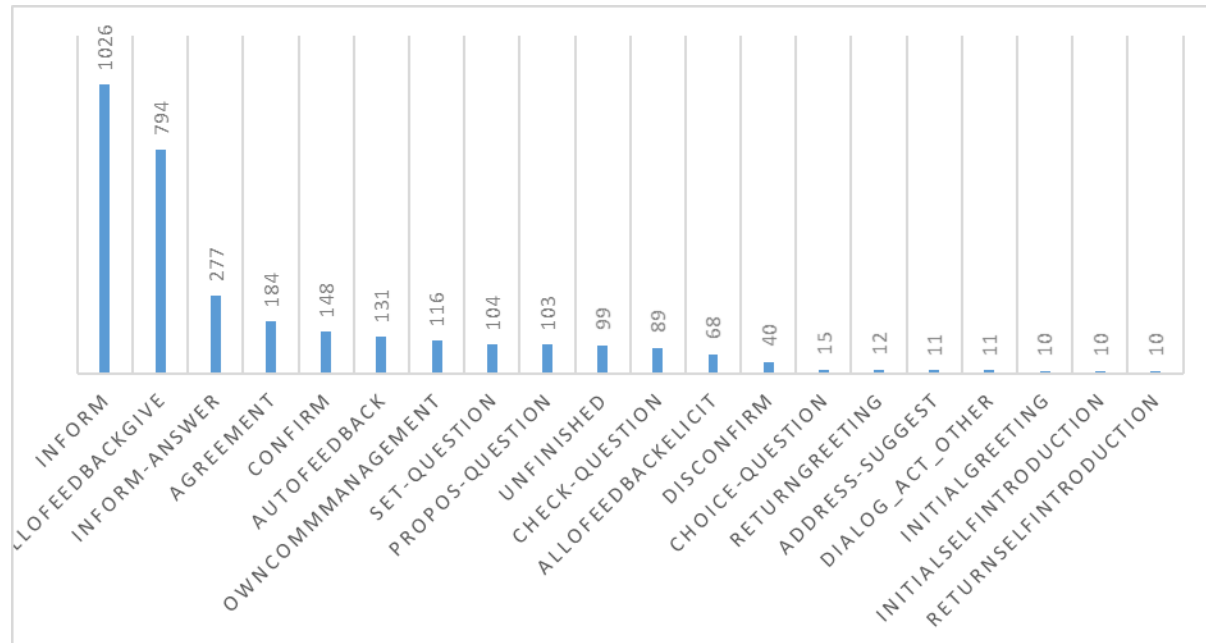


Figure 1: Most frequent dialogue acts

Gesture networks: Introducing dynamic time warping and network analysis for the kinematic study of gesture ensembles

Wim Pouw^{1,2} & James Dixon¹

¹*University of Connecticut, Center for the Ecological Study of Perception and Action*

²*Erasmus University Rotterdam, Department of Psychology, Education, and Child Studies*

[wimpouw@uconn.edu, james.dixon@uconn.edu]

Keywords: motion tracking, co-speech gestures, discourse processes, network analysis, time series analysis

In the current proposed talk we introduce applications of established methods in time series- and network analyses which we jointly apply for the kinematic study of *gesture ensembles*. We define a gesture ensemble as the set of gestures that are produced in a narrative by a single person or a group of persons. Here we are interested in how gestures kinematically relate to one another. We employ a bivariate time-series analyses called Dynamic Time Warping (DTW) to assess how similar gestures are to other gestures in the ensemble in terms of their velocity profiles (as well as studying multivariate cases with gesture velocity and speech amplitude envelope profiles). By relating each gesture event with all other gesture events produced in the ensemble, we obtain a weighted matrix that essentially contains a network of similarity relationships. We can therefore apply network analyses, which can gauge for example how diverse or coherent certain gestures are with respect to the gesture ensemble. Our exploratory analyses with two motion-tracking datasets show that iconic gestures tend to occupy less central positions in the gesture ensemble network. We also obtain that gestures change their position in the network as a function of the novelty of co-occurrent speech. Additionally, we introduce multivariate gesture-speech networks as well provide a method to study how gesture kinematics might change as a function of the language spoken. We think these analyses promise to be of great value for multimodal studies of language, as we can come to understand how low-level gesture features (kinematics of gesture) relate to the higher-order organizational structures present at the level of discourse.

Preprint available at: <https://psyarxiv.com/hbnt2>

Lexical and gestural alignment in collaborative referring

Marlou Rasenberg¹, Mark Dingemanse^{1,2} & Asli Özyürek^{1,2,3}

¹ *Centre for Language Studies, Radboud University, Nijmegen, The Netherlands*

² *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

³ *Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands*

[marlou.rasenberg@mpi.nl, m.dingemanse@let.ru.nl, asli.ozyurek@mpi.nl]

Keywords: co-speech gestures, collaborative referring, mutual understanding, mimicry, alignment

Conversations often seem fluent and effortless, though on closer inspection interlocutors are continuously resolving various communicative challenges. An example is talking about abstract ideas or unfamiliar objects which do not have specific labels. Repetition of lexical items plays an important role in this kind of collaborative referring (Brennan & Clark, 1996). Gestures too can be a powerful interactional resource, as they allow for iconic mappings and can convey meaning beyond what is expressed in speech. Like words, gestures can be repeated, and such repetition has also been shown to aid collaborative referring (Holler & Wilkin, 2011). A key open question is the relation between alignment in these two modalities, which remains understudied (Oben & Brône, 2016). Does alignment at one level co-occur with or even lead to alignment at the other, or do they operate more independently?

To investigate patterns of multimodal alignment, we used a director-matcher task, in which dyads communicate about unfamiliar 3D objects in 6 consecutive rounds. This yielded a corpus of roughly 7 hours of audio-video recordings (20 dyads, $M = 21$ min. per dyad), for which speech is transcribed and co-speech gestures annotated. Before and after this interactive task, participants individually labelled each object (in 1-3 words) for their partner. Using pretrained word2vec word embeddings (based on the NLPL Dutch CoNLL17 corpus), we quantified the semantic similarity of these descriptions, yielding a simple measure of relative convergence in naming practices.

We found that over the course of the interaction, participants establish dyad-specific conceptualizations of the objects. Analyses of the naming task data show convergence in naming practices: there was a significant increase in the semantic similarity of descriptions, when comparing pre-interaction naming ($M = 0.49$, $SD = 0.16$) with post-interaction naming ($M = 0.69$, $SD = 0.21$); $t(157) = -9.24$, $p < .001$ (one-tailed). In contrast, dyads of participants who did not interact with each other did not show similar post-interaction convergence.

Preliminary analyses of a subset of the interactional data reveal extensive use of iconic gestures ($N = 561$ for 8 dyads). We found that interlocutors often depict the same subparts of the objects; 66.5% of the gestures were preceded or followed by the partner gesturing about the same referent. Within this set of 'referentially aligned' gestures we investigate the degree of gestural alignment by scoring overlap in handedness, position, handshape, orientation and movement. Further analyses focus on the relationship between lexical and gestural alignment in the interaction, and how such behavioural alignment relates to semantic convergence in the naming task. By investigating alignment from a within- as well as cross-modal perspective, we shed new light on interlocutor's expressive resources to reach mutual understanding.

References:

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482–1493.

Holler, J., & Wilkin, K. (2011). Co-Speech Gesture Mimicry in the Process of Collaborative Referring During Face-to-Face Dialogue. *Journal of Nonverbal Behavior*, 35(2), 133-153.

Oben, B., & Brône, G. (2016). Explaining interactive alignment: A multimodal and multifactorial account. *Journal of Pragmatics*, 104, 32–51.

Multimodal Contingency in Developing Second Language Learning and Literacy

Alessandro Rosborough & Jennifer Wimmer

Brigham Young University

[alex_rosborough@byu.edu, Jennifer_Wimmer@byu.edu]

Keywords: multimodal contingencies; meaning-making; gesture

This study describes the interrelationship of multiple modalities and contingent interactions during literacy events in second and third grade classrooms for second language learners. The question of how teachers corporeally-materialize, develop, and sustain joint-attention, shared-intentionality, and intersubjective levels of meaning-making with their students was analyzed, and is a needed topic in second language educational-based research (Early, Kendrick, & Potts, 2015). We address how co-coordinated attention, intentionality, and intersubjectivity created meaning for second language learners (i.e., in contrast to morpho-lexico-syntactic language systems). Such co-construction is supported by Vygotsky's (1978, 1987) cultural-historical theory, where meaning-making is created on a shared social plane, more readily understood through the dialectical perspective of how the collective and individual experience are always united. Such dialectics include communication and discourse as an experience always performed through whole-body sense-making (Thibault, 2011). Using dialectical unities and a cultural-historical lens, we describe how participants created extended meanings through acts and enactments that were afforded space through contingent engagement. We define *interactional contingency* as the "unpredictability of sequence and outcome," shared between communicating participants, leading to engagement and learning (van Lier, 1996, p. 169).

Data is based on multiple classrooms of second and third grade children learning English and Spanish in English as a Second Language Classrooms, and Dual Language Immersion classrooms. Three cameras allowed for coverage of the confluence of multimodal texts in communication, included co-speech gesture, gesture itself, prosodics, facial expressions, the teacher's act of drawing, drawings themselves, and handling of classroom objects. Coding and transcripts are based on McNeill (1992) with regard to types of gesture, but discourse analysis includes descriptions of the interface between speech-bodily-object use in communication as "ensembles" of meaning. Specifically, "ensembles" identified and described in the data include how *coacting agents* created and engaged with multimodal texts as distributed meaning-making acts.

Analysis revealed how participants and their multimodal ensembles (merged timescales), created new meanings and sense-making in the targeted second language. Some results demonstrated how the multiple modalities consolidated to provide a strong foundational meaning from which further discourse could emerge. However, some activities and discourse limited the students' abilities to participate and engage in the task.

References:

- Early, M., Kendrick, M., & Potts, D. (2015). Multimodality: Out from the margins of English language teaching. *Tesol Quarterly*, 43(3), 447-460.
- McNeill, D. (1992). *Hand and mind*. Chicago, US: Chicago University Press.

Thibault, P. (2011). First-order languaging dynamics and second-order language: The distributed language view. *Ecological Psychology*, 23(3), 210-245.

van Lier, L. (1996). *Interaction in the language curriculum: Awareness, autonomy, & authenticity*. New York: Longman Publishing.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Vygotsky, L. (1986). *Thought and language*. Cambridge, MA: MIT press.

Where should I begin? Calibrating the design of first actions in conversation

Federico Rossano

University of California, San Diego

[frossano@ucsd.edu]

Keywords: action formation, turn design, accountability, longitudinal change, primates

In deciding what social actions are being produced during a conversation, conversation analysts tend to adopt an *emic* perspective (a participants' perspective, see Pike, 1967), and, thus, have developed a procedure that has been called the "*next-turn proof procedure*" (Sacks et al. 1974). The claim is that the interactional nature of conversation provides an obligation among participants in an interaction to display to each other their understanding of the previous conversational turn and if no correction occurs, then the assumption should be that B has correctly understood A, and therefore that A's turn was aimed at eliciting the kind of response that B produced. This procedure has been labeled the "central methodological resource for the investigation of conversation" (Sacks et al. 1974: 728).

Erving Goffman famously rejected the idea that this proof procedure would be sufficient to account for the interpretation of social action in social interaction by noting that: "an account of second utterances in terms of their contingency on a first leaves unexplained how there could be any firsts; after all, from where could they draw their design? Conversation could never begin" (1983: 50). Yet in his criticism he stopped short of any empirical investigation on the matter of calibration of an individual's first action and the origin of those designs.

Moreover, apart from recent efforts in institutional settings (Gonzales, Pekarek-Doheler & Wagner, 2018) most research on ordinary interactions in ordinary settings has neglected investigating how communications between individuals might change through time because of the development of shared knowledge and the establishment of clear epistemic boundaries via repeated interactions with each other. Most importantly, by focusing on the "next-turn proof procedure", we have mostly lost track of what leads people to select specific first turn designs (though see e.g. Curl & Drew, 2008 for an example of research on the design of first pair parts). How do we calibrate our first actions given what we know about others, the social situation we are in, our communicative abilities and given our cognitive abilities?

This paper utilizes both observational and experimental data to present preliminary findings on how human and non-human primates calibrate requests for actions and for objects, to what degree communicative practices change through development and what affects such change.

When pointing becomes more than pointing: multimodal evaluation in product pitches

Julia Valeiras-Jurado¹ & Noelia Ruiz-Madrid²

¹*Ghent University*, ²*Jaume I University*

[julia.valeirasjurado@ugent.be, madrid@uji.es]

Keywords: mutimodality, evaluation, persuasion, product pitches, deictic gestures

This paper presents a study of the use of evaluation as a persuasive strategy in product pitches. Product pitches are oral presentations characterized by a strong persuasive component in their communicative goal, since their main function is to convince an audience of the validity of a product or service, often by presenting it as the best solution to a problem (Allison et al. 2017; Daly and Davy 2016; Davies et al. 2017; Rackham 1998; Woods et al. 2014). In order to build a persuasive pitch speakers resort to different strategies, among which we can find evaluation. From a multimodal perspective, evaluation is interesting because it is frequently realised through an orchestration of different semiotic modes (Kress 2003; Valeiras-Jurado & Ruiz-Madrid 2015; Valeiras-Jurado in press).

This study combines Multimodal Discourse Analysis (MDA) and ethnographic methods to elucidate how speakers in product pitches use evaluation as a multimodal persuasive strategy. In particular, the focus lies on words, intonation (Brazil 1997) and gestures (Kendon 2004) as modes and on how they are used in combination to convey a persuasive evaluation of a product.

Results show that some deictic gestures can acquire an evaluative function when used within a skilfully orchestrated multimodal ensemble. This multimodal evaluation can in turn become persuasive if speakers subtly prompt their audience to share it.

A better understanding of how evaluation can be conveyed persuasively to an audience can greatly improve our knowledge of product pitches as an oral, multimodal genre. In addition, it can help practitioners improve their performance as they use this genre in their everyday tasks.

References:

- Allison, T. H., B. C. Davis, J.W. Webb, J. C. Short. 2017. "Persuasion in crowdfunding: an elaboration likelihood model of crowdfunding performance." *Journal of Business Venturing*, 32 (6): 707–725.
- Brazil, D. 1997. *The Communicative Value of Intonation in English*. 2nd ed. Cambridge: Cambridge University Press.
- Daly, P. and D. Davy. 2016. "Structural, linguistic and rhetorical features of the entrepreneurial pitch: Lessons from Dragons' Den." *The Journal of Management Development* 35 (1): 120-132.
- Davies, B., M. Hmieleski, J.W. Webb and J. E. Coombs. 2017. "Funders' positive affective reactions to entrepreneurs' crowdfunding pitches: the influence of perceived product creativity and entrepreneurial passion." *Journal of Business Venturing* 32(1):90–106.
- Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Kress, G. 2003. *Literacy in the New Media Age*. London: Routledge.
- Rackham, N. 1998. "From Experience: Why Bad Things Happen to Good New Products." *Journal of Product Innovation Management* 15: 201–207.

Valeiras-Jurado, J. & Ruiz-Madrid, N. 2015. "A multimodal approach to product presentations". *Procedia - Social and Behavioral Sciences* 173: 252-258.

Valeiras-Jurado, J. In press. "Modal coherence in specialised discourse: a case study of persuasive oral presentations in business and academia". *Iberica*

Woods, R., K. Rafferty, J. Murphy, and P. Hermon, eds. 2014. *Engineering Innovative Products*. John Wiley.

The PAT Annotation Model for Multimodal Instructions

Ielka van der Sluis & Gisela Redeker

Center for Language and Cognition Groningen

[i.f.van.der.sluis@rug.nl, g.redeker@rug.nl]

Keywords: annotation, model, application, corpus, instructions

We present the development, implementation and application of the PAT annotation model, which describes instructional documents that consist of pictures and text (PAT). In document design research (Schrifer, 1997) the combination of pictures and text has been noted but not so much investigated in terms of function and content. Although useful starting points have been provided (Bateman, 2014), we are unaware of a standard methodology to describe and evaluate picture-text relations. This leaves document designers without specific guidelines, while readers and users may experience difficulties in effectively processing multimodal content due to mismatches with their expectations and cognitive capacities.

Development: As the possibilities to describe multimodal documents are infinite, we advocate conducting corpus studies and reader/user studies in tandem to determine the relevance of annotation categories and their values. Based on preliminary analyses of a corpus with first-aid instructions (currently 297), we conducted multiple reader and user studies to investigate the effectiveness of particular design features. With the results of these studies we developed and fine-tuned the specification of 51 types of functional and content relations between textual elements and between text and pictures.

Implementation: Our corpus annotation is supported by the PAT Workbench, a custom-made online tool that provides a flexible environment to systematically investigate multimodal designs by facilitating storage, annotation, retrieval and evaluation of documents (See, Figures 1, 2 and 3 and <https://cosmo.service.rug.nl/patworkbench/login/>). The workbench includes 'smart' OCR for uploaded documents, user-defined specification of annotation categories, and a tool for creating gold standard annotations based on multiple annotations.

Application: As a worked example, we will present the results of a comparative study that involved the application of the PAT annotation model to a subcorpus of 46 first-aid instructions from two editions of *Het Oranje Kruis Boekje* 2011 and 2016. *Het Oranje Kruis* is a Dutch organisation that provides learning materials for first-aid certification trainings. A comparison of multimodal instructions (117 pictures and 9416 words in total) for 23 tasks in both editions of *Het Oranje Kruis Boekje* allows us to conclude that the two editions are similar in terms of the visualised actions, but differ in terms of: text content (preambles, alternative actions, control information); the type of shot used in the pictures (close-up/medium shot versus long shot); and text-picture relations in terms of layout (alignment versus proximity).

Future work: The PAT project (<http://www.rug.nl/let/pat>) will deliver theoretical results in terms of empirically validated models for effective multimodal presentations and authoring guidelines for multimodal documents. Future work will include more comprehensive textual analysis and finer-grained analysis of the pictorial materials, coverage of a greater number and a wider variety of instructions, (semi-)automatic annotation, more empirical evaluation, and (semi-)automatic generation of potentially effective text-picture combinations for multimodal instructions.

References:

Bateman, J. (2014). Using multimodal corpora for empirical research. In *The Routledge handbook of multimodal analysis*, pp. 238–252. Routledge, London.

Schriver, K. (1997). *Dynamics in document design: Creating text for readers*. Wiley, New York.

Figures:

The screenshot shows the 'Paragraph Types' configuration page in the PAT workbench. The page has a red header with 'pat workbench' on the left and 'Dashboard' and 'First Aid Instructions' in the center. Below the header, there are navigation tabs: 'General', 'Documents', 'Annotation definition', 'Annotation manual', 'Paragraph types' (selected), and 'Import / Export'. The main content area is titled 'Paragraph Types' and contains a table with columns for 'Name', 'CSS', 'Shortcut', and 'Delete'. The table lists five paragraph types: Title, Heading, Paragraph, Picture, and Caption. Each row includes a text input field for the name, a text input field for the CSS style, a dropdown menu for the shortcut, and a checkbox for the delete option. Below each row is an 'Explanation' text area.

Name	CSS	Shortcut	Delete
Title	text-transform: none; font-size: 12pt; color: #000; font-weight: bolder;	Y	<input type="checkbox"/>
Explanation	The Title of a document is usually placed at the top of the document and presents information about the content of the document. The font size of a Title is usually larger than the main text. The font colour may also differ from...		
Heading	text-transform: none; font-size: 12pt; color: #000; font-weight: bolder;	H	<input type="checkbox"/>
Explanation	Headings precede one or more paragraphs. Headings contain information about the content or place of a paragraph within the procedure described in the document.		
Paragraph	text-transform: none; font-size: 12pt; color: #000;	P	<input type="checkbox"/>
Explanation	A Paragraph consists of one or more sentences and is separated from the previous paragraph or heading with a hard return or white line.		
Picture	text-transform: none; font-size: 12pt; color: #000;	F	<input type="checkbox"/>
Explanation	Photograph, drawing or other graphical presentation.		
Caption	text-transform: none; font-size: 12pt; color: #000; font-style: italic;	C	<input type="checkbox"/>

Figure 1: Screenshot of description of Paragraph Types in the PAT workbench.

The screenshot shows the 'Documents' page in the PAT workbench. The page has a red header with 'pat workbench' on the left and 'Dashboard' and 'First Aid Instructions' in the center. Below the header, there are navigation tabs: 'General', 'Documents' (selected), 'Annotation definition', 'Annotation manual', 'Paragraph types', and 'Import / Export'. The main content area is titled 'Documents' and features a large dashed box with the text 'Click to upload your files'. Below this, there is a section for '(PDF) Documents for annotation within this project'. This section includes search filters: 'Search in document content', 'Search in title and filename', and 'Search in annotations'. There are also buttons for 'Search', 'Clear', 'Export results', 'Lock all 297', and 'Unlock 297'. A list of documents is displayed below, with columns for document name, size, and various icons (lock, share, report, etc.).

Document Name	Size	Icons
MI983_OK2016_Deel3.8_Actieve_bloedingen_Behandelen_van_actief_bloedverlies.pdf	824.5 KB	lock, share, report, pdf
MI984_OK2016_Deel3.8_Levensbedreigende_letsels_en_ziekten_Behandelen_van_een_hitteberoerte.pdf	746.9 KB	lock, share, report, pdf
MI985_OK2016_Deel3.8_Levensbedreigende_letsels_en_ziekten_Behandelen_van_een_vergiftiging.pdf	802.0 KB	lock, share, report, pdf
MI986_OK2016_Deel3.8_Levensbedreigende_letsels_en_ziekten_Vaststellen_van_een_hersenvliesontsteking.pdf	725.7 KB	lock, share, report, pdf
MI987_OK2016_Deel3.8_Overige_letsels_en_ziekten_Behandelen_van_een_lichte_onderkoeling.pdf	795.5 KB	lock, share, report, pdf
MI988_OK2016_Deel3.8_Wonden_Behandelen_van_een_brandwond.pdf	747.8 KB	lock, share, report, pdf
MI989_OK2016_Deel3.8_Wonden_Behandelen_van_een_bevriezingswond.pdf	820.2 KB	lock, share, report, pdf
MI990_OK2016_Deel3.8_Botbreuken_en_ontwrichtingen_Behandelen_van_kneuzing_en_verstuiking.pdf	791.7 KB	lock, share, report, pdf
MI991_OK2016_Deel3.8_Botbreuken_en_ontwrichtingen_Behandelen_van_botbreuken.pdf	759.2 KB	lock, share, report, pdf
MI992_OK2016_Deel3.8_Neus_en_oortelsels_Dichtdrukken_van_de_neus.pdf	771.4 KB	lock, share, report, pdf

Figure 2: Screenshot of corpus documents in the PAT workbench.

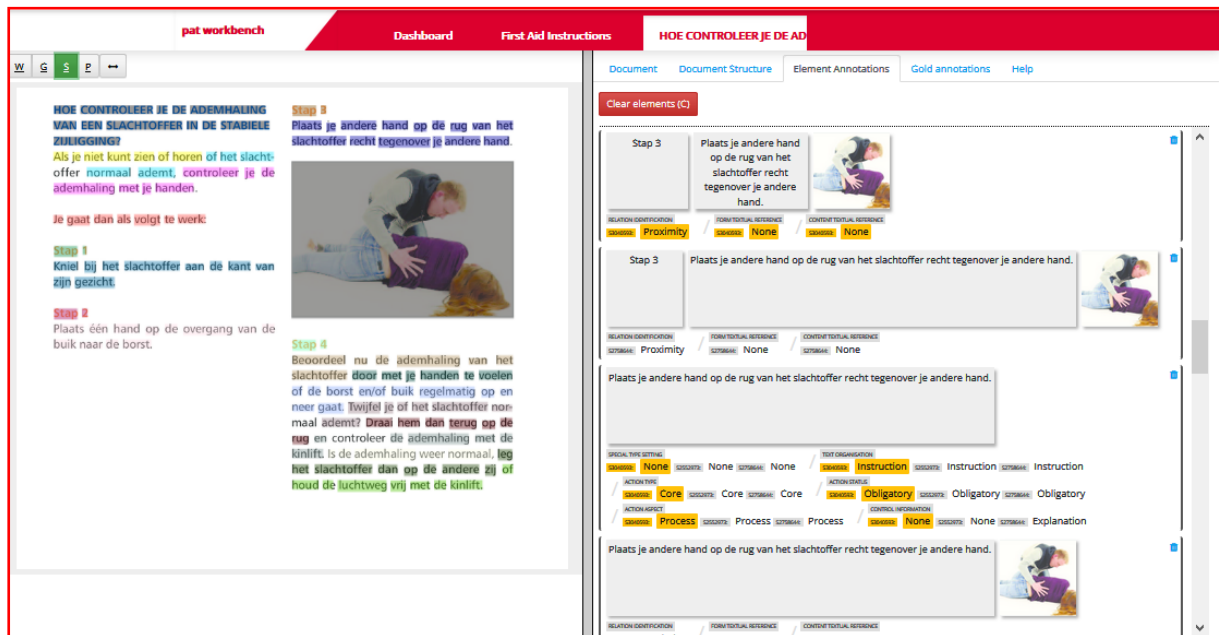


Figure 3: Screenshot of document annotation in the PAT workbench.

Are emotions hemispheric specialized?

Sabrina von Au, Katharina Hümmer, Marie Garling, Hedda Lausberg, Ingo Helmich

Department of Neurology, Psychosomatic Medicine and Psychiatry, Institute of Health Promotion and Clinical Movement Science, German Sport University (GSU) Cologne

[s.vonau@dshs-koeln.de]

Keywords: emotions, tachostoscopy, NEUROGES®-Elan, hemispheric specialization, hand movements

Objectives: Two theories exist for the hemispheric lateralization of emotions H1: Differential hemispheric specialization for positive and negative emotions only in the right hemisphere; H2: Differential hemispheric specialization for positive emotions in the left hemisphere and negative emotions in the right hemisphere for right-handers and vice versa for left-handers (Hartje, 2001; Casasanto, & Jasmin, 2010; Reuter-Lorenz et al., 1983; Ley & Bryden, 1979). Previous research revealed that hand movements and gestures are differently expressed with regard to emotional processes (Lausberg et al, 2010). The hemispheric specialization for certain movement types are associated with certain emotional processes (Lausberg, 2013), thus we hypothesize that emotional hand movement behavior is associated to hemispherically specialized functions.

Design: Ten healthy right-handed individuals were examined (M=23.60±3.03 years; male: 2, female: 8). The German version of the Positive and Negative Affect Schedule (PANAS) was used to record the current emotional state. Ten images of emotional sports scenes were presented tachostoscopically to each subject (five negative / five positive stimuli per right and left field of view). Individuals were asked to indicate the emotion expressed by the sport scene and described the images to a third person. The NEUROGES Coding System was used for the analysis of the hand movements. The frequency of hand movements and the structure of hand movements were examined.

Results: There was a significant difference between the hand movement Structure ($F(4,36)=23,219$ $p<0,001$). When examined the hypotheses, no significant influence of the two factors visual field of vision and emotions could be demonstrated. A minimal but not significant higher frequency of the left hand movement structure phasic could be detected.

Discussion: Minimal higher frequency of left hand structure phasic for all positive and negative images indicate H1, i.e., emotions is right hemispherically processed. These findings are in line with the fact that left hand movements are controlled by the right hemisphere and that information generated in the left hemisphere are transferred via corpus callosum to the right hemisphere. The generation of left hand movements is associated with emotional processes that are lateralized in the right hemisphere.

Conclusion: The data could maybe support the hypothesis of differential hemispheric specialization for positive and negative emotion only in the right hemisphere. Besides there could be an influence on emotions through increased hand movement structure phasic or increased left hand movements.

References:

Hartje, W. (2001). Emotion und Hemisphärendominanz. *Zeitschrift für Neuropsychologie*, 12(4), 276-290.

Casasanto, D., & Jasmin, K. (2010). Good and bad in the hands of politicians: Spontaneous gestures during positive and negative speech. *Plos one*, 5(7), e11805.

Lausberg, H. (2013). *Understanding Body Movement: A Guide to Empirical Research on Nonverbal Behaviour-With an Introduction to the NEUROGES Coding System*. Peter Lang International Academic Publishers.

Ley, R. G., & Bryden, M. P. (1979). Hemispheric differences in processing emotions and faces. *Brain and language*, 7(1), 127-138.

Reuter-Lorenz, P. A., Givis, R. P., & Moscovitch, M. (1983). Hemispheric specialization and the perception of emotion: Evidence from right-handers and from inverted and non-inverted left-handers. *Neuropsychologia*, 21(6), 687-692.

Sassenberg, U., Helmich, I. & Lausberg, H. (2010). Awareness of emotions: Movement behaviour as indicator of implicit emotional processes in participants with and without alexityhmia. *Proceedings of KogWis: 10th Biannual Meeting of the German Society for Cognitive Science*, Potsdam, Germany; Haack, Wiese, Abraham, Chiarcos (eds.), p. 169.