Multimodal Processing in Bilingualism and Second Language Acquisition

In the context of language processing, multimodality refers to communication via more than one sensory modality. For example, auditory cues such as speech may be accompanied by visual cues such as hand gestures, facial expressions, or text. Precursor research on communication strategies, which are used to compensate for unknown words, indicate that multimodality is particularly helpful for communicating in a second language (L2) because the secondary modality provides a means of overcoming challenges in communicating in the primary modality due to limited familiarity with or use of the L2 (Dörnyei, 1995). For example, if someone were to say to an L2 English speaker, “Can you hand me the fidget spinner?” while pointing or gesturing as if using one, these visual cues would increase the possibility that the L2 English speaker, who may be unfamiliar with the term fidget spinner, would hand the correct object to the requestor. Likewise, if an L2 English speaker were to similarly point or gesture towards the fidget spinner while saying, “Can you hand me that thing?”, it would increase the possibility that the addressee would retrieve the fidget spinner even though the speaker didn’t produce its label via speech.

Two important theories characterizing the effects of multimodality on learning with relevance to L2 acquisition are dual coding theory (Paivio, 1990) and multimedia learning theory (Mayer, 2002). Dual coding theory posits that images and speech can be simultaneously encoded via analog perceptual codes in the visual channel and symbolic conceptual codes in the verbal channel, resulting in robust, enriched representations. Similarly, multimedia learning theory posits that deeper learning is achieved when speech and relevant images are presented together than when speech is presented alone (multimedia principle) or than when relevant images are presented with text (modality principle). These theories provide clear recommendations for how
information should be conveyed via speech and images to optimize L2 word learning and narrative comprehension. Moreover, their recommendations can be extended to representational gestures, which accompany speech and convey information related to it iconically via their form and motion, as well as deictic (pointing) gestures, which direct attention to referents. Thus, they have implications for how bilinguals, L2 learners, and their interlocutors can employ co-speech gestures to communicate more effectively.

This entry will provide an overview of key insights from research on multimodal processing in bilingualism and L2 acquisition. In particular, the impacts of text, images, facial cues, and hand gestures in relation to speech will be discussed, as will hand and air writing, which combine haptic and visual input. Because most extant research examining the impacts of these stimuli has focused on L2 vocabulary acquisition and spoken discourse comprehension, this entry will focus primarily on these aspects of L2 processing. Thus, this entry will cover multimodal L2 processing across the auditory, visual, and haptic modalities, providing insight into how it can be conjointly leveraged by bilinguals, L2 learners, and their L1 interlocutors to communicate effectively.

**Text and Speech**

Research examining the impact of text in relation to L2 speech has shown that comprehension of L2 spoken discourse as well as L2 vocabulary acquisition in videos is facilitated by L1 subtitles and L2 captions (Bianchi & Ciabattoni, 2008). Moreover, L2 reading comprehension is enhanced when spoken voice-overs are present (Marzban, 2011). These findings are consistent with the redundancy principle of multimedia learning theory, which maintains that presentation of both text and speech can enhance comprehension if it does not increase cognitive load.
**Images and Speech**

With respect to L2 vocabulary learning, viewing pictures of referents in addition to hearing L2 words during learning enhances memory for them (Morett, 2019), with some work suggesting that text in addition to images of referents accompanied by spoken L2 words further enhances memory for them (Jones & Plass, 2002). Similar beneficial effects of observing images and videos of referents have been found for L2 vocabulary acquisition in the context of reading (Al-Seghayer, 2001; Chun & Plass, 1996). These findings suggest that images of referents may facilitate memory for them, particularly when verbal labels are present.

**Facial Cues and Speech**

Facial cues consist of movements of the lips, jaw, eyebrows, and other facial components produced while speaking, sometimes referred to as visible speech. Given that L1 speech perception is enhanced by observing facial cues, particularly when speech quality is degraded, L2 speech perception should be similarly enhanced by observing facial cues. The McGurk effect, in which the auditory component of one speech sound (e.g., /b/) is paired with the visual component of another speech sound (e.g., /g/), resulting in an illusory perception of a third speech sound (e.g., /d/), provides a means of examining the influence of facial cues on auditory L2 speech. Research probing the McGurk effect in L2 indicates that facial cues are combined with L2 speech and that that listeners’ L1s influence their weighting of visual and auditory cues (Hardison, 1996). From another perspective, opaque face masks permit examination of how the absence of facial cues affects L2 speech perception. Given comparable acoustic quality, opaque face masks impair L2 word recognition and speech comprehension relative to unmasked faces (Smiljanic et al., 2021). Eye-tracking evidence of visual attention reveals that, during L2 speech perception, listeners attend primarily to facial cues from the eyes and mouth rather than cues
from other parts of the body (Batty, 2021). Together, these findings indicate that facial cues enhance L2 speech perception.

**Hand and Air Writing**

Hand writing refers to producing letters, words, or characters with a handheld utensil that leaves a visible trace, whereas air writing refers to using the hands themselves to produce letters, words, or characters without a visible trace. In L2, hand writing facilitates learners’ acquisition of orthography, or written language (Guan et al., 2011), and air writing facilitates learners’ acquisition of logographic characters, which represent words or morphemes via their forms, to an even greater extent than hand writing in some cases (Thomas, 2015). These findings support the lexical quality hypothesis (Perfetti & Hart, 2002), which maintains that associations between orthographic, phonological, and semantic information contribute to high-quality lexical representations.

**Gesture and Speech**

Gesture consists of hand movements conveying the content and/or rhythm of concurrent speech via its form and motion (McNeill, 2005). Gesture provides visual input when observed and haptic input when produced, enriching L2 speech representations via these additional modalities. In an unfamiliar L2, observing and producing gestures conveying referents facilitates acquisition of unknown words (Allen, 1995); likewise, observing and producing gestures conveying novel speech sounds visually facilitates their acquisition (Morett & Chang, 2015). Moreover, for listeners with limited L2 proficiency, observing gestures facilitates L2 listening comprehension (Sueyoshi & Hardison, 2005). Indeed, L1 speakers produce more iconic and deictic gestures when recounting narratives to L2 speakers (Adams, 1998), as well as larger, faster, and longer gestures when describing pictures to L2 speakers (Prové et al., 2022),
suggesting that L1 speakers adjust their gestures to facilitate the comprehension of L2 interlocutors. Moreover, bilinguals gesture more frequently than monolinguals, reflecting cross-linguistic gestural transfer (Nicoladis, 2007) and facilitating speech production and communication (Gullberg, 2011). These findings concerning gesture in the context of L2 acquisition and bilingualism support theories of embodied cognition, which postulate that the body influences language processing and language is grounded in embodied experience (Shapiro, 2019).

Conclusion

In conjunction with auditory input from speech, visual and haptic input from text, images, facial cues, hand and air writing, and gestures enriches L2 representations in bilinguals and L2 learners. Thus, multimodal representations resulting from these inputs facilitate L2 processing, thereby improving communication between L2 and L1 speakers and bilinguals.

Suggested Readings


See Also

wbeal0455.pub2
Biography

Laura M. Morett is Assistant Professor of Educational Psychology and Neuroscience at the University of Alabama. Her research focuses on the role of embodiment in language processing, with a particular focus on gesture and its role in L2 acquisition. Noteworthy recent publications include Morett et al. (2022), published in *Cognition*; Morett et al. (2021), published in *JEP:LMC*; and Morett et al. (2020), published in *Brain Research*. 
References


https://doi.org/10.1017/S0142716407070245


https://doi.org/10.1121/10.0005191


https://doi.org/10.1111/j.0023-8333.2005.00320.x


https://doi.org/10.1111/lang.12128