INTERNATIONAL WORKSHOP
"MULTIMODAL LANGUAGE. THEORETICAL PERSPECTIVES AND RESEARCH METHODS"

DEPARTMENT OF HUMANITIES, UNIVERSITY OF CATANIA

13-14 MARCH 2025

BOOK OF ABSTRACTS

Keynote

A Multimodal Language Framework (MLF) best situates language in cognition and communication: Insights from Italian and other spoken and signed languages

Aslı Özyürek

Multimodal Language Department, Max Planck Institute for Psycholinguistics
Donders Institute for Brain Cognition and Behavior, Radboud University

asli.ozyurek@mpi.nl

Since the publication of seminal works of Kendon and McNeill it is now an accepted view that language is inherently multimodal and that many properties of visual expressions in spoken and sign languages constitute fundamental design features of language. These include modality specific features such as visual iconicity, indexicality and simultaneity expressed by the manual, bodily and facial expressions. The next challenge in linguistics is then how to define a framework that explains how these can be integrated along with arbitrary, categorical and sequential properties of linguistic expressions typical of spoken/textual expressions and in a way that grounds language in human cognition and communication.

To do so I will sketch a M*ultimodal Language Framework (MLF)* (e.g., Holler and Levinson 2019; Hagoort and Özyürek, 2024) that attempts to offer a way for how characterize language in such a multidimensional way. Based on empirical data from spoken and sign languages, including Italian and Italian Sign language, I will outline how producing and perceiving linguistic expressions with this added complexity of multimodal expressions allows us to revisit typologies of language structures, language-(neuro)cognition interface, production and comprehension processes, models of communication (based on efficiency, relevance, conversation) and language acquisition. The take home message from more than a decade of such research is that a multimodal but not a unimodal view of language structure and use best characterizes human language as an adaptive system to human cognition, learning and communication principles. I will end the talk with a call for more multimodal language research in Italian which is rooted typologically in rich multimodal use of language.

Keynote

Aspects of multimodal language development

Maria Graziano

Lund University Humanities Lab

maria.graziano@humlab.lu.se

Given the substantial evidence supporting the close link between speech and gesture in human communication, research on language development has increasingly embraced a multimodal approach. While earlier studies primarily concentrated on the initial stages of language acquisition, more recently research has shifted to examining later stages, aiming to understand how children learn to integrate linguistic structures with bodily resources in extended complex discourse, such as storytelling.

I will illustrate how children's ability to integrate the expressive potentials of spoken and gestural modalities in narrative discourse changes over time and increases with the development of linguistic and pragmatic skills. I will argue that, in later stages of language acquisition, the development of speech and gesture proceed in parallel, and I will maintain that this supports the view that two modalities form an integrated system. Finally, I will also highlight the need for further research on children's multimodal behaviour, particularly the importance of extending the studied age range beyond childhood, as well as the sort of discourse. Additionally, I will suggest that combining a qualitative and a quantitative perspective could offer valuable insights into how the two modalities develop and interact in children's spoken discourse.

TALKS

Multimodal adaptations in storytelling: Differences in gesture and speech between child-Directed and adult-directed narratives

Jiali Li[1], Ingrid Vilà-Giménez[2], Sara Coego[1], Pilar Prieto[3,1]
Universitat Pompeu Fabra[1], Universitat de Girona[2], Institució Catalana de Recerca i Estudis Avançats[3]

Jiali.li@upf.edu

There is evidence that speakers use multimodal linguistic adaptation strategies when communicating with children (child-directed speech, CDS) and adults (adult-directed speech, ADS). Previous studies have generally found that CDS involves more referential iconic gestures and longer speech production than ADS (Campisi & Özyürek, 2013; Kandemir et al., 2024; Zhang & Gu, 2024). While the increased use of referential iconic gestures in CDS has been highlighted, the role of gesture viewpoint or informativeness has been neglected. The present study investigates how future preschool teachers adapt their multimodal language when telling stories to children compared to adults, focusing on differences in narrative length, gesture use, and the semantic-pragmatic characteristics of gestures between CDS and ADS.

The study employed a within-subjects design based on the TEACH-TALK audiovisual narrative corpus (Vilà-Giménez et al., 2024), which consists of 80 narratives performed by 40 first-year Early Childhood Education students. Each participant narrated the same Pingu wordless cartoon to two simulated audiences: one adult and one child (1 narrative x 40 participants x 2 conditions). All the narratives were analyzed for word count, and all co-speech gestures were coded across multiple levels: gesture referentiality (referential vs. non-referential gesture) and dimensions of referentiality (referential iconic vs. deictic vs. metaphoric gesture), following the M3D labeling system (Rohrer et al., 2023); gesture viewpoint (character vs. observer vs. dual vs. narrator viewpoint gesture; following McNeill, 1992); and gesture informativeness in the case of referential iconics (redundant vs. non-redundant referential iconic gesture). From a methodological standpoint, this study introduced a novel approach by analyzing gesture viewpoint across all gesture types, including non-referential gestures.

Statistical analyses involved six linear mixed-effects models examining differences in narrative length (in terms of word count) and gesture use (i.e., overall gesture rate, gesture rate by referentiality, dimensions of referentiality, viewpoint, and informativeness) between CDS and ADS conditions. Results revealed significant differences in multimodal storytelling strategies across conditions. Narratives directed at children were significantly longer than those for adults, suggesting more elaborative and explanatory language use. While no significant differences were found in overall gesture rate or gesture informativeness, speakers produced significantly higher rates of referential iconic gestures in CDS. Importantly, the use of character viewpoint (CVPT) gestures increased significantly in CDS narratives. Although infrequent, non-referential CVPT gestures were found exclusively in CDS, potentially conveying abstract qualities of the characters' behavior beyond explicit dialogue.

The increased use of referential iconic gestures and CVPT gestures when narrating in CDS indicates a tendency to embody characters' perspectives more frequently when addressing children, suggesting that these gestures may serve to enhance story comprehension and engagement for younger audiences. These findings contribute to our understanding of how speakers adapt their multimodal communication strategies based on audience characteristics (CDS or ADS) and offer practical insights to enhance storytelling techniques through effective multimodal communication strategies for storytelling in educational contexts.

References

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics*, *47*(1), 14-27.

Kandemir, S., Özer, D., & Aktan-Erciyes, A. (2024). Multimodal language in child-directed versus adult-directed speech. *Quarterly Journal of Experimental Psychology*, *77*(4), 716-728.

McNeill, D. (1992). Hand and mind1. *Advances in Visual Semiotics*, *351*.

Rohrer, P. L., Tütüncübasi, U., Vilà-Giménez, I., Florit-Pons, J., Esteve-Gibert, N., Ren-Mitchell, A., Shattuck-Hufnagel, S. & Prieto, P. (2023). The MultiModal MultiDimensional (M3D) labeling system. https://doi.org/10.17605/osf.io/ankdx

Vilà-Giménez, I., Coego, S., Geladó, S., Fraser, A., & Prieto, P. (2024). TEACH-TALK: An audiovisual corpus of future teachers' child- and adult-directed narratives. https://osf.io/ae3xr/

Zhang, Y., & Gu, Y. (2024). Individual differences in multimodal child-directed language: Unraveling individual style, empathy and the Big Five personality traits. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 46).

From discourse to dialogue: exploring dialogic gestures in French interaction

Loulou Kosmala[1], Corrado Bellifemine[2]

[1]Paris Est-Créteil University
[2]University of Lorraine

loulou.kosmala@u-pec.fr

For a long time over the past fifty years, a distinction has often been made between 'referential' and 'pragmatic' gestures (Kendon, 2004, 2017; Lopez-Ozieblo, 2020) to refer to the two main functions of manual gestures in interaction: some are used to represent objects or spatial relations (referential), while others display a stance, embody a speech act, or indicate the relationship between different discourse segments (pragmatic). The latter are also known as 'speech-handling' gestures (Streeck, 2009) and share many recurrent (Muller, 2017) formal features such as the palm-up open hand, the finger bunch, the cyclic gesture, or the shrug, among others. Fewer studies have been dedicated to *interactive* gestures, which can be viewed as a subtype of pragmatic gestures (Kendon, 2017). One notable exception is the seminal work of Bavelas et al. (1992, 1995) who presented a typology of gestures 'specialized for dialogue', including common-ground gestures, gestures used to cite the interlocutor's previous contribution, or turn-taking gestures. Additionally, while all pragmatic gestures can be considered as interactive, some referential gestures (i.e. representational and deictic) can also serve interactive functions, especially in contexts of co-construction and alignment. From an interactionist and dialogical approach (Bakhtine 1978, 1984; Couper-Kuhlen & Selting, 2018), any gesture can be considered interactive to a certain extent, in the sense that they are 'addressed to the interlocutor' during situated interaction. For the present study, we adopt the term 'dialogic gestures' to refer to those gestures which embody 'dialogic moves' (Hudelot & Salazar-Orvig, 2003) and are thus primarily oriented to the dialogue rather than discourse.

This study aims at examining the dialogic dimension of referential and pragmatic gestures in six 20-minute recordings of semi-spontaneous French conversations between university students. Building and expanding on Bavelas and colleagues' initial typology, we first distinguish between gestures that are oriented to the dialogue and the interlocutor ("dialogic") and those that are more discourse- and self-oriented ("discursive") before labeling them 'pragmatic' or 'referential'. Gesture form was also annotated (pointing, palm up, cyclic gesture, finger snap, tracing, molding, acting…) as well as other bodily articulators such as eyebrows, torso and head movements, which contribute both to interactional dynamics and discourse construction.

Results show that speakers mainly rely on pragmatic gestures during face-to-face interaction, confirming their interactive nature in dialogue but also their structuring role during discourse elaboration. Dialogic gestures, which are mostly pragmatic, are highly influenced by the level of familiarity and engagement found within the pairs; while some pairs are highly engaged and rely a lot on common ground, others are more withdrawn and thus rely less on dialogic gestures. In particular, (dis)alignment, common ground, and citing are among the most frequent subtypes of dialogic gestures.

More detailed qualitative analyses show that some of the functions put forth by Bavelas and colleagues (e.g. "delivery" gestures) can also be applied to discourse-oriented gestures, thus highlighting the intricate relation between discourse and dialogue in interaction. At the formal level, Finger pointing and palm ups, depending on the context, can both be used to point towards or introduce/maintain a referent (referential) or establish common ground and signaling agreement or disapproval (pragmatic). The orientation of the hands, in particular, can further help distinguish between a discourse-oriented gesture and a dialogic one. This further questions the complex relationship between form, shape, orientation, and function.

Although pragmatic gestures can always be considered interactive (according to Bavelas and colleagues), this study shows that some sequential and contextual features (turn sequencing, turn positioning, and other dialogic moves) can further influence the way speakers make use of different types of gestures to anchor and ground themselves in conversation and participate in discourse co-construction.

References

Bakhtine M. (1934[1975/1978]). Du discours romanesque, dans *Esthétique et théorie du roman*, Paris : Gallimard, p. 83-233,.

Bakhtine M. (1952[1979/1984]). Les genres du discours, dans *Esthétique de la création verbale*, Paris : Gallimard, p. 265-308,.

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes, 15*(4), 469-489.

Bavelas, J. B., Chovil, N., Coates L., & Roe L. (1995). Gestures specialized for dialogue. *Personality and Social Psychology Bulletin, 21*(4), 394–405.

Couper-Kuhlen, E., & Selting, M. (2018). *Interactional Linguistics: Studying Language in Social Interaction*. Cambridge: Cambridge University Press.

Hudelot, C., & Salazar Orvig, A. (2003). Influence du mode de garde sur le développement du langage chez l'enfant: quelques repères sur le développement dialogique. *La langue et les parlants*, 49-63.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge University Press.

Kendon, A. (2017) Pragmatic functions of gestures: Some observations on the history of their study and their nature. *Gesture*, *16*(2), 157-175.

Lopez-Ozieblo, R. (2020). Proposing a revised functional classification of pragmatic gestures. Lingua, 247, 102870.

Müller, C. (2017). How recurrent gestures mean: Conventionalized contexts-of-use and embodied motivation. *Gesture*, *16*(2), 277-304.

Streeck, J. (2009). Depicting gestures: Examples of the analysis of embodied communication in the arts of the West. *Gesture*, *9*(1), 1-34.

Beyond honest:
The pragmatic dimensions of the Hand(s) on Heart Gesture in Hebrew face-to-face interaction

Anna Inbar[1, 2], Yuval Geva[2]
[1]Levinsky-Wingate Academic College; University of Haifa

inbara9@gmail.com

Using Hebrew data, the present study examines the pragmatic functions of the Hand(s) on Heart Gesture which is performed by placing either one hand or two hands on the upper torso, with the two-handed variant including the crossing of one hand over another. In Hebrew culture, the gesture is strongly associated with the concepts of *honesty* and *sincerity*, as evident, for example, in the Hebrew idiom *'im yad 'al ha-lev* (lit. 'with a hand on the heart') meaning 'honestly', in which the chest metonymically represents a heart (cf., Baranyiné Kóczy and Sipőcz 2023). However, we will show that the analysis of the gesture as it is used in naturally occurring conversation, reveals other meanings.

The Hand(s) on Heart Gesture has sparked the interest of researchers in various fields, including art history (Lange 1887), embodied cognition (e.g., Cantarero et al. 2017; Parzuchowski et al. 2014), and nonverbal communication (Farley et al. 2021). In the field of multimodal interaction analysis, however, the gesture was mentioned only in passing (e.g., Quasinowski et al. 2022; Streeck 2009). The present study seeks to fill this lacuna, using data drawn from the *Haifa Multimodal Corpus of Spoken Hebrew*, which consists of video recordings of naturally occurring casual conversations collected from 2016 through 2023, with approximately 18 hours. We identified 91 occurrences of the gesture in total. However, in 49 cases participants used the gesture referentially, usually for a metonymic representation of the first person or heart, or as a means of indexing the chest. The current study will focus on the 42 remaining occurrences of the gesture. To explore what this gestural resource accomplishes in interaction in the moment in which it is produced, we employed the methodologies of interactional linguistics (Couper-Kuhlen and Selting 2018) and multimodal conversation analysis (e.g., Goodwin and Goodwin 2000; Mondada, 2016).

The analysis of the collection revealed the following contexts in which the gesture was used: (1) conveying an affective stance (18%), (2) conveying content assumed to violate expectations, such as concessive statements (20%); in these contexts, the gesture was often co-produced with verbal expressions of sincerity, such as *be-'emet* (lit. 'in true'), (3) intensification (41%), (4) "self-defense sequences" (13%), in which the speaker rejects the criticism and produces a "self-defense" statement claiming that the interlocutors misinterpreted his or her intentions, and (5) other contexts in which the gesture accompanied uncontrollable physiological responses involving physical exertion, such as laughing, coughing, and also imitation of vomiting (8%). Based on these findings, on the principles of grammaticalization (e.g., Hopper and Traugott 1993), and on the principles of the 'emancipation' of gestures as semiotic resources used in interaction (Müller 2014; Streeck 2009, 2021), we will suggest and elaborate on the grammaticalization path of the gesture, tracing its evolution from a physiological response to excitement-induced palpitations to its role as a pragmatic marker of intensification.

References

Baranyiné Kóczy, J. and Sipőcz, K. (eds.). (2023), *Embodiment in cross-linguistic studies: The 'Heart',* Brill.

Cantarero, K., Parzuchowski, M., and Dukala, K. (2017), White lies in hand: Are other-oriented lies modified by hand gestures? Possibly not, *Frontiers in Psychology*, 8:814.

Couper-Kuhlen, E. and Selting, M. (2018), *Interactional linguistics*, Cambridge University Press.

Farley, S. D., Akin, K., and Hedgecoth, N. (2021), Exploring the meanings of the "heartfelt" gesture: A nonverbal signal of heartfelt emotion and empathy, *Journal of Nonverbal Behavior* 45, 567–585.

Goodwin, C. and Goodwin, M. H. (2000), Emotion within situated activity, in A. Duranti (ed.), *Linguistic Anthropology: A Reader*, Basil Blackwell, 239–257.

Hopper, P. J. and Traugott, E. C. (1993), *Grammaticalization*, Cambridge University Press.
Lange, J. (1887/2021), *The hand on the breast tr. Karl Johns, Journal of Art Historiography*, 25.
Mondada, L. (2016), Challenges of multimodality: Language and body in social interaction, *Journal of Sociolinguistics* 20(3), 336–366.
Müller, C. (2014), Gestural modes of representation as techniques of depiction, in C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem (eds), *Body—language—communication: An international handbook on multimodality in human interaction*, De Gruyter Mouton, 1687–1702
Parzuchowski, M., Szymkow, A., Baryla, W., and Wojciszke, B. (2014), From the heart: Hand over heart as an embodiment of honesty, *Cognitive Processing* 15, 237–244.
Quasinowski, B., Solmaz, A., Bachmann, C., Chen, W., Elcin, M., Kamisli, C. Liu, T., Maass, A. H., Merse, S., Morbach, C., Neumann, A., Neumann, T., Sommer, I., Stoerk, S., Weingartz, S., Weiss, A., and Wietasch, G. (2023), Hearts in their hands – physicians' gestures embodying shared professional knowledge around the world, *Sociology of Health and Illness* 45, 1101–1122.
Streeck, J. (2009), *Gesturecraft: The manu-facture of meaning*, John Benjamins.
Streeck, J. (2021), The emancipation of gestures, *Interactional Linguistic*

Uncovering the multimodal nature of figurative meaning using the EEG: the neural dynamics of metaphor-evoked mental imagery

Federico Frau, Paolo Canal, Maddalena Bressler, Chiara Pompei, Valentina Bambini

Laboratory of Neurolinguistics and Experimental Pragmatics - IUSS Pavia

federico.frau@iusspavia.it

Theoretical accounts of metaphor comprehension argue whether our ability to understand figurative expressions involves mainly imagery processes (Lepore & Stone, 2010) or pure amodal propositions (Sperber & Wilson, 2008), or a multimodal combination of both imagistic and propositional representations (Carston, 2018; Gibbs & Matlock, 2008). Evidence of imagery involvement in metaphors can be found in the literature, but this is limited to literal aspects of the metaphorically used words (e.g., the activation of sensory-motor information of 'snake' when hearing the metaphor 'highways are snakes'; Al-Azary & Katz, 2021), while it is unclear whether perceptual features associated with the mental image of the global metaphorical meaning (e.g., highways being dangerous) are activated in the brain. Here, we used EEG recordings to investigate the psychological reality of complex mental images associated with metaphorical expressions.

We designed a novel experiment where a picture (e.g., a woman with messy hair) was preceded by one of four different tasks, either matching or mismatching with the picture (**Fig. 1**): i) a metaphoric description (*some hairstyles are bushes*, MET); ii) a literal description (*some hairstyles are uncombed*, LIT); iii) an adjective prompting mental imagery (*uncombed*, IMA); iv) another picture (PHY). ERP and Time-Frequency (TF) analysis were conducted using Montecarlo cluster-based permutation tests: using the ERP responses of PHY as the benchmark, we contrasted MET *minus* LIT and IMA *minus* MET. Based on the expectation that in PHY the matching condition would evoke P300 responses while the mismatching condition N200 responses (Folstein & Van Petten, 2008; Friedman et al., 1988), we shaped the following predictions for the key contrasts: i) if metaphors trigger richer images than literal sentences, we would expect P300 (in matching) and N200 (in mismatching) responses to be greater in MET and IMA compared to LIT; ii) if metaphors do not benefit from greater imagistic processes than literal sentences, we would expect the P300 (in matching) and N200 (in mismatching) responses to be similar in MET and LIT, both differing from IMA.

ERPs results (**Fig. 2**) showed that, in the matching condition, MET differed from LIT (but not IMA) showing a greater P300-like response (i.e., aligning to PHY), while it differed from IMA (but not LIT) in the late positivity. In the mismatching condition, MET again differed from LIT (but not IMA) for a smaller N200-like response, being also different from IMA in later negative responses. For the TF analysis (**Fig. 3**), MET revealed greater alpha desynchronization compared to LIT and reduced theta synchronization compared to IMA across conditions. The difference between MET and IMA is also supported by exploratory source estimation analysis, which highlighted reduced medial-parietal activations in MET compared to IMA (**Fig. 4**).

These findings support a multimodal model of metaphor processing (Gibbs & Matlock, 2008; Paivio & Walsh, 1993) by showing that metaphoric meanings blend both amodal and imagistic representations. In particular, metaphoric representations were closer to pure imagistic ones than literal ones in the P300 response, being also easier to accommodate compared to literal ones as shown by the reduced N200 response in mismatching conditions. However, metaphoric representations differed from pure imagery in later components indicating that imagery might require extra effort to be integrated with abstract conceptual representations when not constrained by linguistic material (Dudschig et al., 2016; Hirschfeld et al., 2011). Similarly, oscillation analysis revealed differences from literal in terms of greater alpha decrease (indexing greater verbo-pictorial integration, Scharinger et al., 2020) and from imagery in terms of reduced theta synchronization (reduced visual properties retrieval from long-term memory, Canales-Johnson et al., 2021). In supporting a multimodal model of figurative meaning, metaphor-evoked

images are shown to be also more relevant and perceptually richer compared to those aroused by literal sentences.

References

Al-Azary, H., & Katz, A. N. (2021). Do metaphorical sharks bite? Simulation and abstraction in metaphor processing. *Memory & Cognition*, *49*(3), 557–570. https://doi.org/10.3758/s13421-020-01109-2

Canales-Johnson, A., Lanfranco, R. C., Morales, J. P., Martínez-Pernía, D., Valdés, J., Ezquerro-Nassar, A., Rivera-Rei, Á., Ibanez, A., Chennu, S., Bekinschtein, T. A., Huepe, D., & Noreika, V. (2021). In your phase: neural phase synchronisation underlies visual imagery of faces. *Scientific Reports*, *11*(1), 2401. https://doi.org/10.1038/s41598-021-81336-y

Carston, R. (2018). Figurative Language, Mental Imagery, and Pragmatics. *Metaphor and Symbol*, *33*(3), 198–217. https://doi.org/10.1080/10926488.2018.1481257

Dudschig, C., Mackenzie, I. G., Strozyk, J., Kaup, B., & Leuthold, H. (2016). The Sounds of Sentences: Differentiating the Influence of Physical Sound, Sound Imagery, and Linguistically Implied Sounds on Physical Sound Processing. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(5), 940–961. https://doi.org/10.3758/s13415-016-0444-1

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Friedman, D., Sutton, S., Putnam, L., Brown, C., & Erlenmeyer-Kimling, L. (1988). ERP Components in Picture Matching in Children and Adults. *Psychophysiology*, *25*(5), 570–590. https://doi.org/10.1111/j.1469-8986.1988.tb01893.x

Gibbs, R. W., & Matlock, T. (2008). Metaphor, imagination, and simulation. In R. W. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 161–176). Cambridge University Press. https://doi.org/10.1017/CBO9780511816802.011

Hirschfeld, G., Zwitserlood, P., & Dobel, C. (2011). Effects of language comprehension on visual processing – MEG dissociates early perceptual and late N400 effects. *Brain and Language*, *116*(2), 91–96. https://doi.org/10.1016/j.bandl.2010.07.002

Lepore, E., & Stone, M. (2010). Against Metaphorical Meaning. *Topoi*, *29*(2), 165–180. https://doi.org/10.1007/s11245-009-9076-1

Paivio, A., & Walsh, M. (1993). Psychological processes in metaphor comprehension and memory. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 307–328). Cambridge University Press. https://doi.org/10.1017/CBO9781139173865.016

Scharinger, C., Schüler, A., & Gerjets, P. (2020). Using eye-tracking and EEG to study the mental processing demands during learning of text-picture combinations. *International Journal of Psychophysiology*, *158*, 201–214. https://doi.org/10.1016/j.ijpsycho.2020.09.014

Sperber, D., & Wilson, D. (2008). A Deflationary Account of Metaphors. In R. W. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 84–106). Cambridge University Press. https://doi.org/10.1017/CBO9780511816802.007

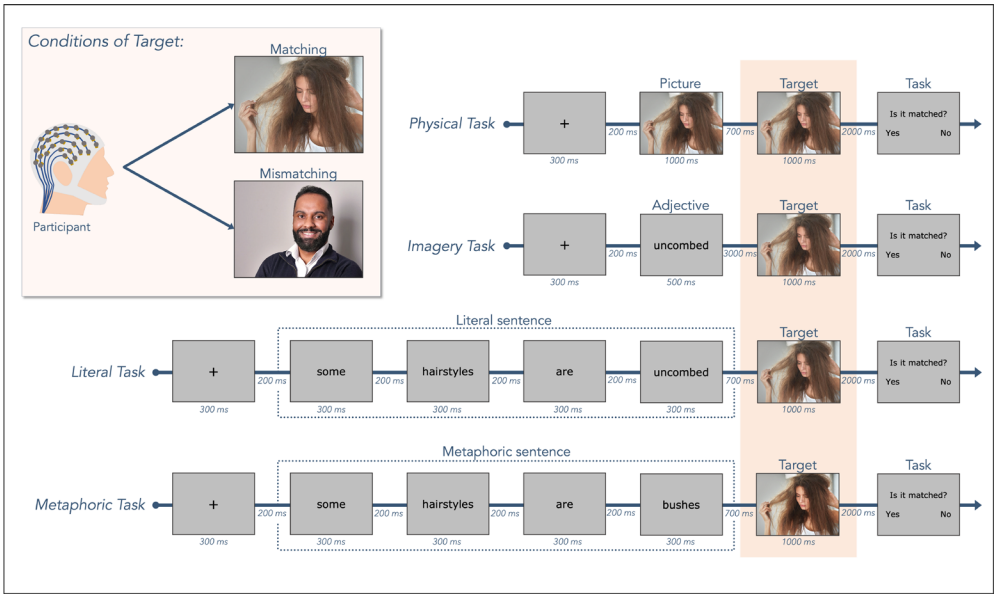


Fig 1. Trial structure of the four tasks included in the EEG experiment and example of matching and mismatching picture stimuli.

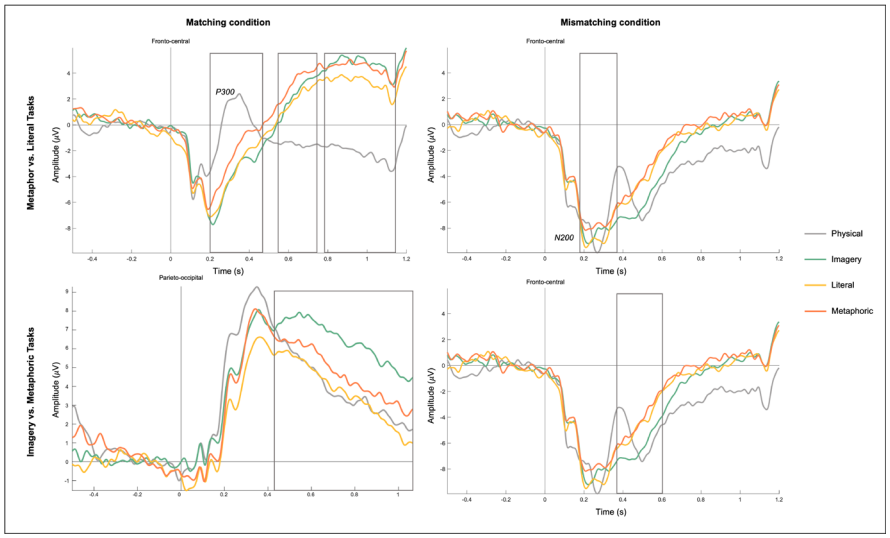


Fig 2. ERPs grand averages across tasks and significant clusters as identified by the cluster-based permutation tests on the ERPs. Fronto-central sites: Fz, FC1, FC2, F1, F2, FCz; Parieto-occipital sites: CP1, Pz, CP2, P1, PO3, POz, PO4, P2, CPz. The P300 and N200 responses are labeled for reference as the expected ERP patterns for the Physical Task.

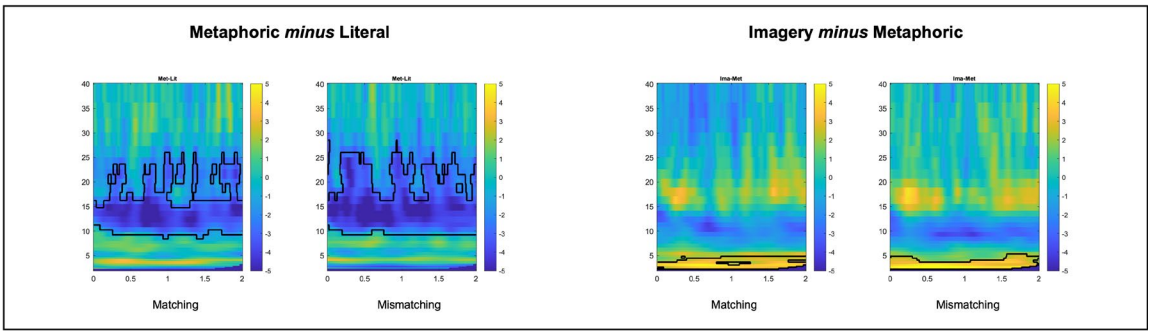


Fig 3. Time-frequency representation analysis with significant clusters as identified by the cluster-based permutation tests.

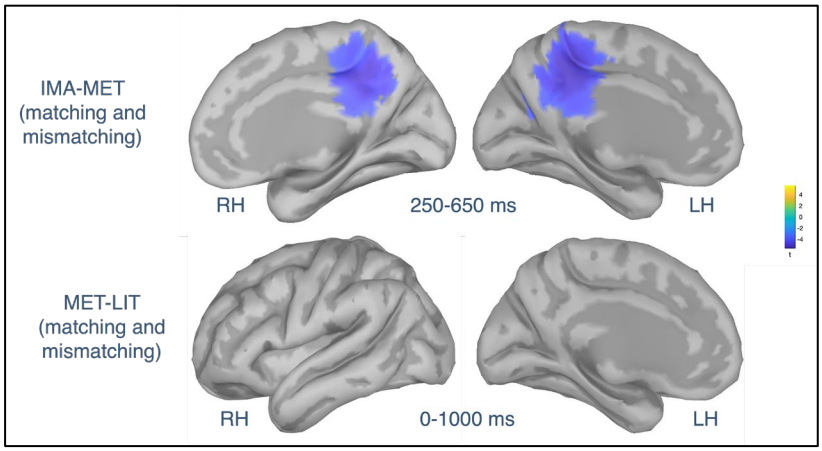


Fig 4. Exploratory source estimation analysis computed on ERP grand averages using sLORETA with unconstrained orientation, based on OpenMEEG BEM. Task comparisons showed reduced medial-parietal activations in MET compared to IMA, while no differences in source activations emerged between MET and LIT.

Distinct temporal dynamics of speech and gesture processing: Insights from ERPs across L1-Turkish and L2-English

Demet Özer[1*], Efe Soyman[2*], Ayşe Nur Badakul[2], Burcu Arslan[2], Fatma Sena Yılmaz[3], & Tilbe Göksun[2]

[1]Department of Psychology, Bilkent University, Ankara - Türkiye
[2]Department of Psychology, Koç University, Istanbul - Türkiye
[3]Department of Psychology, Boğaziçi University, Istanbul – Türkiye

demet.ozer@bilkent.edu.tr

During face-to-face communication, listeners integrate various communicative signals such as speech and speech accompanying iconic gestures that represent object attributes, actions, and events (for a review, see Özyürek, 2014). Earlier ERP studies with native listeners demonstrate the N400 effect in gesture-speech mismatch paradigms with more negative N400 amplitudes to gestures that convey mismatching than matching information in relation to the accompanying speech or an earlier context (e.g., Kelly et al., 2007; Özyürek et al., 2007; Wu & Coulson, 2005). Evidence also suggests distinct behavioral and neural dynamics of speech-gesture integration across first (L1) and second languages (L2), where the processing of visual semantic cues might be different in L1 and L2 (e.g., Dahl & Ludvigsen, 2014; Drijvers & Özyürek, 2018; Zhang et al., 2023). Building on prior work, this study examined the neural and behavioral underpinnings of multimodal language by examining how listeners process speech and co-speech gestures in their first (L1-Turkish) and second (L2-English) languages under varying attentional conditions (i.e., when attended the speech versus gesture channel).

Across two experiments, we recorded EEG activity (Experiment 1, 24 participants) and reaction times (Experiment 2, 24 participants) during a mismatch task where participants watched a 1-second video of an action prime (e.g., a man *chopping* vegetables with a knife) followed by a 1-second video of another actress describing the action in the preceding action prime by uttering a word and making a gesture across 4 major conditions: (1) both-match (e.g., saying "chop" and gesturing "chop"), (2) gesture-mismatch (e.g., saying "chop" and gesturing "twist"), (3) speech-mismatch (e.g., saying "twist" and gesturing "chop"), (4) both-mismatch. The both-mismatch condition was further divided into two sub-conditions: (4.1) local match (e.g., saying "twist" and gesturing "twist") and (4.2) local mismatch (e.g., saying "twist" and gesturing "shake") to investigate the effect of the mismatch between concurrent speech and gestures independently. Participants were asked to detect whether the gesture (gesture-focused task) or the speech (speech-focused task) in the second video was related to the preceding action prime. Speech was presented in Turkish or English in separate blocks. In Experiment 1, we specifically focused on N400 and N2 components as indices of late semantic processing and early sequential matching, respectively.

ERP waveforms elicited by the concurrent speech and gesture across languages, task focus, and mismatch conditions at the central electrode sites (average of Cz, C3, and C4) can be seen in Figure 1. In the gesture-focused task, there was a clear mismatch effect with more negative N400 amplitudes to mismatching than matching gestures only when the accompanying speech was matching (Figure 2A). This effect was comparable across L1 and L2. For the speech-focused task, on the other hand, the mismatch effect was present in N2 amplitudes with more negative N2 amplitudes to mismatching speech than matching speech, regardless of the accompanying gesture (Figure 2B). Again, this effect was comparable across L1 and L2. Moreover, the reaction times in Experiment 2 yielded parallel results with N400 amplitudes in the gesture-focused task and with N2 amplitudes in the speech-focused task.

When combined, our results provide compelling evidence for distinct neural and temporal dynamics in processing speech and gestures, as well as their asymmetric influences on one another. Processing speech might be instantiated earlier and independent of the accompanying visual signals,

which is evident in N2 amplitudes. Gestures, on the other hand, might engage late semantic processes and be influenced by the accompanying spoken signal. Notably, these patterns were comparable across first and second languages. Our findings highlight a complex interplay between modality, modality focus, and neural processing.

Figure 1. ERP waveforms across languages, task focus, and mismatch conditions at the central electrode sites (Cz, C3, and C4). The N400 amplitude was computed as the average voltage within the 390-517 ms and the N2 amplitude was computed as the average voltage within the 258-324 ms. The N2 and N400 boundaries indicate the time windows used to calculate time-averaged normalized N2 and N400 amplitudes, respectively. Shadings depict the standard errors of the means across 24 participants.

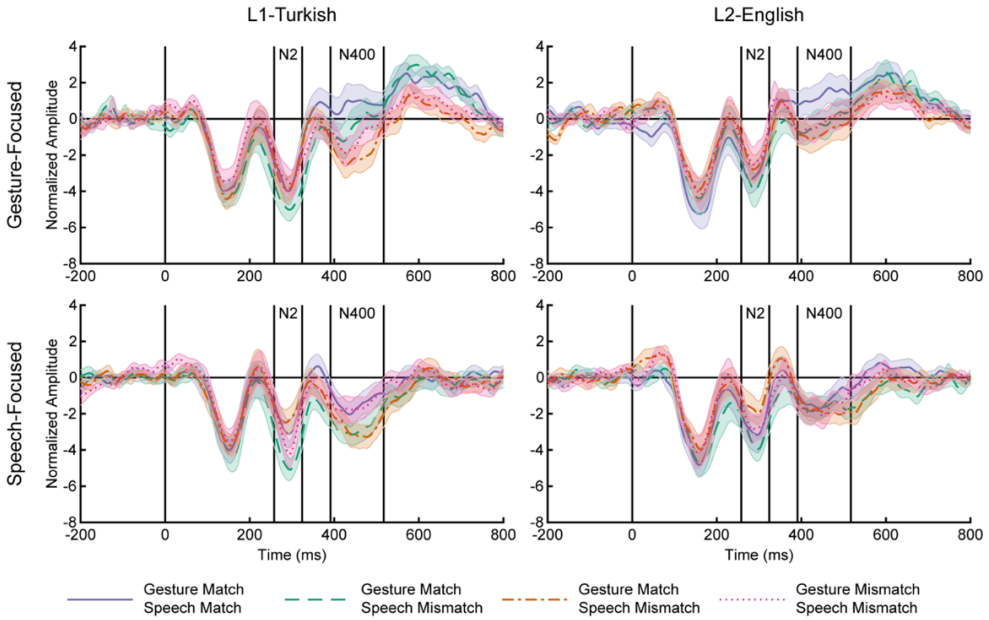
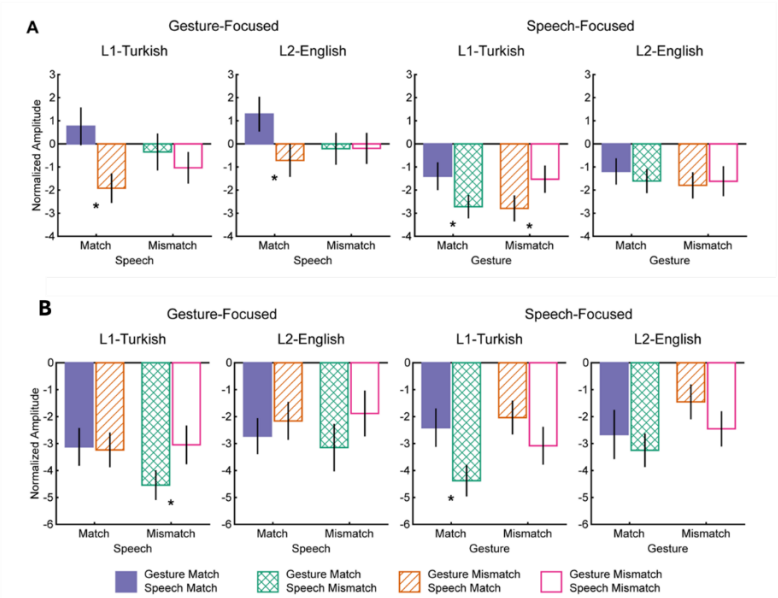


Figure 2. Normalized N400 (panel A) and N2 (panel B) amplitudes across languages, and task focus, and mismatch conditions.



References

Dahl, T. I., & Ludvigsen, S. (2014). How I see what you're saying: The role of gestures in native and foreign language listening comprehension. *The Modern Language Journal*, *98*(3), 813-833.

Drijvers, L., & Özyürek, A. (2018). Native language status of the listener modulates the neural integration of speech and iconic gestures in clear and adverse listening conditions. *Brain and Language*, *177*, 7-17.

Kelly, S. D., Ward, S., Creigh, P., & Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and Language*, *101*(3), 222-233.

Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*(4), 605-616.

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130296.

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, *42*(6), 654-667.

Zhang, Y., Ding, R., Frassinelli, D., Tuomainen, J., Klavinskis-Whiting, S., & Vigliocco, G. (2023). The role of multimodal cues in second language comprehension. *Scientific Reports*, *13*(1), 20824.

Multimodal cues as predictors in feedback categorization

Tobias-Alexander Herrmann[1], Sonja Gipper[2], Jana Hosemann[3], & Anastasia Bauer[2]

[1] Slavic Department, Faculty of Arts and Humanities, University of Cologne, Germany.
[2] Department of Linguistics, General Linguistics, Faculty of Arts and Humanities, University of Cologne, Cologne, Germany.
[3] Department of Rehabilitation and Special Education, Faculty of Human Sciences, University of Cologne, Cologne, Germany.

anastasia.bauer@uni-koeln.de

Any conversation among humans is rife with feedback, interactional moves that display some kind of stance towards another interlocutor's utterance [1]. Feedback serves various conversational functions: it can indicate passive recipiency to signal that the speaker can continue (continuer), acknowledge or agree with a claim (acknowledgment), mark information as new (newsmark), or evaluate information (assessment) [2].

Recent research studying naturalistic conversational data unveiled feedback as a fundamentally multimodal phenomenon involving the coordination of different channels [3]–[7]. However, the role of multimodal cues in differentiating feedback categories in face-to-face interactions remains underexplored. It is known that various multimodal feedback signals are combined to larger feedback events [8]. Feedback signals may take the form of lexical cues (words like *ja* 'yes' or signs such as STIMMT 'right'), non-lexical cues (vocalizations like 'mm' or manual gestures such as palm-up), and non-manual cues such as head nods, eyebrow raise, smiles, laughter and most importantly the combination thereof. Previous research has shown that, i.e. the head is the most often used articulator in the production of feedback [9]. Not known is yet the specific contribution of each multimodal cue to feedback categorizing as our understanding of how vocal, visual, manual and non-manual signals combine into complex feedback events in everyday conversation across different language modalities is limited.

This study investigates which multimodal cues best predict feedback categories, using data from four languages: German Sign Language (DGS), Russian Sign Language (RSL), spoken German (GER) and spoken Russian (RUS) [10]–[13]. We analyzed 45 minutes of dyadic conversations per language, identifying 1,900 feedback events comprising 3,400 feedback signals. Using parallel annotation and analysis, we modeled the relationship between multimodal signal combinations and feedback categories ('continuer,' 'newsmark,' 'assessment,' 'acknowledgment'). Modeling this relation helps to tap into the way how a speaker/signer arrives at the interpretation of their interlocutor's feedback event, providing quantifiable metrics. We use random forests to determine the relevance of each multimodal signal for feedback events. Based on the (most) relevant predictors, we then run multinomial regression models. We extract coefficients that indicate the strength and direction of the relationship between predictors and each feedback category, which are then exponentiated to calculate odds ratios for each predictor across feedback categories, representing the multiplicative change in the odds of a category's occurrence associated with a one-unit change in the predictor. In this fashion, we provide empirical evidence demonstrating how individual feedback signals influence the likelihood of a larger feedback event being interpreted as belonging to a specific feedback category.

The data reveal that analyzing multimodal cues as predictors for differentiating feedback categories holds significant potential to advance our understanding of feedback dynamics, offering deeper insights into the interplay of these cues. Our findings underscore the importance of cross-linguistic and cross-modal research in understanding how multimodal cues shape face- to-face interaction.

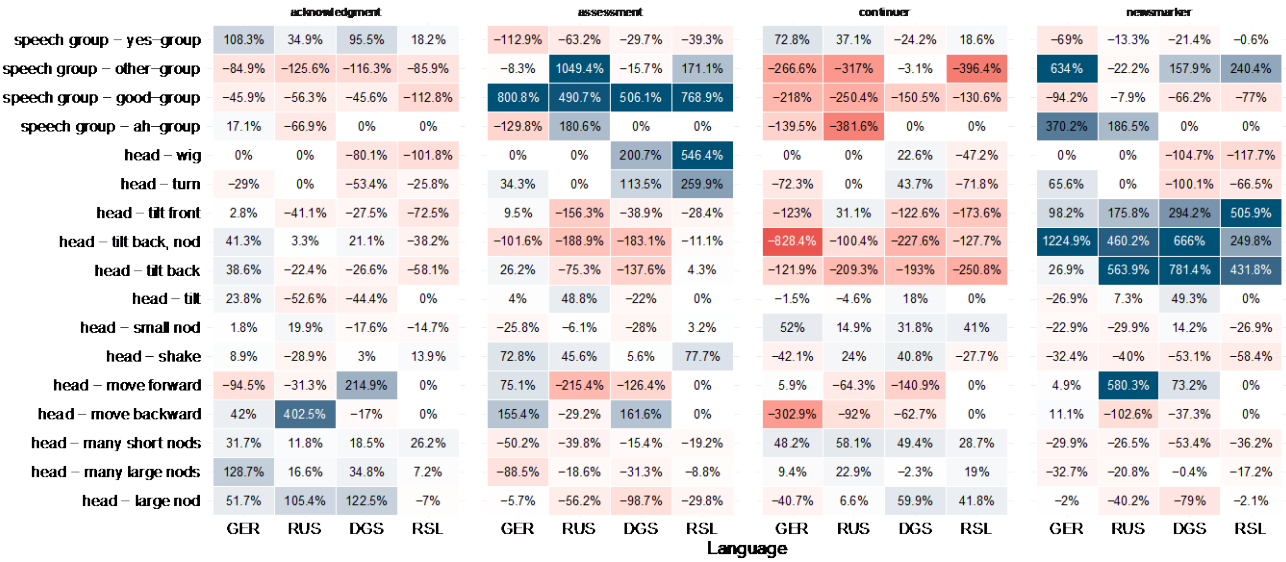


Figure 1: *Likelihood of the two predictors' (*SPEECH GROUP *and* HEAD*) feedback signals to affect the interpretation of the feedback category. The direction of change is indicated by signed percentages and the colored cells, with blue showing an increase, red a decrease.*

References

[1] J. Allwood, J. Nivre, and E. Ahlsén, "On the semantics and pragmatics of linguistic feedback," *Journal of Semantics*, vol. 9, no. 1, pp. 1–26, 1992.

[2] R. Gardner, *When listeners talk: Response tokens and listener stance* (Pragmatics & Beyond New Series). Amsterdam: John Benjamins Publishing Company, 2001, vol. 92.

[3] J. Allwood and L. Cerrato, "A study of gestural feedback expressions," in *First Nordic Symposium on Multimodal Communication.*, Copenhagen: Gothenburg University Publications, 2003, pp. 7–22.

[4] J. Allwood, S. Kopp, K. Grammer, E. Ahlsén, E. Oberzaucher, and M. Koppensteiner, "The analysis of embodied communicative feedback in multimodal corpora: A prerequisite for behavior simulation," *Language Resources and Evaluation*, vol. 41, no. 3-4, pp. 255–272, 2007.

[5] K. P. Truong, R. Poppe, I. D. Kok, and D. Heylen, "A multimodal analysis of vocal and visual backchannels in spontaneous dialogs," in *Proc. Interspeech 2011*, Aug. 2011, pp. 2973–2976.

[6] C. Navarretta and P. Paggio, "Classification of feedback expressions in multimodal data," in *Annual Meeting of the Association for Computational Linguistics, 48th ACL*, Uppsala, Sweden: Association for Computational Linguistics, 2010, pp. 318–324.

[7] Z. Malisz, M. Włodarczak, H. Buschmeier, J. Skubisz, S. Kopp, and P. Wagner, "The ALICO cor- pus: Analysing the active listener," *Language Resources and Evaluation*, vol. 50, no. 2, pp. 411– 442, Jun. 2016.

[8] A. Bauer, S. Gipper, T.-A. Herrmann, and J. Hosemann, *Multimodal feedback signals: Comparing response tokens in co-speech gesture and sign language*: Poster presented at the 1st International Multimodal Communication Symposium: Barcelona, 2023, https://sslac.uni-koeln.de/sites/sslac/PDFs/ MMSYM.pdf.

[9] H. Lutzenberger, L. D. Wael, R. Omardeen, and M. Dingemanse, "Interactional infrastructure across modalities: A comparison of repair initiators and continuers in British Sign Language and British English," *Sign Language Studies*, vol. 24, no. 3, pp. 548–581, 2024.

[10] R. Konrad, T. Hanke, G. Langer, *et al.*, *MY DGS – annotated. Public Corpus of German Sign Language, 3rd release*, 2020.

[11] S. Burkova, *Russian Sign Language Corpus*, Novosibirsk University, 2015. Available: http://rsl.nstu.ru/.

[12] A. Bauer, *Russian multimodal conversational data*, 2023. Data Center for the Humanities. https://doi.org/10.18716/dch/a.00000016.

[13] A. Bauer and R. Poryadin, *Russian Sign Language conversations*, 2023. Data Center for the Humanities, https://doi.org/10.18716/dch/a.00000028.

The emergence of multimodal grammatical construct(ion)s through pointing gestures

Kurt Feyaerts
University of Leuven

kurt.feyaerts@kuleuven.be

This contribution is situated at the crossroads between the paradigms of interactional and cognitive linguistics, from where it focuses on pointing (or deictic) gestures as a crucial semiotic resource for the multifaceted interpretation of a usage event. Accordingly, the overall objective of this paper is to demonstrate how pointing gestures may actively contribute to the multimodal realization of grammatical construct(ion)s. The empirical basis for this qualitative study is a corpus of two plenary debates in the Flemish Parliament, of which four video excerpts will be analyzed more closely.

It is generally accepted that pointing gestures cannot be categorized as mere reference markers (Kita 2003). Several studies have convincingly shown that these gestures may occur with many formal variations as well as in multimodal co-occurrences (Fricke 2007; Ladewig 2020). Mondada (2014) demonstrates that pointing gestures are dynamically adapted in function of different socio-material circumstances and interactional recipients whereas the dialogue-based account of Ginzburg & Lücking (2021) elaborates the existing semantic taxonomy of pointing gestures by four types of addressees pointing.

The present contribution will demonstrate by video excerpts taken from the corpus and transcribed as in the following examples (our translation), how pointing gestures may play a decisive role in the multimodal realization of a ditransitive construction (example 1), an argument structure underlying an attributive participle (*shared* in 2) and, finally, in the realization of a parenthetical construction. The underlined elements in these transcriptions mark co-occurrence with a pointing gesture.

(1) …a lot of questions have been asked here…
(2) …because I think that is a shared concern…
(3) … because (--) government investments also have a positive effect…

In (1) the speaker uses both hands to point at himself, thus indicating that all questions have been asked to him. In doing so, the gesture clearly impacts the syntactic organization of the argument structure as it provides a perfect realization of a multimodal ditransitive construction. In (2) the speaker points at herself and the previous speaker thus identifying two referents of the argument structure of the underlying verb 'share', which factors into the specific pragmatics of this utterance. In (3), during a short pause following the Dutch conjunct 'want', the speaker points to the previous speaker thus expressing a multimodal realization of the parenthetical construction along with a clear impact on the discursive and the pragmatic organization of the usage event.

Our focus on the integration of pointing gestures along with locally situated aspects of interaction in grammatical construct(ion)s may feed into a new debate about a) the status of non-verbal and multimodal structures within construction networks (Diessel 2020: 12; Zima 2014; Schoonjans et al. 2015; Bergs & Zima 2017), and b) the relative status (in terms of prominence) of different types of formal information (verbal vs. gestural) within a construction. With regard to the c) semantic pole of a construction, the integration of pointing gestures raises the cognitive linguistic issue of objective vs. subjective construal as a highly relevant and refining, but hitherto largely ignored dimension on the CxG agenda.

References

Diessel Holger. 2020. A Dynamic Network Approach to the Study of Syntax. Frontiers in Psychology. 11:604853. doi: 10.3389/fpsyg.2020.604853

Fricke, Ellen. 2007. *Origo, Geste und Raum: Lokaldeixis im Deutschen*. Berlin: Mouton De Gruyter.

Fried, Mirjam & Jan-Ola Östman. 2005. Construction Grammar: A thumbnail sketch. In: M. Fried & J-O. Östman (eds.), *Construction Grammar in a cross-language perspective*, 11-86. John Benjamins.

Ginzburg, Jonathan & Andy Lücking. 2021. I thought pointing is rude: A dialogue-semantic analysis of pointing at the addresse In: P. G. Grosz, L. Martí, H. Pearson, Y. Sudo, & S. Zobel (eds.) *Proceedings of Sinn und Bedeutung 25*, 276–291.

Kita, Sotaro (ed). 2003. *Pointing: Where Language, Culture and Cognition Meet*. Mahwah, NJ: Lawrence Erlbaum.

Ladewig, Silva. 2020. *Integrating Gestures The Dimension of Multimodality in Cognitive Grammar*. Berlin: Mouton De Gruyter.

Mondada, Lorenza. 2014. Pointing, talk and the bodies. Reference and joint attention as embodied interactional achievements. In M. Seyfeddinipur & M. Gullberg (Eds.). *From Gesture in Conversation to Visible Action as Utterance: Essays in honor of Adam Kendon* (pp. 95-124). Amsterdam – Philadelphia: Benjamins.

Schoonjans, Steven, Geert Brône & Kurt Feyaerts. 2015. Multimodalität in der Konstruktionsgrammatik: Eine kritische Betrachtung illustriert anhand einer Gestikanalyse der Partikel einfach. In J. Bücker, S. Günthner & W. Imo (eds.), Konstruktionsgrammatik V. Konstruktionen im Spannungsfeld von sequenziellen Mustern, kommunikativen Gattungen und Textsorten, 291–308. Tübingen: Stauffenburg

Zima, Elisabeth. 2014. Gibt es multimodale Konstruktionen? Eine Studie zu [V(motion) in circles] und [all the way from X PREP Y]. Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 15. 1–48.

Zima, Elisabeth & Alexander Bergs. 2017. Multimodality and construction grammar. Linguistics Vanguard 2017; 3(1)

A visually annotated multimodal corpus of LIS

Chiara Bonsignori, Alessio Di Renzo, Gabriele Gianfreda, Tommaso Lucioli, Barbara Pennacchi, Luca Lamano, Olga Capirci
ISTC-CNR Roma

chiara.bonsignori@cnr.it

The present study aims to present a visually annotated Italian Sign Language (LIS) corpus and to propose it as a tool for multimodal corpora annotations.
The corpus has been developed in the context of the Learning for ALL (L4ALL) Project, which seeks to develop a sign language ontology, enabling simultaneous access to different sources of linguistic information, reflecting the multilinear nature of sign language structure. The overall goal of the project is to make use of the LIS corpus in an AI environment.
Signed utterances are constructed by the simultaneous combinations of different articulators (i.e., mouth, eye gaze, facial expressions, head movements, shoulder movements, hands), each one conveying a semantic value in the utterance constructions. Signed languages, being realized in the three-dimensional space, should not be forced into linear representations.
Because of this multilinear nature - a feature inherent not only to signed languages but to all types of face-to-face communication - researchers struggled for a long time to identify the proper way to annotate signs (Antinoro Pizzuto *et al.* 2010, Slobin, 2008). SignWriting (SW) system has been proposed since the ’80 by Valerie Sutton as a promising solution to annotate the multilinear features of sign languages (Sutton, 1995). SW uses graphic symbols to represent handshapes, movements, facial expressions, and eye gaze, allowing a spatial arrangement of the units that refer to the face-to-face sign, and visually showing their simultaneous articulation (Di Renzo et al. 2010). Recently, research on artificial intelligence (AI) opened new discussions in the sign annotation field, and SW proved to be a good candidate for being an intermediate language for automatic sign language recognition and translation (Yin et al., 2021; Jiang et al., 2023). Starting from these recent and promising results, we developed a corpus of more than 4000 signs from 240 LIS videos across different linguistic genres (e.g., descriptive, narrative, expository, and argumentative text).
The corpus was annotated by deaf research assistants through the ELAN software, using Formal SignWriting (FSW), a standard system of SW encoding with plain text (ASCII). The corpus has been annotated using four tiers: segmentation; the matching FSW for each signed unit; Italian descriptions for each signed unit; and an Italian translation of the entire LIS sentence. Our approach utilizes the link between each sign and its description, maximizing and allowing our corpus to overcome the limitations that annotated signed corpora currently display.
Considering that SW allows the simultaneous representation of the articulators, we linked it to a semantic description, aiming to represent the co-construction of meaning. For doing so, we developed a list of shared rules to take detailed note of the iconic aspects of the units and enable semantic networks to be created between them, that can be later exploited by language models.
Finally, we discuss the impact that our work on multilinear data annotation could have in the field of multimodal research in face-to-face communication.

References:

Antinoro Pizzuto, E., Chiari, I., Rossini, P. (2010). Representing Signed Languages: Theoretical, Methodological and Pratical Issues. *Spoken Communication*. P. Massimo, G. Antonella, C. Isabella and D. Francesca. Cambridge, MA, Cambridge Scholars Publishing**,** 205-241

Di Renzo, A., Lamano, L., Lucioli, T., Pennachi, B., Gianfreda, G., Petitta, G., Bianchini, C.S., Rossini, P., e Antinoro Pizzuto, E. (2011). *Scrivere la LIS con il SignWriting: manuale introduttivo.* Roma: ISTC CNR.

Jiang, Z. Moryossef, A., Müller, M., and Ebling, S. (2023). Machine Translation between Spoken Languages and Signed Languages Represented in SignWriting. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 1706–1724, Dubrovnik, Croatia. Association for Computational Linguistics.

ELAN (Version 6.9), [computer software]. (2024), Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive.

Yin, K., Moryossef, A., Hochgesang, J., Goldberg, Y., & Alikhani, M. (2021). Including Signed Languages in Natural Language Processing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, Vol.1, 7347-7360.

Slobin, D.I. (2008). Breaking the Molds: Signed Languages and the Nature of Human Language. *Sign Language Studies 8*(2), 114-130.

Sutton V. (1995). Lessons in Sign Writing: textbook & workbook. La Jolla (California, USA), Deaf Action Committee for Sign Writing.

Multimodality and Illocution:
The Role of Gestures in the Construction of Linguistic Action

Luca Lo Re
Università degli Studi di Firenze

lucalore@unifi.it

Language production adheres to the principles of the linguistic system to formulate utterances that convey the intended meaning. Nonetheless, in addition to expressing meanings, our utterances also execute actions. As per the model established by Austin, linguistic action comprises three distinct acts: the locutionary act, the illocutionary act, and the perlocutionary act (Austin, 1962).

Within Austin's theoretical framework, the illocutionary act is pivotal, signifying an intermediate level between the act of stating something and its subsequent effects. Over the years, this concept has prompted various interpretations among scholars: the conventionalist perspective, which views illocution as the generation of conventional effects through activities regulated by convention, and the intentionalist perspective, which regards illocution as the articulation of communicative intent (Sbisà, 2013). Nearly all scholarly contributions concerning illocution are predicated on an exclusive analysis of verbal language, neglecting the significance of the gestural channel's contribution.

Kendon (1995, 2004), by adopting a multimodal perspective, was among the pioneers in observe that certain emblematic gestures mark the illocutionary acts conveyed by utterances. In the years following, additional studies have been conducted on this subject. Bressen and Müller (2014) identified a repertoire of recurrent gestures in the German language that express various illocutionary values. Furthermore, Wilson (2024) acknowledges that representational gestures can fulfill specific pragmatic functions, thereby influencing the illocutionary interpretation of verbal expressions and contributing to the manifestation of complex linguistic structures acts.

It is unequivocally acknowledged, as evidenced by numerous studies, that emblematic and representational gestures possess specific pragmatic and illocutionary functions. Moreover, alongside these gestures, which are distinguished by a significant degree of iconicity or conventionality, ANONYMIZED noted that in spontaneous speech, multimodal strategies for the construction of linguistic action may incorporate baton gestures. Specifically, three distinct strategies can be identified: the synchronization of speech and gesture in conveying the same illocutionary value, gestures that modify the illocution expressed verbally, and gestures that independently convey the illocution. This raises the question: Are gestures merely indicators of illocution, or do they play a contributory role in its expression?

The study aims to address the inquiry by evaluating the agreement on the classification of illocution. It encompasses three groups of participants: the initial group was assigned the task of identifying the type of illocutionary act conveyed solely through the verbal channel (by listening to an audio file); the second group undertook the same task but focused exclusively on the gestural channel (by viewing a video file), while the third group was exposed to multimodal data. The illocutionary values employed are Assertion, Expression, Refusal, and Rite (Searle, 1969; Cresti, 2020). The spoken data utilized for the agreement assessment were derived from the CorMIP corpus, which compiles spontaneous interactions among speakers annotated pragmatically in accordance with the L-AcT model. The study illustrates that linguistic action is multimodal and speakers employ all available channels to convey illocution. Moreover, it appears plausible to hypothesize that such orchestration is influenced by the speaker's intention, adhering to an agentivity model predicated on feedback.

References

Austin J. L. (1962), *How to Do Things with Words*. Oxford: University Press.

Bressem J. e Müller C. (2014). “A Repertoire of German Recurrent Gestures with Pragmatic Functions”, in Handbücher Zur Sprach-Und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (Hsk) 38/2,. Berlino, De Gruyter Mouton. Pp. 1575–91.
Cresti E. (2020), *The pragmatic analysis of speech and its illocutionary classification according to the Language into Act Theory*, in Izre´el S., Mello H., Panunzi A., Raso T. (a cura di). In search of basic units of spoken language: A corpus-driven approach, Amsterdam, John Benjamins, 181-219.
Kendon A. (1995), “Gestures as Illocutionary and Discourse Structure Markers in Southern Italian Conversation”, in Journal of Pragmatics, 23, 3, pp. 247-279.
Kendon A. (2004), *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
Sbisà M. (2013), «1. Locution, Illocution, Perlocution». In *Pragmatics of Speech Actions*, a cura di Marina Sbisà e Ken Turner, 25–76. DE GRUYTER, 2013. https://doi.org/10.1515/9783110214383.25.
Searle J.R. (1969), *Speech Acts. An essay in the philosophy of language*, Cambridge, Cambridge University Press.
Wilson J. (2024), *Pragmatics, Utterance Meaning, and Representational Gesture*, Cambridge, Cambridge University Press.

Where the sidewalk begins: searching for the origins of multimodal linguistic structures

Laura Sparaci
ISTC-CNR Roma

laura.sparaci@istc.cnr.it

Multiple studies have shown that language development is an inherently multimodal process, encompassing not only cognitive-symbolic skills, but also visuo-motor and/or proprioceptive competencies (Northrup & Iverson 2020; Yu & Smith 2013; Sparaci & Volterra 2017; Volterra et al. 2017). In child development this has led to reconsider the question of the origins of linguistic structures, such as narratives. In fact, once we accept that language extends well beyond speech, we find ourselves questioning where does multimodal language begin and we are faced with the difficult task of defining new boundaries for the emergence of specific linguistic skills.

Considering narratives, some have traced their origins in pre-verbal action formats characterized by a four stage structure (i.e., a canonical state, a precipitating event, a restoration and a coda marking the end) (Bruner 1990; Bruner & Feldman 1993). For example, the game of peek-a-boo, where: (1) mutual gaze is established, (2) the face is hidden, (3) it is shown again, and (4) "boo" marks the end of the game (Dautenhanh 2002). However, there are currently two contrasting theories on pre-verbal action formats and their relation to narratives. The first one, states that pre-verbal action formats, with a four-phase structure, are to be considered *per se* as early forms of embodied narratives (Delafield-Butt & Trevarthen 2015). The second one, suggests that while narratives may well be embodied and anchored in pre-verbal acts, the two must be kept well distinct (Gallagher & Hutto 2019; Gallagher 2020; Sparaci & Gallagher 2023). For the former approach, pre-verbal actions are already a form of multimodal linguistic structures, for the latter a distinction should be made. So the ultimate question seems to be: where does the sidewalk of multimodal language begin?

Starting from this question I will analyze the emergence of embodied narratives from early pre-verbal acts. In particular, by assuming a multimodal approach and considering the pivotal role of gestures, my attempt will be to show that pre-verbal action formats are purely performative structures (i.e., actions selected to convey meaning, but still linked to a specific here and know, which require presence of contextual or contingent background knowledge to be understood, Bates et al. 1975). On the other hand, narratives imply selecting specific contents from the continuous flow of everyday interactions, parsing them out and restructuring them in time and space. In this process content or plot is distinguished from the way in which it is conveyed, with an important shift in perspective (Sparaci & Gallagher 2023). In this sense, while we can still establish an important continuity from pre-verbal actions to multimodal linguistic structures such as narratives, we must also allow for relevant differences, outlining a perspective in which the passage from actions to multimodal language can be described as a form of a continuity through change.

References

Bates, E., Camaioni, L. & Volterra, V. (1975). The acquisition of performatives prior to speech, *Merrill-Palmer Quarterly of Behavior and Development*, 21(3), 205–226.

Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Bruner, J. & C. Feldman (1993). Theories of mind and the problem of autism. In S. Baron-Cohen, H. Tager-Flusberg, D. J. Cohen (Eds.), *Understanding other Minds: Perspectives from Autism* (267-291). Oxford: Oxford University Press.

Dautenhahn, K. (2002). The origins of narrative: In search of the transactional format of narratives in humans and other social animals. *International Journal of Cognition and Technology*, *1*(1), 97-123.

Delafield-Butt, J. T., & Trevarthen, C. (2015). The ontogenesis of narrative: from moving to meaning. *Frontiers in psychology*, *6*, 1157.

Gallagher, S. (2020). How moving is sometimes thinking. *Idea Journal*, *17*(02), 58-68.

Gallagher, S., & Hutto, D. D. (2019). Narratives in embodied therapeutic practice: Getting the story straight. In *The Routledge International Handbook of Embodied Perspectives in Psychotherapy* (28-39). Routledge.

Northrup, J. B., & Iverson, J. M. (2020). Multimodal coordination of vocal and gaze behavior in mother–infant dyads across the first year of life. Infancy, 25(6), 952-972.

Sparaci, L., & Gallagher, S. Continuity through Change: How Gestures Inform Current Debates on the Ontogeny of Embodied Narrative, In G. Maddalena, F. Ferrucci, M. Bella and M. Santarelli (Eds.), *Gestures. Approaches, Uses and Developments* (251-271), Berlin/Boston: De Gruyter.

Sparaci, L., & Volterra, V. (2017). Hands shaping communication: From gestures to signs. In Bertolaso, M. & Di Stefano, N. (Eds.), *The hand: Perception, cognition, action* (29-54), Springer.

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. Cognition, 125(2), 244-262.

Volterra, V., Capirci, O., Caselli, M.C., Rinaldi, P. & Sparaci, L. (2017). Developmental evidence for continuity from action to gesture/sign. *Language Interaction and Acquisition*, 8(1), 13-42.

Multimodal utterance in the autistic spectrum. A case study

Valentina Saccone[1], Giorgina Cantalini[2], Massimo Moneglia[1]
[1]Università degli Studi di Firenze
[2]Scuola di Teatro Paolo Grassi, Milano

valentina.saccone@unifi.it

Our study investigates the multimodal characteristics of communication in an autistic adolescent (19 y/o) by analyzing spontaneous and semi-spontaneous spoken productions in monological and interactive contexts. Autistic individuals often face challenges in pragmatic communication, frequently attributed to "deficient or delayed mind-reading abilities" (Happè, 1993; Baron-Cohen, 1995; Sperber & Wilson, 2002). These challenges can lead to breakdowns in conveying messages effectively and establishing interpersonal connections (Volkmar, 2020). Previous studies report impairments in both verbal and non-verbal pragmatic skills (McCann *et al.*, 2007; So *et al.*, 2015), including atypical gestures, issues with synchrony between gesture and speech (de Marchena & Eigsti, 2010), limited variety (Wetherby & Prutting, 1984; Colgan *et al.*, 2006) and difficulties with spatial referencing gestures (Hobson *et al.*, 2010). This contributes to reduced communicative effectiveness and limits social interaction (Duffy & Healy, 2011), impacting the development of communication skills and non-verbal behaviors (Eigsti, 2013; Sparaci *et al.*, 2019).

Autism is also characterized by repetitive sensorimotor movements or self-adaptors (e.g., fidgeting, Mahmoud *et al*., 2013; Froiland & Davison, 2016). While often perceived as distracting to social interaction and bizarre (Lewis & Bodfish, 1998; Leekam *et al.*, 2011), they may serve communicative purposes (Bascom, 2012) and are increasingly recognized as integral to autistic embodiment (Nolan & McBride, 2015; Chen, 2024).

Our analysis addresses the following questions: How do social engagement patterns differ between monological and interactive contexts, considering the supposed "disconnection from others"? What occurs when co-speech gestures are absent? Can stimming movements fulfill pragmatic functions? Lastly, is the lack of articulation of information identified in previous research (Saccone *et al.*, 2023) more pronounced in monological contexts?

Using the Language into Act Theory framework (Cresti, 2000; Cresti & Moneglia, 2018), we analyze speech information structure and prosodic parameters, including utterance complexity, dysfluency, mean length of prosodic units, pitch variation, and speech rate. Gestures are segmented hierarchically into Gesture Units, Phrases, and Phase (Ladewig & Bressem, 2013); Kita et al., 1998). Specific focus is given to self-adaptors such as self-manipulators (e.g., scratching one's leg; Chan *et al.*, 2016) and gestures associated with cognitive processes (e.g., "Butterworths"; McNeill, 2005). Small gestures and stimming movements are meticulously transcribed, revealing otherwise unnoticeable social actions (Saccone *et al.*, 2023; Chen, 2024). Non-verbal cues, like head and leg movements, are also annotated.

Preliminary findings reveal monotonous information structures and prosody, characterized by slow, elongated prosodic units with limited speech rate and information type variation. In the monologue, there is also a clear prevalence of scanning the locutive content into prosodic units rather than aligning it with the speech's pragmatic architecture and textual hierarchy. This results in information management that is more connected to the prosodic parsing of the locutive content than to its pragmatic composition and in the repetition of the same prosodic patterns (Fig.1).

Gesture reduction is evident across multiple gesture properties and is more pronounced in monological contexts. Overall, self-adaptors tend to replace gestures in accompanying speech.

In our 9-minute sample, the subject performs 541 Information Units, recording only 53 gestures (22 beats), always limited to the Stroke phase, and occurring in peripheral locations (e.g., on the belly or lap). In the monologue, gestures are never gathered within Gesture Units. Considering the Kendon's continuum, no Emblems are observed, and the few idiosyncratic Iconic gestures never

depict objects. Recurrent gestures (Ledwing, 2014; 2024; Bressem & Muller, 2014) are strongly reduced. For instance, Palm Up Open Hand is reduced to hand turn-out. In the monologue, just finger-lifting (Cienki, 2021) (Fig.2). Similarly, the gesture for negation (Palm down horizontally across the body; Harrison, 2009) is minimized in trajectory or with the hand stuck on the lap (Fig.3).

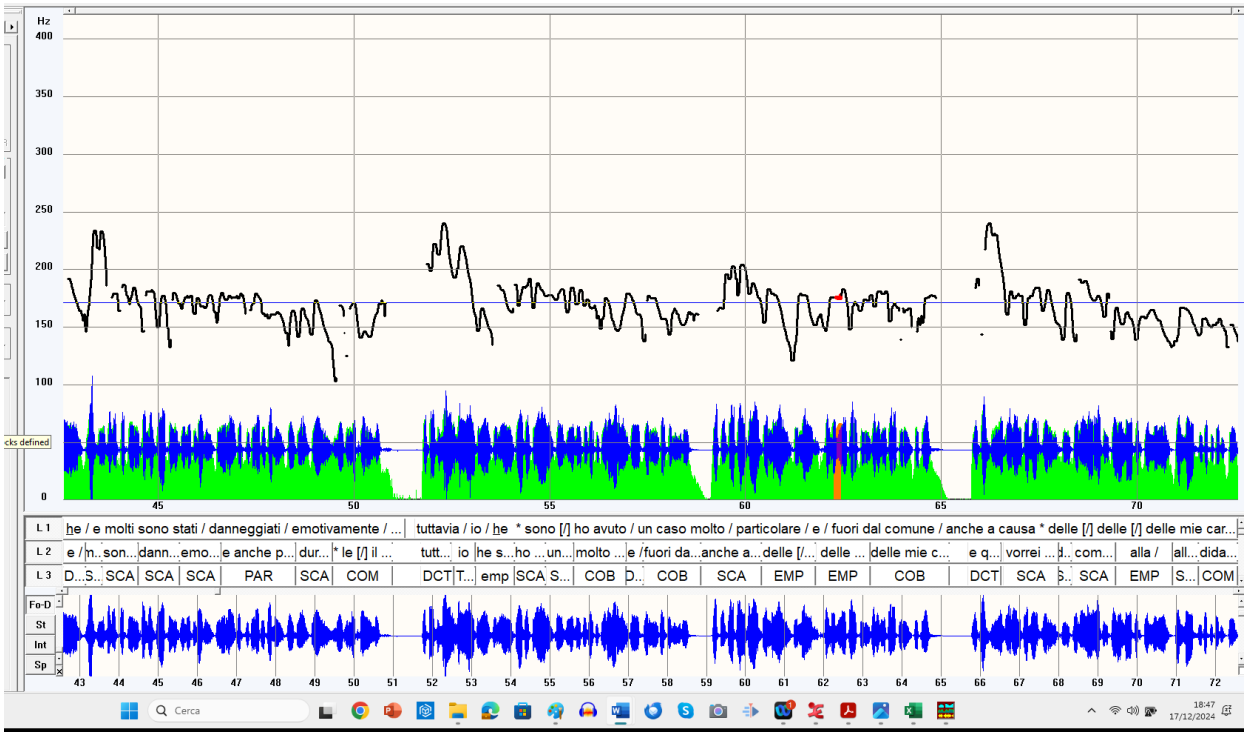


Fig.1 Monotonous repetition of prosodic patterns in the monologic performance (30 seconds)

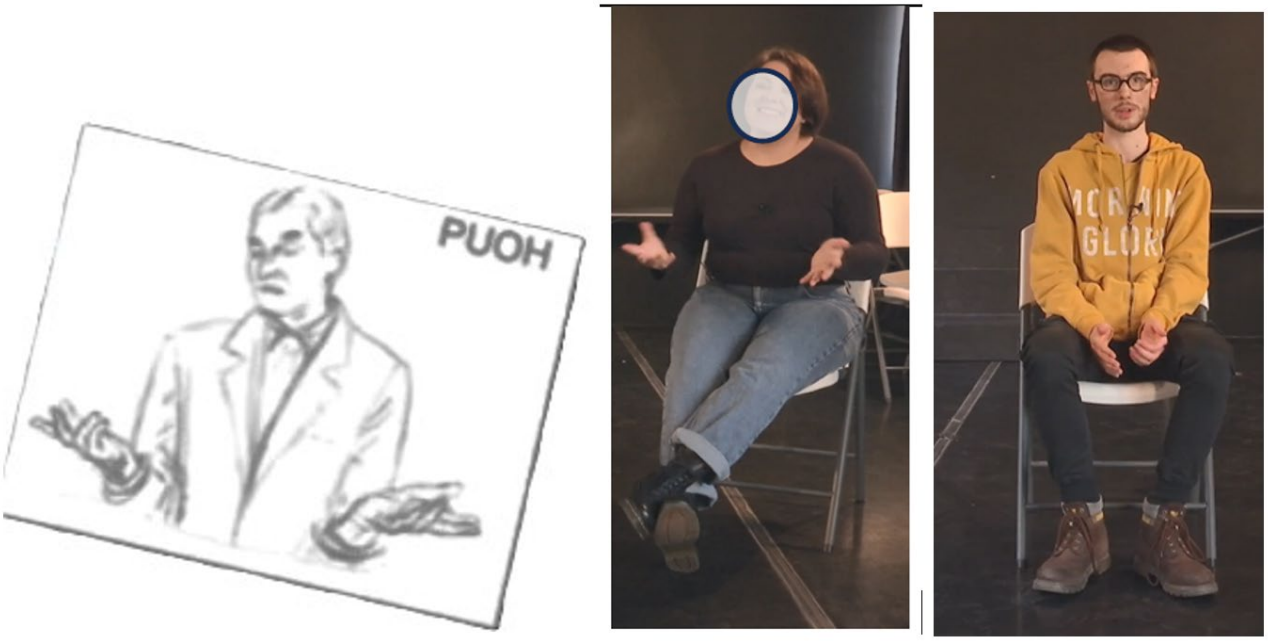


Fig.2 Palm Up Open Hand (Bressem & Muller 2014) performed by a non-autistic (left) and an autistic person (right)

Fig.3 Recurrent gesture of negation (Harrison 2009): reduced trajectory and position.

References

Baron-Cohen, S. (1995). Mindblindness: An essay on autism and theory of mind. The MIT Press.

Bascom, J. (2012). Loud hands: Autistic people, speaking. Washington DC, USA: The Autistic Press.

Bressem, J., Ladewig, S.H., Müller, C. (2013). Linguistic Annotation System for Gestures (LASG). In Müller C., Cienki A., Fricke E., *et al.* (eds.) «Body - Language - Communication: An International Handbook on Multimodality in Human Interaction (Handbooks of Linguistics and Communication Science 38)» Vol. 1. Berlin: De Gruyter Mouton, 1098–1125.

Chan, S., Khader, M., Ang, J., *et al.* (2016). To behave like a liar: Nonverbal cues to deception in an Asian sample. Journal of Police and Criminal Psychology, 31, 165-172.

Chen, R.S.Y. (2024) Bridging the gap: fostering interactive stimming between non-speaking autistic children and their parents. Front. Integr. Neurosci. 18:1374882. doi: 10.3389/fnint.2024.1374882

Cienki, A. (2021). From the finger lift to the palm-up open hand when presenting a point: A methodological exploration of forms and functions. *Languages and Modalities*, 1(1), 17–30. https://doi.org/10.3897/lamo.1.68914

Colgan, S.E., Lanter, E., McComish, C., *et al.* (2006). Analysis of social interaction gestures in infants with autism. Child Neuropsychology 12, 307–319.

Cresti, E. (2020). The pragmatic analysis of speech and its illocutionary classification according to the Language into Act Theory. In Izréel S., Mello H., Panunzi A., Raso T. (eds.) «In search of basic units of spoken language: A corpus-driven approach». Amsterdam: John Benjamins, 181-219.

Cresti, E., Moneglia, M. (2018). The illocutionary basis of information structure. Language into Act Theory (L-AcT). In Adamou, E., Haude, K., Vanhove, M. (eds.) «Information structure in lesser-described languages: Studies in prosody and syntax». 359-401. Amsterdam: John Benjamins.

Duffy, C., Healy, O. (2011). Spontaneous Communication in Autism Spectrum Disorder: A Review of Topographies and Interventions. Research in Autism Spectrum Disorders 5, 977-983.

Eigsti, I. (2013). A review of embodiment in autism spectrum disorders. Frontiers in Psychology 30; 4: 224. doi: 10.3389/fpsyg.2013.00224. PMID: 23641226; PMCID: PMC3639406.

Froiland, J.M., Davison, M.L. (2016). Home literacy, television viewing, fidgeting and ADHD in young children. Educational Psychology 36:8, 1337-1353, DOI: 10.1080/01443410.2014.963031

Happé, F. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. Cognition 48, 101-19.

Harrison, S. (2009). The temporal coordination of negation gestures in relation to speech. TiGeR 2013 - Tilburg Gesture Research Meeting At: Tilburg University, The Netherlands, 1-4.

Hobson, R.P., García-Pérez, R.M., Lee, A. (2010). Person-centered (deictic) expressions and autism. Journal of Autism and Developmental Disorders 40(4): 403–415.

Kita, S., van Gijn, I., van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In Wachsmuth, I., Fröhlich, M. (eds), «Gesture and Sign Language in Human-Computer Interaction». Berlin: Springer, 23–35.

Ladewig, S.H., Bressem, J. (2013). A linguistic perspective on the notation of gesture phases. In Müller, C., Cienki, A., Fricke, E., *et al.* (eds) «Body - Language - Communication: An International Handbook on Multimodality in Human Interaction (Handbooks of Linguistics and Communication Science 38)» Vol. 1. Berlin: De Gruyter Mouton, 1060–1079.

Ledwing, S. H. (2014) Recurrent Gestures. In Müller, C., Cienki, A., Fricke, E., *et al.* (eds) «Body - Language - Communication: An International Handbook on Multimodality in Human Interaction Vol. 2. Berlin: De Gruyter Mouton, 1558-1575.

Ledwing, S.H (2024) Recurrent Gestures: Cultural, Individual and Linguistic dimensions of Meaning-Making. In A. Cienki (ed.) *Gesture Studies*, Cambridge: Cambridge University Press, 32-55.

Leekam, S.R., Prior, M.R., Uljarevic, M. (2011). Restricted and repetitive behaviors in autism spectrum disorders: a review of research in the last decade. Psychol. Bull. 137, 562–593. doi: 10.1037/a0023341

Lewis, M.H., Bodfish, J.W. (1998). Repetitive behavior disorders in autism. Mental Retard. Develop. Disabil. Res. Rev. 4, 80–89. doi: 10.1002/ (SICI)1098-2779(1998)4:2<80::AID-MRDD4>3.0.CO;2-0

Mahmoud, M., Morency, L.P., Robinson, P. (2013). Automatic multimodal descriptors of rhythmic body movement. In Proceedings of the 15th ACM on International conference on multimodal interaction. 429-436.

de Marchena, A., Eigsti, I. (2010). Conversational Gestures in Autism Spectrum Disorders: Asynchrony but not Decreased Frequency. Autism Research 3: 311-322.

McCann, J., Peppé, S., Gibbon, F.E., *et al.* (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. Int. J. Lang. Commun. Disord. 42, 682–702. DOI: 10.1080/13682820601170102.

McNeill, D. (2005). Gesture and Thought. Chicago: University of Chicago Press. (chapters 3, 4, 5, 6)

Nolan, J., and McBride, M. (2015). Embodied semiosis: autistic 'stimming' as sensory praxis. In Trifonas, P. (ed.) «International handbook of semiotics». Dordrecht, Netherlands: Springer, 1069–1078.

Saccone, V., Cantalini, G., Moneglia, M. (2023). Prosody, gesture and self-adaptors. A case study of ASD for large corpora collection. CHIMERA. Romance Corpora and Linguistic Studies. 10, 211-235.

So, W., Wong, M.K., Lui, M., Yip, V. (2015). The development of co-speech gesture and its semantic integration with speech in 6- to 12-year-old children with autism spectrum disorders. Autism, 19(8), 956-968. DOI: 10.1177/1362361314556783

Sparaci, L., Lasorsa, F.R., Capirci, O. (2019). More Than Words: Gestures in Typically Developing Children and in Children with Autism. In Grove, N., Launonen, K. (eds.) «Manual Sign Acquisition in Children with Developmental Disabilities». NOVA Science Publishers.

Sperber, D., Wilson, D. (2002). Pragmatics, modularity, and mind-reading. Mind & Language, 17(1-2), 3–23. https://doi.org/10.1111/1468-0017.00186

Volkmar, F.R. (2020). Disturbi dello spettro autistico. III Edizione. EDRA, Milano.

Wetherby, A.M., Prutting, C.A. (1984). Profiles of communicative and cognitive-social abilities in autistic children. Journal of Speech and Hearing Research, 27, 364–377.

Investigating Linguistic and Motor Influences in Signed Second Language Production

Monica Xiao Wei[1], Diar Abdlkarim[1], [2] Sang-Hoon Yeo, Robin L. Thompson[1]
[1] School of Psychology, University of Birmingham
[2]School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham

xxw278@student.bham.ac.uk

Second language (L2) signers exhibit 'accents' when signing, reflecting the influences of prior linguistic and motor experience. While sign language accents are under-researched, a key observation is that the accents of L2 sign language learners differ depending on whether they already have a signed first language (L1) or a spoken L1 (e.g., hearing non-signers). This study examines the production of never-before-seen Hong Kong Sign Language (HKSL) signs to understand the impact of linguistic and motor skills on L2 learning.

Overview: Spoken L2 production is shaped by both phonological transfer from the L1 and L1-specific speech motor patterns (Gass et al., 2020; Chakraborty & Shanmugam, 2011). However, such motor patterns cannot transfer directly to a signed L2, which relies on a manual modality with different motor and phonological demands. Research suggests that hearing L2 signers exhibit greater kinematic variability (Hilger et al., 2015) and less precise motor control (Mirus et al., 2001) than fluent Deaf signers, possibly due to limited experience coordinating fine motor movements in 3D space or unfamiliarity with how these movements map onto sign phonology.

This study integrates linguistic (phonological features) and motor (movement kinematics) analyses to explore whether motor proficiency alone can facilitate L2 sign production or whether specific sign language knowledge is essential. We compare three groups: native BSL signers with developed L1 sign-specific motor patterns, musicians who have fine-tuned bimanual motor control (Sobierajewicz et al., 2018) and experience with hand movements in space to play their instruments, and L1 English speakers with no sign language or musical training.

Hypotheses:

1. Motor Generalisation Hypothesis: If generalised motor skills are sufficient for L2 signing, musicians and L1 signers will perform similarly, highlighting motor proficiency as a transferable modality.
2. Linguistic-Specific Hypothesis: If sign-specific knowledge underpins L2 performance, L1 signers will outperform all non-signers, regardless of their motor expertise.

This approach enables a broader investigation of the impact of L1 and tests the hypothesis that manual dexterity, even without a signed L1, can influence L2 learning.

Methodology and Analysis: Participants viewed 96 HKSL sign videos one time and were then asked to produce the signs, including facial expressions, as quickly and accurately as possible. HKSL signs were controlled across six levels of phonological complexity, balancing one- and two-handed signs, body contact, symmetry, and movement intricacy (Ortega & Morgan, 2015). Responses are analysed both linguistically (accuracy in phonological parameters; Stokoe, 1960) and motorically (kinematic measures from 2D/3D pose estimation via an opensource python library, OpenMMLab).

Preliminary results: Linguistic analyses using linear mixed-effects models (LMER) with 25 participants per group revealed a significant Group × Complexity interaction ($F(2, 6909.3) = 12.263$, $p < 0.001$). While musicians did not reach signers' performance, their fine-tuned motor skills gave them an advantage over non-musicians when producing signs of greater motoric and linguistic complexity (Figure 1). These results so far highlight that motor expertise can facilitate L2 sign learning, particularly as linguistic and motor demands increase, although sign-specific linguistic experience remains the strongest predictor of accuracy.

Further Analysis: Wrist and finger movements kinematics and Dynamic Time Warping (DTW) will be used to analyse motion data, including the velocity and number of velocity peaks of key points (e.g. wrist, fingertips). DTW provides an index of similarity between HKSL model signers and participants' movements by optimally aligning two time series for similarity analysis (Mueen & Keogh, 2016; Müller, 2007). We aim to assess motor similarity across groups, with smaller DTW distances indicating greater similarity, further exploring in motor contributions to L2 sign production.

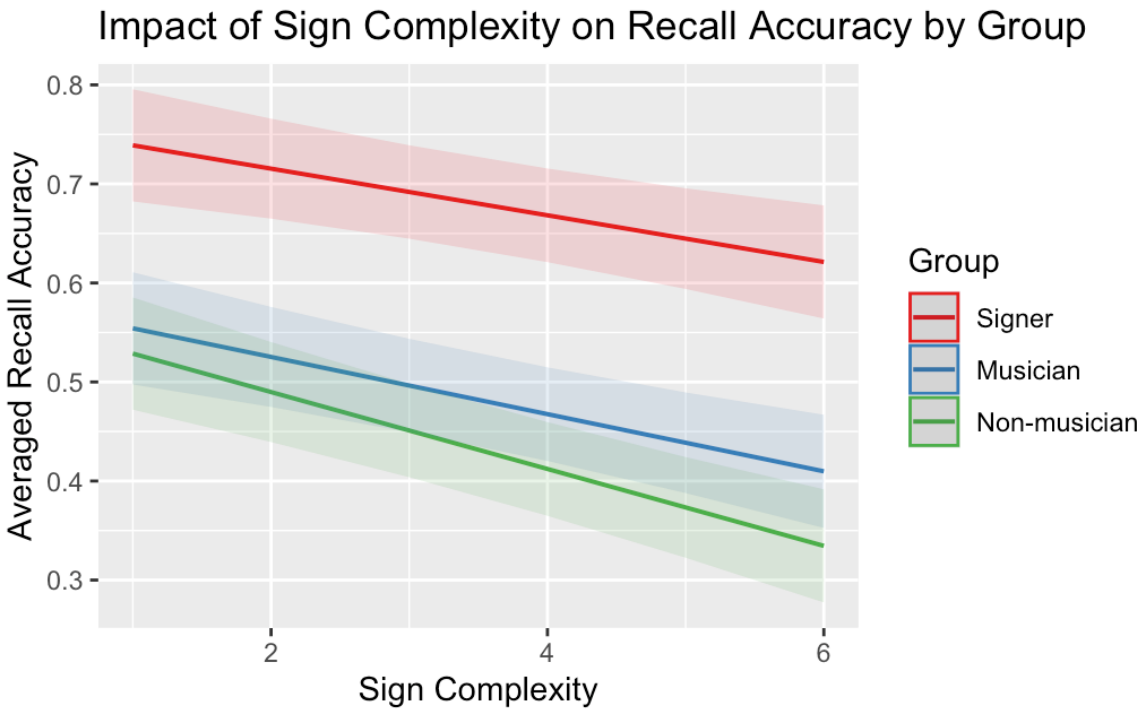


Figure 1. Articulation accuracy in different complexity among groups

References

Chakraborty, R., & Shanmugam, R. (2011). Influence of L2 proficiency on kinematic duration of single words: Real and novel word production by Bengali-English speakers. *International Journal of Speech-Language Pathology*, 13(6), 536–548. https://doi.org/10.1080/17549507.2011.595824

Gass, S. M., Behney, J., & Plonsky, L. (2020). *Second Language Acquisition: An Introductory Course* (5th ed.). Taylor & Francis Group

Hilger, A. I., Loucks, T. M., Quinto-Pozos, D., & Dye, M. W. (2015). Second language acquisition across modalities: Production variability in adult L2 learners of American Sign Language. *Second Language Research*, 31(3), 375–388. https://doi.org/10.1177/0267658315570648

Mirus, G., Rathmann, C., & Meier, R. P. (2001). Proximalization and distalization of sign movement in adult learners. In V. Dively, M. Metzger, S. Taub, & A. M. Baer (Eds.), *Signed languages: Discoveries from international research* (pp. 103–119). Gallaudet University Press

Mueen, A. K., & Keogh, E. (2016). Extracting optimal performance from dynamic time warping. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA (pp. 2129–2130). https://doi.org/10.1145/2939672.2945383

Müller, M. (2007). *Information retrieval for music and motion*. Heidelberg, Germany: Springer.

Ortega, G., & Morgan, G. (2015). Phonological Development in Hearing Learners of a Sign Language: The Influence of Phonological Parameters, Sign Complexity, and Iconicity. *Language Learning*, 65(3), 660-688. https://doi.org/10.1111/lang.12123

Sobierajewicz, J., Naskręcki, R., Jaśkowski, W., & Van Der Lubbe, R. H. J. (2018). Do musicians learn a fine sequential hand motor skill differently than non-musicians? *PLOS ONE*, 13(11), e0207449. https://doi.org/10.1371/journal.pone.0207449

Spencer, K. A., & Rogers, M. A. (2005). Speech motor programming in hypokinetic and ataxic dysarthria. *Brain and Language*, 94, 347–366

Stokoe, W. (1960) *Sign language structure: An outline of the visual communication systems of the American Deaf*. (1993 Reprint ed.). Silver Spri

POSTERS

Children's multimodal feedback practices in interaction: Insights from three diverse languages

Anastasia Bauer[1], Alice Mitchell[2], Birgit Träuble[3], Pamela Perniss[4] & Birgit Hellwig[1]

[1]Department of Linguistics, General Linguistics, Faculty of Arts and Humanities, University of Cologne, Cologne, Germany.
[2]African Studies, Faculty of Arts and Humanities, University of Cologne, Germany.
[3]Department of Psychology, Faculty of Human Sciences, University of Cologne, Cologne, Germany.
[4]Department of Rehabilitation and Special Education, Faculty of Human Sciences, University of Cologne, Cologne, Germany.

anastasia.bauer@uni-koeln.de

Feedback occupies a central position in the field of language acquisition, where it is understood to facilitate language learning. While previous research has predominantly focused on how adults, particularly mothers, orient to and correct children's speech errors (Clark, 2020), feedback signals serve a much broader range of functions in interaction. Adult listeners are understood to be active participants in conversation who co-construct talk through their response tokens (Gardner 2001). However, the ways in which children acquire feedback-giving practices to signal their active participation in interaction remains an open question in developmental pragmatics, with limited research addressing this topic (Tykkyläinen 2010).

To address this gap, this paper investigates children's use of multimodal feedback signals in video corpora recorded in naturalistic contexts representing three very different languages and linguacultures: Datooga (Tanzania); Qaqet (Papua New Guinea); and American English (USA, CHILDES talkbank). Prior studies of adult interaction show that feedback signaling the absence of communicative trouble can take various forms, including non-lexical vocalizations e.g., 'mm', lexical acknowledgments e.g. 'yes', manual lexical signs in sign languages, gestures, head nods, body posture, facial expressions, smiles and laughter, or even long blinks (Yngve 1970; Brunner 1979; Allwood 1992; Dingemanse et al. 2022; Bauer et al. submitted). Conversely, signals indicating trouble may include frowns, raising/lowering of eyebrows, head forward movements, gaze shifts, freeze-look, vocalizations (e.g., 'uh?'), or repair initiators (e.g., 'what?') and combinations thereof (Kendrick 2015; Dingemanse 2015, Manrique 2016, Skedsmo 2020). This study examines which of these feedback formats children employ, focusing on embodied practices, and how children's feedback compares to that of adults in interaction.

Our findings reveal that young children (3-6) have a relatively high tolerance for non-progressivity and break-downs in interaction. They appear to make fewer explicit attempts to achieve shared understanding or to indicate that they are monitoring their interlocutor's speech. Nonetheless, children across our sample, do provide feedback: they assist with word searches, employ minimal repair initiators and produce extended repair formats. Additionally, children occasionally orient to other children's non-target-like forms, thus providing metalinguistic feedback. However, preliminary observations indicate that young children use fewer embodied practices and less non-repair feedback—such as continuers, acknowledgments, newsmarks, and assessments—than adults (Gardner 2001; Bauer et al. under review) or school-age children (6-12) in comparable interactions (Bodur et al. 2023).

We highlight this research gap in the developmental literature and provide two speculative explanations for the observed patterns and their implications for cognitive development and language acquisition. First, integrating insights from developmental psychology, we propose that the acquisition of non-repair feedback aligns with the development of Theory of Mind, as these practices

require a nuanced understanding of others' mental states (de Villiers 2007). Second, we tentatively suggest that the organization of children's interaction differs from adults' in ways that makes non-repair feedback less important, resulting in its later acquisition. The limited use of embodied practices in children's feedback production hints at the progressive development of multimodal languaging in children (Cartmill et al. 2012; Özyürek 2018; Morgenstern 2022; Capirci et al. 2022; Arslan et al. 2023; Karadöller et al. 2024). These findings contribute to our understanding of cognitive development and language acquisition, shedding light on the development of multimodal interactional mechanisms in childhood.

References

Allwood, J. (1992). On the semantics and pragmatics of linguistic feedback. *Journal of Semantics, 9*(1), 1–26. https://doi.org/10.1093/jos/9.1.1

Arslan, B., Aktan-Erciyes, A., & Göksun, T. (2023). Multimodal language in bilingual and monolingual children: Gesture production and speech disfluency. *Bilingualism: Language and Cognition, 26*(5), 971–983. https://doi.org/10.1017/S1366728923000196

Bauer, A., Gipper, S., Hosemann, J., & Herrmann, T.-A. (under review). Rethinking linguistic feedback: A modality-agnostic and holistic approach to multimodal addressee signals in signed and spoken dyadic interaction.

Bodur, K., Nikolaus, M., Prévot, L., & Fourtassi, A. (2023). Using video calls to study children's conversational development: The case of backchannel signaling. *Frontiers in Computer Science, 5*, 1088752. https://doi.org/10.3389/fcomp.2023.1088752

Brunner, L. J. (1979). Smiles can be back channels. *Journal of Personality and Social Psychology, 37*(5), 728–734. https://doi.org/10.1037/0022-3514.37.5.728

Cartmill, E. A., Demir, Ö. E., & Goldin-Meadow, S. (2012). Studying gesture. In *Research Methods in Child Language*(pp. 208–225).

Capirci, O., Caselli, M. C., & Volterra, V. (2022). Interaction among modalities and within development. In A. Morgenstern & S. Goldin-Meadow (Eds.), *Gesture in language: Development across the lifespan* (pp. 113–133). De Gruyter Mouton; American Psychological Association. https://doi.org/10.1037/0000269-005

Clark, E. V. (2020). Conversational repair and the acquisition of language. *Discourse Processes, 57*(5–6), 441–459. https://doi.org/10.1080/0163853X.2020.1719795

de Villiers, J. (2007). The interface of language and Theory of Mind. *Lingua, 117*(11), 1858–1878. https://doi.org/10.1016/j.lingua.2006.11.006

Dingemanse, M., Liesenfeld, A., & Woensdregt, M. (2022). Convergent cultural evolution of continuers (mmhm). In *The Evolution of Language: Proceedings of the Joint Conference on Language Evolution (JCoLE)* (pp. 61–67).

Dingemanse, M. (2015). Other-initiated repair in Siwu. *Open Linguistics, 1*(1). https://doi.org/10.1515/opli-2015-0001

Dingemanse, M. (2022). Language and the infrastructure of social interaction. *Annual Review of Linguistics, 8*, 241–261. https://doi.org/10.1146/annurev-linguistics-031821-095740

Gardner, R. (2001). *When listeners talk: Response tokens and listener stance*. John Benjamins Publishing.

Karadöller, D. Z., Sümer, B., & Özyürek, A. (2024). First-language acquisition in a multimodal language framework: Insights from speech, gesture, and sign. *First Language*. https://doi.org/10.1177/01427237241290678

Kendrick, K. H. (2015). The intersection of turn-taking and repair: The timing of other-initiations of repair in conversation. *Frontiers in Psychology, 6*, 250. https://doi.org/10.3389/fpsyg.2015.00250

Manrique, E. (2016). Other-initiated repair in Argentine Sign Language. *Open Linguistics, 2*(1). https://doi.org/10.1515/opli-2016-0001

Morgenstern, A. (2023). Children's multimodal language development from an interactional, usage-based, and cognitive perspective. *WIREs Cognitive Science, 14*(2), e1631. https://doi.org/10.1002/wcs.1631

Özyürek, A. (2018). Cross-linguistic variation in children's multimodal utterances. In M. Hickmann, E. Veneziano, & H. Jisa (Eds.), *Sources of variation in first language acquisition: Languages, contexts, and learners* (pp. 123–138). Amsterdam: Benjamins.

Skedsmo, K. (2020). Other-initiations of repair in Norwegian Sign Language. *Social Interaction - Video-Based Studies of Human Sociality, 3*(2). https://doi.org/10.7146/si.v3i2.117723

Multimodal Language in the Foreign Language Classroom
An Analysis of Data from Empirical Research

Moira de Iaco

Università degli Studi di Bari Aldo Moro

moira.deiaco@uniba.it

The starting point of this proposal is the hypothesis that gestures and body movements spontaneously play an essential role in the context of language education, since they are an integral part of the human communication system, as it has been demonstrated in several psycholinguistic studies (McNeill 1992, 2005; Krauss 1998; Kendon 2004; Capirci, Contaldo, Caselli, Volterra 2005). The main aim is to substantiate this hypothesis by analysing data from ongoing empirical research and to argue the need to raise awareness of the potential of multimodality among language teachers and students and to promote kinaesthetic learning.

The theoretical framework of this contribution is derived from the theories of Embodied Cognition in combination with the Total Physical Response approach (Hung, Fang, Chen 2014; Kuo, Fang, and Chen 2014). The latter considers the union of physical and verbal answers with the use of student's body movements to prepare and learn verbal expressions during the language acquisition process. It encourages the physical involvement of the learner (Asher 1969). This approach can be integrated into Embodied Cognition studies to gain a broader perspective on the cognitive role of the body's influence on the mind in learning and to include the effects of the teacher's bodily actions during instruction on learning (Paloma, Ascione, Tafuri 2016; Stolz [ed.] 2021). Body actions and physical experience affect mental processes, and mental processes leave traces in the body. Body and mind interact with each other as two distinct, but not separate, independent systems (Barsalou 2008; Gibbs 2005). The result of this interaction is that we think, speak, and learn with the body as well as the mind or even we think, speak, and learn through the influence of the body on the mind and the influence of the mind on the body.

Gestures and body movements are fundamental to cognitive and communicative functions such as conceptualisation, thought organisation, speech production and comprehension, development and management of emotions and pragmatic aspects (Tellier 2006; Swellera, Shinooka-Phelana, & Austin 2020; García-Gámez, Cervilla, Casado, & Macizo 2021; Goldin-Meadow 2023). The use of gestures and body movements by language teachers and learners corresponds to each of the cognitive and communicative functions mentioned above, as this paper aims to demonstrate by analysing data from an empirical study. It is a study in which thirty foreign language teachers who agreed to participate in the study answered a questionnaire about their teaching experience. Questions focused on the perceptions of frequency, the impact on teaching and learning by students, and the composition of gestures and body movements they use in class. It turns out that teachers use gestures both consciously and unconsciously, especially to explain and illustrate meanings, convey contents and give instructions for activities, represent pragmatic and sociolinguistic elements. Gestures and body movements make it possible to promote comprehension, memory, and attention, avoid verbalisation and translation into learners' first language, support student participation in the lesson, convey cultural aspects, manage discomfort and anxiety. Examples of gestures and body movements used by the teachers involved in the study were included in the analysis of the data. Consequences for language education are derived from this.

References

J. J. Asher, "The Total Physical Response Approach to Second Language Learning", in The Modern Language Journal, 53, 1, 1969: 3-17

L. W. Barsalou, Grounded Cognition. Annu. Rev. Psychol, 59, 617–645, 2008.

O. Capirci, A. Contaldo, M. C. Caselli, V. Volterra, “From Action to Language through Gesture: A Longitudinal Perspective,” Gesture 2005, 5, 155–177
K. J. Chao, H. W. Huang, W. C. Fang, and N. S. Chen, “Embodied play to learn: Exploring Kinect-facilitated memory performance,” British Journal of Educational Technology, 44(5), E151-E155, 2013.
N. S. Chen, and W. C. Fang, “Gesture-Based Technologies for Enhancing Learning New Developments in Technology Enhanced Learning: Concept, Research and Best Practices,” Springer’s Lecture Notes in Educational Technology, 2014.
A. B. García-Gámez, Ó. Cervilla, A. Casado, & P. Macizo, “Seeing or Acting? The Effect of Performing Gestures on Foreign Language Vocabulary Learning,” Language Teaching Research, 2021: 1–32.
S. Goldin-Meadow, Thinking With Your Hands: The Surprising Science Behind How Gestures Shape Our Thoughts. New York: Basic Books, 2023.
R. W. Gibbs, Embodiment and cognitive science. Cambridge: Cambridge University, 2008.
I. C. Hung, L. L. Lin, W. C. Fang, and N. S. Chen, “Learning with the body: An embodiment-based learning strategy enhances performance of comprehending fundamental optics,” Interacting with Computers, 2014.
A. Kendon, Gesture: visible action as utterance, Cambridge: Cambridge University Press, 2004.
R. M. Krauss, “Why do we gesture when we speak?” Current Direction in Psychological Science, 7, 54, 1988.
F. R. Kuo, C. C. Hsu., W. C. Fang, and N. S. Chen, “The effects of Embodiment-based TPR approach on student English vocabulary learning achievement,” Journal of King Saudi University - Computer and Information Sciences, 26, 63-70, 2014.
D. McNeill, Hand, and Mind: What Gestures reveal about thought, [M] Chicago: The University of Chicago Press, 1992.
D. McNeill, Gesture & thought, [M] Chicago: The University of Chicago Press, 2005.
F. G. Paloma, A. Ascione, D. Tafuri, “Embodied Cognition: il ruolo del corpo della didattica,” Formazione & Insegnamento, XIV (1), 2016: 75-87.
S. A. Stolz [ed.], The Body, Embodiment, and Education. An Interdisciplinary Approach. Taylor & Francis, 2021.
N. Swellera, A. Shinooka-Phelana, & E. Austin, The effects of observing and producing gestures on Japanese word learning. Acta Psychologica, 207, 2020.
M. Tellier, “Cognitive linguistics and second language learning: Theoretical basics and experimental evidence,” Language Teaching, 39(4), 2008: 207-235.

In the light on constructed action: corpus-cases of *signs with two faces*

Noora Hoffrén
University of Jyväskylä, Finland

notehoff@jyu.fi

In this presentation, I question: if we consider constructed action (CA) as being part of some sign's inherent property, is it possible to re-analyse these signs as *signs with two faces*? I will give one to three examples of these signs, selected from a large, multimodal Finnish sign language (FinSL) corpus data and annotated with ELAN to address this question.

Signs with two faces are iconic sign languages signs, which may manifest as descriptive, lexical signs in one context and as depictive, partly or non-lexical in another (e.g. Ferrara and Halvorsen 2017; Johnston and Ferrara 2012). For example, as Ferrara and Halvrosen (2017) presents from Norwegian's sign language (NTS) corpus data, a sign SWIM (similar to FinSL sign SWIM) possesses this inherent property of dual potential and can be signed in context mediating both sides. This dual-reading of the sign SWIM rises from the non-manual cues, such as the word-like mouthing (descriptive) and elements expressed with face and head (depictive) (Ferrara and Halvorsen 2017:382–83).

The dual potential of iconic, lexicalized signs are widely recognized in the literature (Cormier et al. 2012; Zeshan 2003) among others). Interestingly, most of the researchers aim to make a distinction between the ones depicting and the ones describing (e.g. Cormier et al. 2012; Dudis 2011), even though the fuzzy boundaries are well acknowledged (e.g. Johnston 2013) and signs are known to lexicalize and de-lexicalize (e.g. Jantunen 2018; Johnston and Ferrara 2012). The gradience and detailed analysis of CA has brought a new perspective to this matter.

CA is *showing something by enacting it,* with different articulators such as the head, gaze, hands and torso. When doing CA, the signer (or the speaker) takes a role of another character and enacts this character's thoughts, feelings, sayings or actions from this point of view. CA is gradient and subtly changing, depending on the amount and the overtness of used articulators and it can be roughly divided to overt, reduced and subtle CA (Cormier et al. 2012).

The connection between CA and certain signs are found so strong, that it has been suggested, that CA and these *signs incorporating embodiment* should be analysed as a "continuum of lexicalization of embodiment". This continuum includes three prototypes of sign languages signs: 1. signs incorporating embodiment (lexical), 2. handling depicting signs (partly-lexical) and 3. CA-gestures (non-lexical). The first ones mentioned are fixed lexical signs, such as SWIM presented earlier in this paper (when signed as describing the action). The last mentioned is an ad hod non-lexical gesture, purely *showing* the actions or postures of the character (i.e. CVPT-gestures as in McNeill 1992). Handling depicting signs are in-between cases, with some fixed parts of the sign and some purely context-dependent parts. (Cormier et al. 2012).

To identify different types of signs along this continuum, the authors suggest, that the overtness of CA helps to identify the lexical degree of the sign. Instead of using the continuum only for this purpose, I question: if a sign incorporating embodiment can be lexicalized from instance of CA and de-lexicalize through the use of CA, should these signs be considered as *signs with two faces*? The presentation will focus on the form and the meaning of the example-signs, concentrating on the different parameters and articulators (e.g. torso and gaze) and their participation in the meaning-making and taking CA into account. With this presentation, I aim to participate the discussion about the multimodal model of language, including such perspectives as depiction and gesture as being part of language (following Kendon 2004 among others).

References

Cormier, Kearsy, David Quinto-Pozos, Zed Sevcikova, and Adam Schembri. 2012. 'Lexicalisation and De-Lexicalisation Processes in Sign Languages: Comparing Depicting Constructions and Viewpoint Gestures'. *Language & Communication* 32(4):329–48. doi: 10.1016/j.langcom.2012.09.004.

Dudis, Paul. 2011. 'The Body in Scene Depictions'. Pp. 3–45 in *Discourse in Signed Languages*, edited by C. Roy B.

Ferrara, Lindsay, and Rolf Piene Halvorsen. 2017. 'Depicting and Describing Meanings with Iconic Signs in Norwegian Sign Language'. *Gesture* 16(3):371–95. doi: 10.1075/gest.00001.fer.

Jantunen, Tommi. 2018. 'Viittomakielet hybridisysteemeinä - hämärärajaisuus ja epäkonventionaalisuus osana viittomakielten rakennetta'. *Puhe ja kieli* (3):109–26. doi: 10.23997/pk.77382.

Johnston, Trevor, and Lindsay Ferrara. 2012. 'Lexicalization In Signed Languages—When Is An Idiom Not An Idiom?' Pp. 229–48 in *Selected Papers from UK-CLA Meetings, Vol 1*, edited by G. Rundbland, A. Tytus, O. Knapton, and C. Tang.

Kendon, Adam. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

McNeill, David. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago ; London: University of Chicago Press.

Zeshan, Ulrike. 2003. '"Classificatory" Constructions in Indo-Pakistani Sign Language: Grammaticalization and Lexicalization Processes'. Pp. 113–42 in *Perspectives on Classifier Constructions in Sign Languages*, edited by K. Emmorey. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Supporting Sign Language Acquisition of Deaf Children with Hearing Parents: An Intervention Study

Dilay Z. Karadöller[1,2], Maharrem Ayar[3], Hilal Özdemir[1]
[1] Middle East Technical University, Türkiye
[2] Max Planck Institute for Psycholinguistics, The Netherlands
[3] University of California, Davis, United States of America

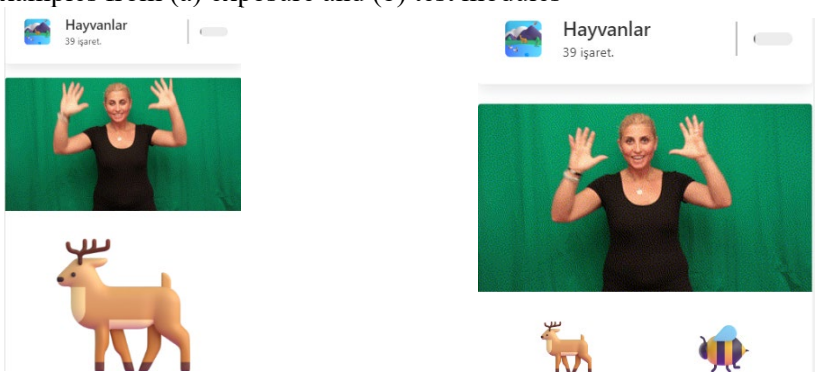
Most deaf children (95%) are born to hearing parents (Michael & Karchmer, 2004). These children often do not have access to sign or spoken language even with hearing aids, children may not get adequate linguistic input (Hall et al., 2019). In such cases, these children (i.e., late signers) acquire sign language late, possibly until they enroll in a deaf school. By contrast, deaf children with deaf parents (i.e., native signers) learn sign language from birth onwards from their deaf parents (e.g., Lillo-Martin & Henner, 2021). Previous studies have shown detrimental effects of late sign language exposure when late signers are compared to native signers on several linguistic capabilities such as grammatical judgment (Boudreault & Mayberry, 2006), mean length of utterance (Berk et al., 2012; Cheng & Mayberry, 2019), and spatial descriptions (Karadöller et al., 2017, 2021, 2013). Less is known about how late-signing children acquire lexical signs upon their first exposure and whether lexical properties of the signs such as iconicity, that is the resemblance between the sign form and what it refers to in the real space (Perniss et al., 2010) and phonological complexity (Ortega & Morgan, 2015), influence this acquisition process. Here, we investigate the lexical development of late-signing children via an intervention study over a mobile-compatible web app that provides a consistent and accessible platform to teach early acquired lexical signs in Turkish Sign Language (Türk İşaret Dili, TİD).

Participants consist of deaf children who acquired sign language late (currently n = 12; Mean Age in months = 83,67; Age Range in months = 54-114). The intervention is carried out via an application developed to teach lexical signs in TİD, consisting of 14 semantic categories (e.g., animals, food, vehicles). For each category, the app has exposure and test modules (see Figure 1 for examples). In the exposure module, children saw a video of a sign and an image corresponding to that sign. In the test module, children saw a video of a sign and two images from the same category, and then, they were expected to choose the correct answer. Videos used in the app were obtained from TİD-CDI (Sümer et al., 2017). At the start of the intervention, children received a general test that measured their baseline accuracy. This task was also repeated at the end of the intervention. Iconicity and phonological complexity ratings of the signs were determined as part of a bigger project launched to create a lexical database for TİD. Iconicity ratings were obtained from hearing Turkish speakers on a 7-point Likert scale based on Caselli and colleagues (2017). Phonological complexity ratings were calculated following Ortega and Morgan (2015).

Results of the glmer models on the preliminary data based on eleven children showed that children significantly developed in their TİD knowledge after 8-week intervention ($p < .001$). This development is more pronounced for learning signs that are high in iconicity ($p < .001$) and irrespective of the phonological complexity of the signs ($p > .05$).

Summarizing, these results showed that children can develop their TİD vocabulary when they are provided with suitable intervention programs. Moreover, their vocabulary expansion can be facilitated for signs that are high in iconicity. However, the phonological complexity of the signs does not facilitate or hinder this acquisition process. Overall, our findings corroborate evidence for iconicity's effect on language acquisition (Caselli & Pyers, 2017; Ortega & Morgan, 2015; Pernis et al., 2010; Sümer et al., 2017; Thompson et al., 2012) and call for multiple practices to help establish alternative platforms of exposure for late-signing children to facilitate their language acquisition (Hall et al., 2019; Lillo-Martin & Henner, 2021).

Figure 1
Examples from (a) exposure and (b) test modules



References

Berk, S., & Lillo-Martin, D. (2012). The two-word stage: Motivated by linguistic or cognitive constraints? *Cognitive Psychology, 65*(1), 118-140.

Boudreault, P., & Mayberry, R. I. (2006). Grammatical processing in American Sign Language: Age of first language acquisition effects in relation to syntactic structure. *Language and Cognitive Processes, 21*(5), 608-635.

Caselli, N. K., & Pyers, J. E. (2017). The road to language learning is not entirely iconic: Iconicity, neighborhood density, and frequency facilitate acquisition of sign language. Psychological science, 28(7), 979-987.

Caselli, N., Sevcikova Sehyr, Z., Cohen-Goldberg, A. M., & Emmorey, K. (2017). ASL-LEX: A lexical database of American Sign Language. Behavior Research Methods, 49(2), 784-801. doi:10.3758/s13428-016-0742-0

Cheng, Q., & Mayberry, R. I. (2019). Acquiring a first language in adolescence: the case of basic word order in American Sign Language. *Journal of Child Language, 46*(2), 214-240.

Hall, M. L., Hall, W. C., & Caselli, N. K. (2019). Deaf children need language, not (just) speech. *First Language, 39*(4), 367-395.

Karadöller, D. Z., Sümer, B. & Özyürek, A. (2021). Effects and Non-Effects of Late Language Exposure on Spatial Language Development: Evidence from Deaf Adults and Children. *Language Learning and Development, 17*, 1-25.

Karadöller, D. Z., Sümer, B., & Özyürek, A. (2017). Effects of delayed language exposure on spatial language acquisition by signing children and adults. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society (CogSci 2017)* (pp. 2372-2376). Austin, TX: Cognitive Science Society.

Karadöller, D. Z., Sümer, B., Ünal, E. & Özyürek, A. (2023). Late sign language exposure does not modulate the relation between spatial language and spatial memory in deaf children and adults. *Memory and Cognition, 51*, 582–600.

Lillo-Martin, D., & Henner, J. (2021). Acquisition of sign languages. *Annual Review of Linguistics, 7*, 395-419.

Mitchell, R. E., & Karchmer, M. (2004). Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the United States. *Sign Language Studies, 4*(2), 12 138-163.

Ortega, G., & Morgan, G. (2015). Phonological development in hearing learners of a sign language: The influence of phonological parameters, sign complexity, and iconicity. *Language Learning, 65*(3), 660-688.

Perniss, P., Thompson, R. L., and Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology*, 1-15.

Sümer, B., Grabitz, C., & Küntay, A. (2017). Early produced signs are iconic: Evidence from Turkish Sign Language. In the 39th Annual Conference of the Cognitive Science Society (CogSci 2017) (pp. 3273-3278). Cognitive Science Society.

Thompson, R. L., Vinson, D. P., Woll, B., & Vigliocco, G. (2012). The road to language learning is iconic: Evidence from British Sign Language. Psychological science, 23(12), 1443-1448

Multimodal language origins:
empirical and theoretical insights from animal communication studies

Costanza Norci

Università degli Studi di Roma – La sapienza

costanza.norci@uniroma1.it

The study of animal communication, especially that of our closest relatives (gorillas, chimpanzees, bonobos and orangutangs), influences and is in turn influenced by research on the evolution of human language. A striking example of this mutual entanglement is offered by the origins of the 'gesture-first' hypothesis (Kendon 2017): already put forward in the 18th and the 19th centuries, the idea of a gestural origin of human language gained new appeal in the light of both the negative and the positive results of two empirical studies, conducted in the 1940s and in the 1960s respectively. The first one was the failure of Hayes and Hayes (1952) to teach a chimpanzee to speak; the second one was the success of Gardner and Gardner (1969) in teaching Washoe, another chimpanzee, to use the symbolic gestures of human signed language. After Washoe, the gesture-first hypothesis outlined by Hewes (1973) became and remains one of the most influential theories of language evolution (supported by e.g. Corballis 2004; Tomasello 2008). One line of research into the evolutionary precursors of human language has thus focused on the intentional and flexible learning and use of gestures by non-human primates – mostly within an empirical framework derived from comparative psychology.

An alternative perspective was offered (albeit more controversially) by the evidence from studies of monkey vocalizations: the research conducted by Seyfarth et al. (1980) on vervet monkeys alarm calls radically discredited a purely motivational and unvoluntary view of primate vocalizations, which instead appeared not only to have symbolic and referential properties, but also to be intentionally produced (Dennett 1983). However, the actual homology of alarm calls with human language was soon questioned, especially with respect to their intentional character, leading to the formulation of the framework of 'functional referentiality' (Macedonia and Evans 1993), which allowed the referential properties of non-human calls to be described without appealing to complex cognitive mechanisms underlying signal production and comprehension – thus retracting the idea of evolutionary continuity between language and primate vocal communication. Relying on neuro-physiological studies of the unvoluntary and biologically fixed nature of monkey vocalizations (e.g. Jürgens 2002), scholars have therefore claimed that homologies with human language should be sought in more cognitively demanding behaviors – for example, our closest relatives' ability to attribute meaning by integrating vocal signals and contextual cues (Wheeler and Fischer 2012) – rather than in the functionally referential properties of signallers' calls. The lack of a unified approach and methodology to the different domains of ape communication has led not only to a supposedly limited understanding of their communicative abilities, but also to what Liebal and colleagues (2022), advocating for a multimodal approach to non-human communication, call a 'void' in the knowledge of the phylogenetic origins of language.

In this presentation, I first outline how a multimodal understanding of human language (Kendon 2004) could benefit animal – and primate – communication studies in developing a theoretical framework within which the different domains of non-human semiotic systems can be held together. Second, I examine how in turn evidence of vocal learning in apes (Lameira 2017) and intentionally referential vocal signals (Crockford et al. 2012; Schel et al. 2013; Townsend et al. 2017) support the claim that both the gestural and the vocal domains exhibit evolutionary precursors of human language (Fröhlich et al. 2019). I hope to highlight how animal communication studies provide fruitful theoretical and empirical insights for the hypothesis of a multimodal evolution of human language (Levinson and Holler 2014), with respect to both proximate (Doherty et al. 2023) and

ultimate questions (Partan and Marler 2005; Fröhlich et al. 2018), as well as to its phylogenetic origins (Dediu and Levinson 2018).

References

Corballis, M. 2002, *From hand to mouth: the origins of language*, Princeton University Press, Princeton, New Jersey.

Crockford, C. et al. 2012, "Wild chimpanzee inform ignorant group members of danger", *Current Biology* 22, 142-146.

Dediu, D. & Levinson, S. C. 2018, "Neanderthal language revisited: not only us", *Current Opinion in Behavioral Science* 21, 49-55.

Dennett, D. 1983, "Intentional systems in cognitive ethology: the 'Panglossian paradigm' defended", *The Behavioral and Brain Sciences* 6, 343-390.

Doherty, E. et al. 2023, "Multimodal communication development in semiwild chimpanzees", *Animal Behavior* 201, 175-190.

Fröhlich, M. & van Schaik, C. P. 2018, "The function of primate multimodal communication", *Animal Cognition* 21, 619-629.

Fröhlich, M. et al. 2019, "Multimodal communication and language origins: integrating gestures and vocalizations", *Biological Reviews* 94, 1809-1829.

Gardner, R. A. & Gardner, B. T. 1969, "Teaching sign language to a chimpanzee", *Science* 165 (3894), 664-672.

Hayes, K. J. & Hayes, C. 1952, "Imitation in a home-raised chimpanzee", *Journal of Comparative and Physiological Psychology* 45, 450-459.

Hewes, G. W. 1973, "Primate communication and the gestural origins of language", *Current Anthropology* 14, 5-24.

Jürgens U. 2002, "Neural pathways underlying vocal control", *Neuroscience and Biobehavioral Reviews* 26, 235-258.

Kendon, A. 2004, *Gesture: visible action as utterance*, Cambridge University Press, Cambridge.

— 2017, "Reflections on the 'gesture-first' hypothesis of language origins", *Psychonomic Bulletin & Review* 24, 163-170.

Lameira, A. R. 2017, "Bidding evidence for primate vocal learning and the cultural substrates for speech evolution", *Neuroscience and Biobehavioral Reviews* 83, 429-439.

Levinson, S. C. & Holler J. 2014, "The origin of human multi-modal communication", *Philosophical Transactions of the Royal Society B* 369, 20130302, 1-9.

Liebal, K. et al. 2022, "The language void 10 years on: multimodal primate communication research is still uncommon", *Ethology, Ecology & Evolution* 34 (3), 274-287.

Macedonia, J. M. & Evans C. S. 1993, "Variation among mammalian alarm call systems and the problem of meaning in animal signals", *Ethology* 93, 177-197.

Partan, S. R. & Marler, P. 2005, "Issues in the classification of multimodal communication signals", *The American Naturalist* 166 (2), 231-245.

Schel, A. M. et al. 2013, "Chimpanzee alarm call production meets key criteria for intentionality", *PLoS One* 8, e76674, 1-11.

Seyfarth, R. M. et al. 1980, "Monkey responses to three different alarm calls: evidence for predator classification and semantic communication", *Science* 210, 801-803.

Tomasello, M. 2008, *Origins of human communication*, MIT Press, Cambridge.

Townsend, S. W. et al. 2017, "Exorcising Grice's ghost: an empirical approach to studying intentional communication in animals", *Biological Reviews* 92, 1427-1433.

Wheeler, B. C. & Fischer, J. 2012, "Functionally referential signals: a promising paradigm whose time has passed", *Evolutionary Anthropology* 21, 195-205.

Languaging-gestures:
A Vygotskian perspective for understanding multimodal languaging for young L2 Learners

Alessandro Rosborough, Lindsay Anderson

Brigham Young University

alex_rosborough@byu.edu

In this study, we look at the concept of languaging as co-speech or silent-gesture which establishes joint-presence that promotes meaning-making through mediation filled exchanges. At issue are current challenges in elementary curriculums that place a macro focus on correct answer (i.e., logocentric) results. This qualitative study focuses on how deictic and iconic gestures in relation to the concept of languaging, enhances the process of content learning for young second language (L2) learners. Using Vygotskian (1987) Sociocultural Theory, languaging can be viewed as a dialectical unity (Lantolf & Poehner, 2014) where speech-thinking includes gesture and other multimodal forms of communication. Especially for children, gesture both materializes and carries concrete signs and perceptions in their environment. Speech-thinking, including gesture, carries actions and reflections that can demonstrate some forms of planning, problem solving, change, and eventual transformation.

Languaging (Swain, 2006; Swain & Watanabe, 2013), is a source for L2 learning and an authentic and initial form of mediation in speaking and writing. Furthermore, it can be viewed as collaborative dialogic discourse where participants (i.e., teachers and students) use personal, emotive, and narrative-based experiences to enhance meaning and mediate their own thinking. However, almost all studies view languaging as functioning through oral and written modalities with less emphasis on gesture and multimodality (Li et al., 2023).

This study attempts to address the challenge of understanding the phenomenon of languaging and gesture for pedagogical purposes, through the study of catchments and communicative dynamism (McNeill, 1992), contingencies (McCafferty & Rosborough, 2023), and forward-oriented gestures (Kendon, 2004; Streeck, 2009). We combine the characteristics of these terms and call them *languaging-gestures*. We define these as co-speech or silent-speech gestures that carry propositional implicatures that [can] activate meaning-making and sense-making between participants to better support young students.

For this workshop, we combined our languaging research questions into the following:

1. What does langauging-gestures look like and how do they mediate understanding for young L2 students?

The data for this study are based on observation, video recording, analysis of interactions and dialogues, and a follow-up interview with the teacher. The settings were a variety of Title 1 schools in mountainous western United States. Data collection covered multiple grades (e.g., 2nd, 3rd, and 4th) with teachers having a variety of multiple years of experience working with second language learners (i.e., English learners). Depending on the classroom, researchers used multiple cameras to capture speech and gesture as naturally occurring between the teacher(s) and students. Gestures were coded using an abbreviated form of McNeill's (1992) coding scheme. Gesture phase and strokes were given primacy as carrying the central meanings in the classroom conversation exchanges. Definition of languaging-gestures (Kendon, 2004; McNeill, 1992; McCafferty & Rosborough, 2023; Streeck, 2009) were analyzed for how multimodal communication and thinking were extended as mediational tools for learning the new language.

Findings from this study presents how languaging-gestures contain forward-orientation and contingency characteristics which actively promoted classroom discourse towards extensions and novel answers. Additionally, findings demonstrated how students did not necessarily follow the teacher's discourse patterns and minimalist style of communicative exchange (Mehan, 1979). Instead,

students provided semantic and pragmatic features not readily available through classroom discourse as demonstrated by the teacher alone.

Conclusions and implications include how languaging-gestures added to new meaning-making acts that supported young L2 students in understanding new concepts in math, science, or language time. Additionally, languaging-gestures can be viewed as novel models produced by students demonstrating how they understand new concepts. Implications include how teachers can read the students' contextualizations and holistic meaning-making, which can support teachers in how they can shift the classroom to better enhance and match their students' mediational needs.

References

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

Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperative. Vygotskian praxis and the research/practice divide.* New York: Routledge.

Li, C., Liu, B., & Wang, X. (2023). A scoping review of research on languaging in second language education. *Heliyon*, *9*(5), e16300–e16300. https://doi.org/10.1016/j.heliyon.2023.e16300

Mehan, H. (1979). Learning lessons: social organization in the classroom. Cambridge, MA: Harvard University Press.

McCafferty, S. G., & Rosborough, A. (2023). Contingency and Multimodal Communication in the Learning Environment: A Second Language Read-Aloud Lesson. *Language & Sociocultural Theory*, *9*(2).

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago press.

Streeck, J. (2009). Forward-Gesturing. *Discourse Processes*, *46*(2–3), 161–179. https://doi.org/10.1080/01638530902728793

Swain, M. (2006). Languaging, agency and collaboration in advanced language proficiency. In H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95–108). Washington, DC: Georgetown University Press.

Swain, M. & Watanabe, (2013). Languaging: Collaborative Dialogue as a Source of Second Language Learning. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics.* Blackwell.

Vygotsky, L. S. (1987). *The collected works of L. S. Vygotsky. Volume 1. Problems of general psychology, including the volume thinking and speech*. New York: Plenum.

Multimodal frame-based civil Antimafia constructions in a local Italian grammar

Paul Sambre

University of Leuven

paul.sambre@kuleuven.be

Context – This contribution analyzes a local grammar of Italian counterhegemonic discourse against the mafia, i.e. a counter-narrative in which an 'Ndrangheta victim tells how a local community denounced acts of violence and symbolic intimidation (1993-2023). Typically, mafia organizations maintain social consensus over a territory by imposing a law of silence. The grammar of Italian civil Antimafia discourse which breaks *omertà* remains unexplored (Author, 2017). Overall, this contribution explores the contribution of multimodal construction grammar (composed of speech and co-speech-gesture) to spatial vectors in a usage-based grammar.
Goals – Our main objective is to document and describe a local multimodal Italian Antimafia grammar in a Southern Italian community, where social resistance was linguistically expressed well before the so-called San Luca-Duisburg Feud. This grammar emerged as a relative of a mafia victim transformed private grieving in a networked set of (public) activities in Bovalino (close to San Luca), counterframing a social geography dominated by crime. We describe our Italian fieldwork interview data based on relevant core predicate-argument (mini-)constructions (for events and agent-patient roles) and non-core spatial and temporal relations encoded in verbal speech and depictive-deictic gestures based on Berkeley FrameNet scenarios and frames relevant to omertà and speaking up against it: Motion_scenario, Speak_on_topic and Silencing Frame (Boas 2021, Boas et al. 2024, Author fc.).
Theoretical tenets – Three theoretical elements complete the descriptive predicate structures in the Italian Antimafia frame-based grammar under study: (1) the CDS model in Cognitive Grammar (R. Langacker, 2008; R. W. Langacker, 2014), with a conceptualizing focus on temporal unfolding and representation of past-present-future windowing of usage events displayed in terms of prospective and/or retrospective viewing frames; (2) intersubjective construal and transitions between the interviewee's first-person individual perspective and (past and ongoing) interactions with a vast array of (mafia and, particularly, Antimafia) social actors, such as perpetrators and a collective network of other victims' relatives (Author, 2022), (3) Discourse Space Theory as a discursive complement to spatiotemporal satellite vectors in FrameNet (Chilton & Kopytowska, 2018, Cap 2021).
Data – We combine corpus linguistics with materialized field interview research, a rich cognitive semiotic and culturally relevant small-scale perspective (Sinha, 2021; Zlatev, 2010) in line with both Fillmore's (2020) seminal ethnographic semantics and later constructicon objectives. The data are a 2 h in situ ethnographic video interview with Deborah Cartisano, the daughter of the last murdered victim of mafia kidnapping, now actively involved in the regional Libera movement.
(Expected) results – We provide a set of predication-based constructions used by the interviewee where three grammatical levels of spatiotemporal and intersubjective construal coincide: (1a) a network of intersubjective positions in which the first-person perspective is extended towards a collective plural *we*, (1b) predicates for expressing antagonisms with perpetrators, (2) temporal constructs for connecting the present story-telling with (tragic) past events and, importantly, future effects of collective memory, (3) multimodal spatial positioning in the local geography, deictically and depictively gestured, and their formerly negative and current positive values for the community. As such we hope to contribute modestly to a socially motivated, cohesive (Michaelis, 2024) Italian FrameNet.

References

Boas, Hans C. “Construction Grammar and Frame Semantics.” In The Routledge Handbook of Cognitive Linguistics, 43–77. Routledge, 2021.

Boas, H. C., Ruppenhofer, J., & Baker, C. (2024). FrameNet at 25. International Journal of Lexicography, 37(3), 263-284.

Cap, P. (2021). “On the Development of the Social-Linguistic Nexus in Discourse Research: A Critical Review.” *Pragmatics and Society 12*, no. 2 (June 3, 2021): 309–33.

Chilton, P., & Kopytowska, M. (2018). *Religion, language and the human mind*. Oxford University Press.

Fillmore, C.J. (2020). *Form and meaning in language. Papers on linguistic theory and constructions* (P. Gras, J.-O. Östman, & J. Verschueren, Eds.). CSLI Publications.

Langacker, R. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195331967.001.0001

Langacker, R. W. (2014). Culture and cognition, lexicon and grammar. In *Approaches to language, culture, and cognition* (pp. 27–49). Springer.

Author E. (2020). Commemorare vittime di violenza: La street art come luogo di resistenza e discorso di positività. *Incontri. Rivista Europea Di Studi Italiani*, *35*(2), Article 2. https://doi.org/10.18352/incontri.10353

Author. (2017). The Multimodal Representation of Italian Anti-Mafia Discourse: Foregrounding Civil Resistance and Interlocution in Two Global English Video Reportages. *The Multimodal Representation of Italian Anti-Mafia Discourse: Foregrounding Civil Resistance and Interlocution in Two Global English Video Reportages*, 38–59.

Author. (2022). Actes définitoires contre-hégémoniques en territoire mafieux-Vers une praxis de la définition aristotélicienne en langue naturelle. *Rencontres*, 219–234.

Author. (fc.). “Il carcere ostativo è una pena che stiamo espiando noi famigliari di vittime della mafia”: vers un recadrage du discours de l’antimafia italienne sur la réclusion à perpétuité sans aménagement de peine. In Julia Kuhn (ed.), *Romanistisches Kolloquium und Kognitive Linguistik*, Sammelbände der Reihe Romanistisches Kolloquium. Narr Francke Attempto Verlag.

Sinha, C. (2021). Culture in language and cognition. In *The Routledge Handbook of Cognitive Linguistics* (pp. 387–407). Routledge.

Zlatev, J. (2010). Phenomenology and Cognitive Linguistics. In D. Schmicking & S. Gallagher (Eds.), *Handbook of Phenomenology and Cognitive Science* (pp. 415–443). Springer Netherlands.

Escaping oral papers: towards multimodality in academia

Magda Stanová

LARGE – Laboratory for Art, Research and Graduate Education
Academy of Fine Arts in Prague

m@magdastanova.sk

In academia, there is a bias towards text: published articles are valued more than conference talks, a text citation is preferred to a citation from a lecture, and even oral presentations at conferences are called “papers”.

Lectures, panel discussions, and conferences are academic genres where multimodal communication could thrive. Yet, these genres took on conventions that diminish the non-verbal modalities and make perception difficult: reading aloud texts that are too complicated for listening, speaking very fast in order to squeeze in as much material as possible, showing slides with long texts, sitting quietly for a long time without moving. When an image gets shown, it is rarely trusted by itself, and a layer of verbal explanation is added. Silence is considered awkward (in art history terms, there’s a kind of horror vacui of silence), so there is little time to think about what has been said.

In my poster, I will dissect the decorum of conventional presentation formats and look into what impact they have on our attention. I will suggest that once a piece of research has been reduced to a text format, it might be difficult to add other modes without them being auxiliary. I will contrast such oral papers with lectures that are natively multimodal and in which multimodality is used in a refined way. I will give examples of mathematicians, scientists (David Deutsch, Roger Penrose), and artists (Laurie Anderson, Daniel Yovino) who use performative elements in their lectures to share some aspects of their research in an implicit way. I will also introduce the relatively recently formalized field of artistic research, which could be seen as a testing ground for multimodal presentation formats. The poster will be accompanied by video citations on the Research Catalogue platform.

I will also provide a brief overview of the history of presentation tools and suggest that their affordances catalyzed different research content in different times, and that, if we don’t pick our presentation tools, the tools we use might be nudging us into doing research that fits to them.