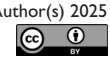
Invited Commentary

FIRST LANGUAGE

First Language
1–5


Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/01427237251326984
journals.sagepub.com/home/fla
Sage

Language acquisition occurs in multimodal social interaction: A commentary on Karadöller, Sümer and Özyürek

Jennifer Sander[1,2], Yayun Zhang[1] and Caroline F. Rowland[1,3]

[1]Language Development Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]Max Planck School of Cognition, Leipzig, Germany
[3]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

Abstract

We argue that language learning occurs in triadic interactions, where caregivers and children engage not only with each other but also with objects, actions and non-verbal cues that shape language acquisition. We illustrate this using two studies on real-time interactions in spoken and signed language. The first examines shared book reading, showing how caregivers use speech, gestures and gaze coordination to establish joint attention, facilitating word-object associations. The second study explores joint attention in spoken and signed interactions, demonstrating that signing dyads rely on a wider range of multimodal behaviours – such as touch, vibrations and peripheral gaze – compared to speaking dyads. Our data highlight how different language modalities shape attentional strategies. We advocate for research that fully incorporates the dynamic interplay between language, attention and environment.

Keywords

Language acquisition, sign language, visual attention, triadic interaction, book reading, pointing

Language is multimodal and this is true not only for adult language comprehension and production but also for language acquisition. Karadöller et al. (2024) provide overwhelming support for this statement in their review article, focussing on evidence from

Corresponding author:
Jennifer Sander, Language Development Department, Max Planck Institute for Psycholinguistics, Wundtlaan 1, Nijmegen 6525 XD, The Netherlands.
Email: jennifer.sander@mpi.nl

child first language production in speech, gesture and sign, but acknowledging that this also applies to language input and comprehension.

We fully agree with the authors' statement and want to support it from a different perspective. Beyond language comprehension and language production, it is important to acknowledge how language use is embedded in multimodal social interaction. Language exchange happens in interactions and these interactions are multimodal beyond the language exchange. As a result, language acquisition takes place in interaction in which every single step is multimodal, in every language modality.

Recognising this enables us not only to acknowledge the language being produced and perceived as multimodal but also to acknowledge the multimodality of the interaction itself and of the environment in which the interaction takes place. Whether it is objects (toys, books), the physical surroundings, actions, or the non-verbal cues given out by the interlocutors – all of these are potential points of reference that have to be integrated to allow for successful language acquisition. In other words, instead of conceptualising interaction as a dyadic exchange between a person producing multimodal language and a person comprehending multimodal language, we need to acknowledge the fact that interactions between children and caregivers are very often triadic, incorporating, in addition, objects or actions in the environment, which are also perceived and referenced multimodally. This perspective has important implications for our understanding of the language acquisition process itself.

We illustrate here with two examples: the role of pointing gestures in shared book reading and the role of visual attention in episodes of joint attention. Both examples stem from our own research on real-time interactions in naturalistic contexts within two language modalities, spoken and sign language. The study of pointing gestures in book reading illustrates how language knowledge is transferred with the support of an external referent. Referencing through pointing creates a uniquely rich environment for children to learn new object labels anchored in the act of pointing, connecting the novel label to the pointed-to-referent. The study of the interaction of visual and tactile attention and language in joint attentional interactions illustrates the relevance of a multimodal understanding of joint attention; it embeds language acquisition into its rich social environment and helps us account for different multimodal realities of dyads using different language modalities.

Our first example focuses on shared book reading, which is a common daily activity for young children, particularly in Western societies, and serves as a key example of how multimodal interactions foster language development. Far beyond a simple exchange of words, shared book reading integrates speech, gestures, visual attention and environmental context, making it an inherently rich and multimodal experience for young children. While much is understood about how linguistic input provided during shared reading effectively supports language acquisition (Noble et al., 2018), we have limited knowledge of whether non-verbal behaviours, especially pointing gestures (Rohlfing et al., 2015), play a role in children's language skills and how they influence language acquisition at the moment. To better understand how caregivers and children jointly create interactional routines during reading to achieve joint focus, we set up a naturalistic study using head-mounted eye-trackers with 16 parents and their children aged 18 to 24 months in the USA (Zhang et al., in prep). We analysed the data from the

eye-trackers to identify exactly where parents and children were looking, moment by moment, as well as keep track of their pointing gesture use. We determined that both parents and children employed a variety of cues to maintain joint attention and facilitate learning. Specifically, parents often paired their verbal descriptions with gestures such as pointing at objects in the book. These gestures helped direct the child's gaze to the named referent, making it easier for the child to form accurate word-object associations. Children's own pointing also contributed to this interaction, often prompting parents to respond with naming or descriptive language, further reinforcing learning opportunities. By weaving together speech, gestures and visual stimuli, parents provide a scaffold that helps children navigate the complexities of language.

Our second example focuses on joint attention. In classic joint attention research, the role of visual attention to objects and actions has been emphasised in that visual attention is considered necessary to connect a (novel) reference to its referent. Thus, a canonical definition of joint attention assumes parallel visual object focus and auditory speech input as the ideal learning environment. However, in signed caregiver-child interactions, visual attention to the object and visual language input compete for a child's visual attention. This means that parallel input of object and language cannot be achieved as effortlessly as in spoken interactions. Thus, it is highly likely that non-vision-based cues (e.g., touch) are used as well and that ignoring these cues will give us a skewed view of how signing dyads establish joint attention. In previous studies, we were able to show that including a wider range of multimodal behaviours in social interactions into our considerations about joint attention improves our understanding of the relationship between joint attention and children's vocabulary size (Sander et al., 2024), and leads us to a better understanding of how caregivers use joint attention to support language acquisition (Sander et al., 2023).

To better understand the role of different multimodal behaviours in joint interactions, we set up a mobile head-mounted eye-tracking study to identify the language-modality-specific contribution of gaze, touch, language and attentional behaviours in joint attentional caregiver-child interactions (Sander et al., in prep). We analysed the recordings from the head-mounted eye tracker and additional third-person video recordings of sign and spoken language interactions between 24 caregivers and their children (aged 1–5 years) in the Netherlands, who were using either spoken Dutch or Sign Language of the Netherlands (NGT). We identified joint attention according to the joint attention coding scheme by Gabouer and Bortfeld (2021) using annotations for gaze, touch, attentional behaviours and language usage. We found that both signing and speaking caregivers and children employed a number of different multimodal behaviours in joint attentional episodes, from overt attention-getting behaviours using vision (like waving), touch (like tapping), vibrations (like banging), to gaze following and tactile attention-guiding behaviours. However, as predicted, signing dyads engaged in more multimodal attention-getting behaviours beyond gaze and language usage than non-signing dyads. We further found that dyads' gaze strategies during naming events differed by language modality. In signing dyads, the named objects were less often fixated by the child and more often in the visual periphery of the child at the time point of naming than was the case for children in speaking dyads.

We concluded that to fully understand the role of joint attention in language acquisition, we need to adopt a more flexible understanding of joint attention that involves all behaviours directed at the interaction partner or object or action of interest. Considering the whole range of multimodal behaviours in joint interactions enables us not only to explore how joint attention is established in speaking or signing dyads but also in interactions of blind individuals as well as individuals using tactile sign language.

In sum, language production as well as language comprehension are multimodal. Both also take place in a multimodal social environment. As demonstrated in our two research examples, language acquisition takes place in interactions, which themselves are highly multimodal beyond the multimodality of the language exchange, and the way in which dyads use the environment in interaction facilitates language acquisition. Our theories of language acquisition need to incorporate the multimodal interactions in which language acquisition takes place to fully understand the mechanisms and processes involved. We join Karadöller et al. (2024) in urging future research to consider a more integrative and multimodal perspective on child language acquisition.

Author contributions

Jennifer Sander: Conceptualisation; Writing – original draft; Writing – review & editing.
Yayun Zhang: Conceptualisation; Writing – review & editing.
Caroline F. Rowland: Conceptualisation; Writing – review & editing.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was funded by the Max Planck Society.

ORCID iDs

Jennifer Sander https://orcid.org/0009-0009-7695-6360
Yayun Zhang https://orcid.org/0000-0002-9017-2740
Caroline F. Rowland https://orcid.org/0000-0002-8675-8669

References

Gabouer, A., & Bortfeld, H. (2021). Revisiting how we operationalize joint attention. *Infant Behavior and Development*, *63*, 101566. https://doi.org/10.1016/j.infbeh.2021.101566

Karadöller, D. Z., Sümer, B., & Özyürek, A. (2024). First-language acquisition in a multimodal language framework: Insights from speech, gesture, and sign. *First Language*, *0*(0). https://doi.org/10.1177/01427237241290678

Noble, C. H., Cameron-Faulkner, T., & Lieven, E. (2018). Keeping it simple: The grammatical properties of shared book reading. *Journal of Child Language*, *45*(3), 753–766. https://doi.org/10.1017/S0305000917000447

Rohlfing, K. J., Grimminger, A., & Nachtigaller, K. (2015). Gesturing in joint book reading. In B. Kümmerling-Meibauer, J. Meibauer, K. Nachtigäller, & K. J. Rohlfing (Eds.), *Learning from picturebooks* (pp. 99–116). Routledge.

Sander, J., Lieberman, A., & Rowland, C. F. (2023). Exploring joint attention in American Sign Language: The influence of sign familiarity. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45), Philadelphia, PA, USA, August 10–13, 2016.

Sander, J., Çetinçelik, M., Zhang, Y., Rowland, C. F., & Harmon, Z. (2024). Why does joint attention predict vocabulary acquisition? The answer depends on what coding scheme you use. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 46), Philadelphia, PA, USA, August 10–13, 2016.

Sander, J., Eikelboom, D., Zhang, Y., & Rowland, C. F. (in prep). *Multimodal attention and gaze strategies in signing and speaking dyads during word learning*. https://osf.io/ha57v/

Zhang, Y., Knabe, M. L., Rowland C. F., & Yu, C. (in prep). *The helping hand: The temporal dynamics of pointing gestures and visual attention facilitate early word learning in shared book reading*.