Special Issue: First Language Acquisition in a Multimodal Language Framework

FIRST LANGUAGE

The high dimensionality of caregiver-child communication: A commentary on Karadöller, Sümer, and Özyürek

First Language
2025, Vol. 45(6) 748–752

DOI: 10.1177/01427237251328056
journals.sagepub.com/home/fla
Sage

Jessica E. Kosie
School of Social and Behavioral Sciences, Arizona State University, USA

Casey Lew-Williams
Department of Psychology, Princeton University, USA

Abstract
Karadöller, Sümer, and Özyürek are doing an important service to the field by emphasizing multimodality in young children's language learning. They integrate research on speech, gesture, and sign to highlight the independent and combined influence of these modalities on how children learn to communicate. In this commentary, we call for scientists to further broaden the study of natural caregiver-child communication by encompassing a dynamic set of interacting signals that facilitate complex information exchange.

Keywords
Multimodality, naturalistic data, infant-directed communication, language learning, caregiver-infant interaction

Research on the development of infants' and young children's communication with adults has long defaulted to the modality of speech. Karadöller, Sümer, and Özyürek (2025), henceforth KS&Ö, provide a detailed and compelling account of why this is too

Corresponding author:
Jessica E. Kosie, School of Social and Behavioral Sciences, Arizona State University, 4701 W Thunderbird Road, Glendale, AZ 85306, USA.
Email: jkosie@asu.edu

limited in scope if we want to understand the natural complexity and multimodality of communication and learning. They overview prior work showing that spoken language learning is facilitated by both speech and gesture (separately and in combination), taking enormous care to elaborate on pointing and iconic gestures as signals that co-occur with speech across time. Further, they discuss how children's sign language learning is facilitated by indexicality, iconicity, and simultaneity. We could not agree more with KS&Ö that "Human language is inherently multimodal in nature" (p. 674) and that language research will inevitably move toward multimodality in the years and decades to come. Yet, a comprehensive account of multimodality will need to consider dynamic interactions between dimensions of communication beyond speech, gesture, and sign.

Recent estimates suggest that approximately 60% of communicative acts directed to infants and young children incorporate cues from two or more modalities (Kosie & Lew-Williams, 2024). Speech and gesture occur together frequently, as KS&Ö describe in detail. However, actions on objects, facial expressions conveying emotion, and infant-directed touch are also prevalent in infants' communicative input (Abu-Zhaya et al., 2017; Brand et al., 2002; Chong et al., 2003). To understand the complex dynamics of language learning, scientists should integrate across communicative dimensions beyond speech, gesture, and sign because multiple modalities are modified in interactions with infants, and these modifications are linked to attention and learning. When interacting with infants (vs. adults), caregivers' speech, actions, gestures, emotion-conveying facial expressions, and social touch are more exaggerated and repetitive (Abu-Zhaya et al., 2017; Brand et al., 2002; Chong et al., 2003; Hilton et al., 2022; Iverson et al., 1999). For example, caregivers use more extreme emotion-conveying facial expressions when describing positively and negatively valenced images to infants versus adults (Wu et al., 2023). In addition, when caregivers talk to their infants about body parts, they also systematically touch the body part as it is named (Abu-Zhaya et al., 2017), and this alignment between speech and social touch increases infants' attention to the named body part (Tincoff et al., 2019). Infants also prefer to view infant-directed over adult-directed action demonstrations (Brand & Shallcross, 2008), and they are more likely to imitate actions and engage longer with objects following demonstrations that feature infant-directed modifications (Koterba & Iverson, 2009; Williamson & Brand, 2014). Thus, prior work clearly demonstrates that infants' communicative input is multidimensional and complex, requiring careful consideration of how multimodal cues are integrated to create meaning.

Though the integration of multimodal communicative cues enables richer understanding than is possible through a single modality (Clark, 2016; Morgenstern, 2023), the majority of research to date has focused on individual cues in isolation or the combination of speech and only one other non-speech cue. There has been substantially less exploration of how multiple cues are combined to produce meaning, though this understanding is critical for theories of early communicative development. Determining how these cues are integrated, however, is a challenge. It is unlikely, for example, that the total information gained by a multimodal communicative event is simply the sum of its parts and, instead, might be better captured by a calculation that incorporates the assignment of customized weights to each of the multimodal cues being used. These weights are likely to vary with features, including the meaning that the communicator intends to

express, the context in which the communicative event takes place, and the developmental stage or abilities of the communicative partner, adding to the complexity of this calculation. It may additionally be informative to explore how interactional features influence the order in which multimodal cues occur, incorporating analytic techniques such as Granger Causality or Cross-Recurrence Quantification Networks (Xu et al., 2020) to assess the moment-to-moment dynamics of multimodal events as they unfold and test links to processing and learning. Understanding how information is integrated across numerous multimodal cues is consequential for efforts to understand how children learn from their everyday multimodal input.

To successfully study multimodal communication, researchers need access to data capturing natural human interaction. Many fields in the behavioral sciences are moving toward a real embracing of natural (and big) data, including developmental science, language science and computational cognitive science. Experiments are great for providing mechanistic insight with rigorous control, but quantified descriptive approaches with natural data promise to provide an essential complement to this: the ability to evaluate phenomena of interest (including speech, gesture, and sign) as they operate in people's everyday lives. Reflecting these complementary approaches, KS&Ö cite both experimental and naturalistic studies in support of their arguments. But going forward, prioritization of studies capturing natural interaction between caregivers and young children – both audio and video – will be key for advancing knowledge of multimodal communication. Many corpora are now being used to evaluate assumptions in the field and to test entirely new hypotheses, including longitudinal corpora (Bergelson et al., 2019; Sullivan et al., 2021), 1-hr interactions across nearly 1000 mother-infant dyads (Soska et al., 2021), cross-cultural speech corpora (Bergelson et al., 2023), and short in-lab and at-home recordings (Kosie & Lew-Williams, 2024). Technologies for computer vision will be developed and refined for many years to come, allowing scientists to adopt a truly broad definition of communicative behaviors and to complete projects in reasonable timeframes. For the coming few years, careful human annotation methods will need to suffice for most research questions. But the benefits of analyzing natural interactions will be borne out in countless ways, including for research on gesture and sign.

The future of research on caregiver-child communication will not only combine vocal and manual behaviors but also a broad range of embodied signals that guide young children's learning in their natural habitats. By harnessing large-scale video datasets and making a habit of analyzing additional signals – including emotion, action, and social touch – the field will be better equipped to understand the high dimensionality of information exchange and promote learning in children with diverse capacities and from diverse environmental contexts.

Author Contributions

Jessica E. Kosie: Conceptualization; Writing – original draft; Writing – review & editing.
Casey Lew-Williams: Conceptualization; Writing – original draft; Writing – review & editing.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Jessica E. Kosie https://orcid.org/0000-0002-2390-0963

References

Abu-Zhaya, R., Seidl, A., & Cristia, A. (2017). Multimodal infant-directed communication: how caregivers combine tactile and linguistic cues. *Journal of Child Language*, *44*(5), 1088–1116.

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, *22*(1), e12715.

Bergelson, E., Soderstrom, M., Schwarz, I.-C., Rowland, C. F., Ramírez-Esparza, N., Hamrick, L. R., Marklund, E., Kalashnikova, M., Guez, A., Casillas, M., Benetti, L., van Alphen, P., & Cristia, A. (2023). Everyday language input and production in 1,001 children from six continents. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(52), e2300671120.

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': Modifications in mothers' infant-directed action. *Developmental Science*, *5*(1), 72–83.

Brand, R. J., & Shallcross, W. L. (2008). Infants prefer motionese to adult-directed action. *Developmental Science*, *11*(6), 853–861.

Chong, S. C. F., Werker, J. F., Russell, J. A., & Carroll, J. M. (2003). Three facial expressions mothers direct to their infants. *Infant and Child Development*, *12*(3), 211–232.

Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review*, *123*(3), 324–347.

Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., Simson, J., Knox, D., Glowacki, L., Alemu, E., Galbarczyk, A., Jasienska, G., Ross, C. T., Neff, M. B., Martin, A., Cirelli, L. K., Trehub, S. E., Song, J., Kim, M., . . . Mehr, S. A. (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*, *6*(11), 1545–1556.

Iverson, J. M., Capirci, O., Longobardi, E., & Cristina Caselli, M. (1999). Gesturing in mother-child interactions. *Cognitive Development*, *14*(1), 57–75.

Karadöller, D. Z., Sümer, B., & Özyürek, A. (2025). First-language acquisition in a multimodal language framework: Insights from speech, gesture, and sign. *First Language*, *45*(6), 673–710. https://doi.org/10.1177/01427237241290678

Kosie, J. E., & Lew-Williams, C. (2024). Infant-directed communication: Examining the many dimensions of everyday caregiver-infant interactions. *Developmental Science*, *27*(5), e13515.

Koterba, E. A., & Iverson, J. M. (2009). Investigating motionese: The effect of infant-directed action on infants' attention and object exploration. *Infant Behavior & Development*, *32*(4), 437–444.

Morgenstern, A. (2023). Children's multimodal language development from an interactional, usage-based, and cognitive perspective. *Wiley Interdisciplinary Reviews. Cognitive Science*, *14*(2), e1631.

Soska, K. C., Xu, M., Gonzalez, S. L., Herzberg, O., Tamis-LeMonda, C. S., Gilmore, R. O., & Adolph, K. E. (2021). (Hyper)active data curation: A video case study from behavioral science. *Journal of Escience Librarianship*, *10*(3). https://doi.org/10.7191/jeslib.2021.1208

Sullivan, J., Mei, M., Perfors, A., Wojcik, E., & Frank, M. C. (2021). SAYCam: A Large, longitudinal audiovisual dataset recorded from the infant's perspective. *Open Mind: Discoveries in Cognitive Science*, *5*, 20–29.

Tincoff, R., Seidl, A., Buckley, L., Wojcik, C., & Cristia, A. (2019). Feeling the way to words: Parents' speech and touch cues highlight word-to-world mappings of body parts. *Language Learning and Development: The Official Journal of the Society for Language Development*, *15*(2), 103–125.

Williamson, R. A., & Brand, R. J. (2014). Child-directed action promotes 2-year-olds' imitation. *Journal of Experimental Child Psychology*, *118*, 119–126.

Wu, Y., Taylor, I., Chen, H., & Frank, M. C. (2023). Adults tailor their emotional expressions to infants through 'emotionese'. *Annual Meeting of the Cognitive Science Society*, *45*(45), 3312–3318. https://escholarship.org/uc/item/9vm4j7bp

Xu, T. L., de Barbaro, K., Abney, D. H., & Cox, R. F. A. (2020). Finding Structure in time: Visualizing and analyzing behavioral time series. *Frontiers in Psychology*, *11*, 1457.