Toward a theoretical framework of multimodal communicative efficiency

Corresponding Author (correspondingauthor@amlap.com)[1], First name and Last name Author one[1],
[1]Department of Something, Some University

Communicative efficiency, defined as the optimization of the cost–benefit trade-off in communication, has been central to understanding language structure and use, albeit primarily from a unimodal, spoken-language perspective [1]. Such approaches overlook the ongoing paradigmatic shift in our understanding of language arising from insights into how language is actually used in interaction: multimodally and adaptively [2,3]. I therefore propose a new theoretical framework of communicative efficiency, which reconceptualizes efficiency as a property of language production and comprehension that flexibly exploits multiple dimensions—time and space, acoustic and visual modalities, sequentiality and simultaneity, arbitrariness and iconicity—in response to contextual demands. These demands can be categorized in terms of the affordances available for communication (linguistic modality), who we are communicating with (the addressee's informational needs), what we are communicating about (content), and the communicative goal.

Drawing on evidence from my recent experimental and corpus work on co-speech gesture, sign languages, and silent gesture, I argue that communicative efficiency cannot be understood without accounting for the multimodal and adaptive nature of language. I present converging evidence from studies on co-speech gesture and sign languages showing that speakers and signers systematically adjust their use of the visual modality across communicative contexts. For example, when teaching new skills to children, speakers produce more sign-language-like [4], highly informative gestures to support the sequentially unfolding information conveyed in speech (Figure 1), and further enhance gesture salience through double ostensive cues (gaze shifts toward the gesture and demonstratives in speech). In the same teaching context with adult addressees, gestures tend to be less informative (Figure 2), and highlighting relies on lower-cost cues such as gaze alone. I argue that these patterns reflect a context-sensitive trade-off: speakers invest greater multimodal effort when they anticipate higher comprehension demands, while relying on less costly representations and cues when demands are lower. Together, these and related findings suggest that the visual modality plays a fundamental role in enabling languages to balance communicative costs and benefits across contexts. As such, multimodality of language provides a novel window into how communicative systems adapt to support efficient communication.

Figure 1: Speech (“you need to arrange these disks”) accompanied by a highly informative gesture representing two disks of different sizes, with the smaller disk positioned on top of the larger disk (adult–child teaching context).

Figure 2: Speech (“you need to arrange these disks”) accompanied by a less informative gesture representing a single disk (adult–adult teaching context).

References:

[1] Levshina, N. (2022). Communicative efficiency. Cambridge University Press.
Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In Advances in psychology (Vol. 9, pp. 287-299). North-Holland.
[2] Hagoort, P., & Özyürek, A. (2025). Extending the architecture of language from a multimodal perspective. Topics in cognitive science, 17(4), 877-887.
[3] Holler, J., & Bavelas, J. (2017). Multi-modal communication of common ground. Why gesture, 213-240.
[4] Slonimska, A., Özyürek, A., & Capirci, O. (2020). The role of iconicity and simultaneity for efficient communication: The case of Italian Sign Language (LIS). Cognition, 200, 104246.