# 2.1. Multimodality in Language Learning and Processing

## Auditorio Hall

Chairs:

Beata Grzyb
Yasamin Motamedi
*University College London, UK*

In the language sciences, there is an increasing recognition of the multimodal nature of language—that human language is more than just decontextualized, rule-governed speech. Languages, both spoken and signed, are primarily used in face-to-face communication, comprising a dynamic system of multiplex components in addition to structured speech and sign: gesture, facial expression, eye gaze and prosodic modulation. A growing body of research aims to model language as a multimodal system, accounting for the effects of multiple channels on language learning and processing beyond speech alone. This symposium aims to highlight the importance of building multimodal models of language, showcasing a range of theoretical and methodological perspectives on multimodality, from child language development, neuroimaging, sign language research and computational modelling.

Speakers:

Reyhan Furman
Henning Holle
David Peeters
Tessa Verhoef
Beata Grzyb

## 2.1.1. Do you see what I mean?: Bilingual and monolingual children use iconic gestures in speech disambiguation

Reyhan Furman
*University of Central Lancashire, UK*

Everyday language is rife with verbal ambiguity. One way listeners can deal with this problem is through using a speaker's gestures as disambiguation cues. For instance, adult listeners use the information in a speaker's iconic gestures to resolve lexical ambiguities in speech and to disambiguate the dominant and subordinate meanings of homonyms (e.g. *bat* to mean an animal or a sports instrument) (Holle & Gunter, 2007). Currently, we know very little about whether children can benefit from a speaker's gestures to resolve lexical ambiguities in speech. The only study that has investigated the developmental role gesture plays in the resolution of lexical ambiguity has focused on how children produce gestures to disambiguate homonym senses. Kidd and Holler (2009) found that while 3-year-olds rarely disambiguated between different senses, 4-year-olds used iconic gestures to do so. In two studies, we asked whether a speaker's iconic gestures help children resolve lexical ambiguity in speech. Monolingual (English) and bilingual (English-French) children aged 4 watched videoclips where a speaker uttered a sentence containing a homonym (e.g. *The glasses fell on the floor and broke*) that was accompanied either by a cospeech iconic gesture, or no gesture. Children were then given pictures that depicted the dominant and subordinate senses of the homonym and asked to choose which sense the speaker referred to. Our results show that children often had a preference for one of the senses of the homonyms, regardless whether this dominant sense was accompanied by a gesture. In contrast, the presence of iconic gestures significantly increased children's preference for the subordinate senses of homonyms. The results indicate that children benefit from speakers' gestures selectively, only using them to comprehend senses that they would otherwise ignore. Seeing different homonym senses represented in gesture might help young children better understand that one-to-many mappings are possible in language, and allow them to zoom in on senses that they have not yet fully acquired.

References

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, *19*, 1175-1192.

Kidd, E., & Holler, J. (2009). Children's use of gesture to resolve lexical ambiguity. *Developmental Science, 12*, 903-913.

## 2.1.2. Identifying linguistic and neural levels of interaction between gesture and speech during comprehension using EEG, fMRI and TMS

Henning Holle
*University of Hull, UK*

Conversational gestures are hand movements that co-occur with speech but do not appear to be consciously produced by the speaker. The role that these gestures play in communication is disputed, with some arguing that gesture adds only little information over and above what is already transmitted by speech alone. My own work has provided strong evidence for the alternative view, namely that gestures add substantial information to the comprehension process. One level at which this interaction between gesture and speech takes place seems to be semantics, as indicated by the N400 of the Event Related Potential. I will also present evidence for a syntactic interaction between gesture and speech (as indexed by the P600 component). Finally, I will present evidence suggesting that the left inferior frontal gyrus and left posterior temporal lobe are crucial components of the brain network for co-speech gesture comprehension. These findings are consistent with the idea that these areas play a joint role in gesture-speech integration, with IFG regulating strategic semantic access via top-down signals acting upon temporal storage areas.

## 2.1.3. Virtual Reality: A Promising Method for the Study of Multimodal Communication

David Peeters
*Tilburg University, the Netherlands*

Virtual Reality (VR) technology is increasingly used as a tool for psychological research. In the language sciences, however, its use has long remained virtually non-existent. In this talk, I will argue that immersive VR technology offers unique possibilities to answer fundamental questions about the neurocognitive mechanisms supporting the production and comprehension of multimodal language and communication. More than other methods, immersive VR combines ecological validity, experimental control, and repeatability in experimental settings. Moreover, as a mode of display, it can relatively easily be combined with the collection of EEG, eye tracking, and motion capture data to track online linguistic and communicative processes. I will discuss ongoing work from our lab that shows the added value of VR in the study of i) the interplay between communicative intentions and actions in the production of multimodal deictic speech acts, and ii) the concurrent processing of facial expressions and speech in language comprehension.

## 2.1.4. Iconic representations of actions and objects: gestural biases and the emergence of patterns in sign language

Tessa Verhoef
*University of Leiden, the Netherlands*

The bodily-visual modality affords the ability to iconically represent actions and objects with the movement and shape of the body and hands. In sign languages, related verbs and nouns can be distinguished, for instance, in their manner of movement or iconic handshapes. This talk examines whether nonsigners are sensitive to representational strategies in natural sign languages when pairing gestures with actions or objects, and how these initial biases may drive the emergene of patterns in conventionalized languages. It has been found that sign-naïve gesturers show a strong preference for representing actions with "handling" handshapes, which show how you hold the object, and a slight preference for representing objects with "instrument" handshapes, which show the shape of the object, an encoding strategy that also distinguishes verbs and nouns in ASL (Padden et al., 2015). In ASL, movement patterns also distinguish related verbs and nouns: verbs are formed with longer/continuous movements, while nouns are formed with faster/constrained movements (Supalla & Newport, 1978). Here, we showed 1175 (online) participants pairs of videos of iconic gestures, and asked them to identify each video as representing either an action (e.g. "using a handsaw") or object (e.g. "a handsaw"). When movement is constant across videos and handshape varies, participants tend to map handling handshapes to actions and instrument handshapes to objects. When handshape is constant and movement varies, the expected movement preference is also found. When both handshape and movement vary between the videos, participants weight their handshape preference over the preference based on movement patterns. A second study investigates the interaction between these biases and pressures to systematize and conventionalize in the learnability of artificial languages. Gestural biases, already present in participants who have no experience with sign language, is found to influence the emergence of language-like patterns.

## 2.1.5. Future directions for multimodality research

Beata Grzyb
*University College London, UK*

In the final session of the symposium, we will highlight the importance of multimodal language models across language research, asserting that a multimodal model of language gives a more comprehensive understanding of the mechanisms which underlie language learning and processing. Using some examples of computational models of multimodal behaviours, we will present possible future directions in multimodal language research: i) complex multimodal models, that account for several interacting behaviours, and ii) how language as a multimodal system can shed light on the cognitive systems beyond language. We will also argue that bringing together computer science and developmental psychology/cognitive science permits to combine the best of human learning and

the best of machine learning in a way that can benefit both. The results of empirical studies can guide the development of computational models for multimodal learning and processing in socially intelligent artificial agents. Computer science, in turn, can inform developmental psychology by generating explicit and embodied computational hypothesis of how and why different cues combine in language learning and processing.