G-M&B's proposed next step of developing new technologies to measure utterances more finely will not clarify the distinction. As even they mention, the same form can be generated by either a gradient or categorical system (sect. 4.2, para. 3). Take their example of wiggling fingers moving along a path to indicate a person running (sect. 5, para. 3). Nothing in the form itself determines whether it has internal combinatorial structure; what matters is whether pieces of that form map holistically and veridically to the world (where legs, internal movement, and path of movement all occur together) or according to a system used to generate this and other utterances, using recombinable handshape, manner, and path elements. Figure 2 illustrates that the same manual utterance can be iconic and holistic in one linguistic context, and morphological and combinatorial in another.

We agree with the importance of creating a unified account of language that includes all aspects of its production, whether manual or vocal. We suggest that the study of spoken and signed languages at moments of change – particularly language acquisition, emergence, and change – offer a better view of the sources of language structure. The dimensions of discreteness and combinatoriality are of interest not because they help define gesture, but because they represent an abstraction and reconfiguration of information from how it is organized in the world. Accordingly, these dimensions are sites of qualitative shifts as language is created and changed. Forms appearing in new contexts constitute evidence of corresponding changes along these dimensions. For example, at some point learners transformed the onomatopoeic verbal gesture "mooo," allowing it to participate in combinatorial expressions like "The cow moo-ed all day." The path that elements follow as they become linguistic reveals human language-making capacity at individual, community, and multigenerational timescales. The world offers a continuous and image-rich stream of experience; representations that derive their structure directly from the world will be correspondingly gradient and holistic. Our own work demonstrates that gradient, context-dependent, and imagistic forms are reshaped by learners into discrete, recombinable elements (Coppola & Senghas 2010; Senghas et al. 2004). Investigating transformation over time along these dimensions will reveal how the nature of language reflects human minds, rather than the world to which language refers.

The physiognomic unity of sign, word, and gesture

Carlos Cornejo and Roberto Musa

Escuela de Psicología, Pontificia Universidad Católica de Chile, 7820436 Santiago, Chile.
cca@uc.cl   rfmusa@uc.cl

Abstract: Goldin-Meadow & Brentari (G-M&B) are implicitly going against the dominant paradigm in language research, namely, the "speech as written language" metaphor that portrays vocal sounds and bodily signs as means of delivering stable word meanings. We argue that Heinz Werner's classical research on the physiognomic properties of language supports and complements their view of sign and gesture as a unified system.

Goldin-Meadow & Brentari's (G-M&B's) view of sign and gesture as a unified system is more revolutionary than it might outwardly seem. They part with the prevailing backdrop of contemporary language studies and embrace instead the long-neglected (if not poorly understood) tradition of conceiving language as a human activity instead of a chain of thing-like lexical pieces. The undergoing paradigm shift the authors call for draws much of its force from leaving behind the extant framework that has dominated language research for the past 300 years, namely, the "speech as

written language" metaphor (Ingold 2007) that portrays vocal sounds and bodily signs as means of delivering stable word meanings. At the dawn of modernity, the invention of the printing press and subsequent availability of books progressively inspired the idea that human speech was ultimately a variant of printed word production: Our mind uses sound units (i.e., words) whose semantic content does not vary across contexts and users. Thus, contextual cues of utterance, melodic aspects of the voice (such as rhythm and prosody), and certainly the accompanying gestures were dismissed from the framework of language production and comprehension except as peripheral information sources.

Shortcomings of the metaphor of speech as written language become blatant whenever the meaning intended by the user differs from the lexical meaning. Co-speech gestures constitute one of these cases, since they modify or complement verbal expression. This leads to the problem of combining two purportedly antinomical types of meaning: lexical and gestural (McNeill 1992). G-M&B contribute to close the artificial theoretic divide between objective meaning bearers (words and signs) and idiosyncratic and contextually dependent meaning (gestures). This divide, pervasive since Saussure's definition of *langue* as the subject matter of modern linguistics, fails to reflect that language, first and foremost, emerges organically among the humans who use it. The meaning of words is wholly dependent on "their always insistent actual habitat, which is not, as in a dictionary, simply other words, but includes also gestures, vocal inflections, facial expression, and the entire human, existential setting in which the real, spoken word always occurs" (Ong 1982, p. 46). Though the heritage remains tacit in the article, G-M&B are heirs to this contextual and organic conception of language; they reveal so by stressing that in real communicative acts, signs and words often behave like gestures and vice versa.

This tradition can be traced back to Wilhelm von Humboldt, who maintained that language "is no product (*Ergon*), but an activity (*Energeia*). Its true definition can therefore only be a genetic one" (Humboldt 1836/1988, p. 49). Language is not a set of prefabricated meaning units ready to be deployed for communication; one cannot replace the living utterance with tokens whose content has been described and preserved outside the stream of the real communicative situation. While the *language-as-action* tradition has known remarkable advocates during the 20th century – Peirce, Bühler, and Wittgenstein among the most prominent – it is in the writings of Heinz Werner, a lesser-known but crucial figure (Valsiner 2005; Wagoner 2013), where several of G-M&B's claims find the most relevant support.

Through a variety of experiments, Werner identified "physiognomic features" of words that grant them connotative values in a direct, immediate way (Werner 1978b). Just as when we see a face, we perceive words as hard or soft, sweet or dry, energetic or tired, and so on. Such physiognomic features of words do not correspond to their semantic representation; they rather include dynamic contents, imprecise but intermodal and synesthetic. Therefore, when the words "climbing" and "raising" are displayed centered on a monitor, they are perceived upwardly, while "falling" and "plunging" are perceived downwardly (Kaden et al. 1955). For Werner, the physiognomic qualities are what make symbol formation possible: "We submit that even the most conventionalized units of speech – words and sentences – are still part and parcel of an articulatory process, bodily postural activity, which, through its dynamic features, links those conventionalized units to their referents" (Werner & Kaplan 1963, p. 207). The pervasiveness of the physiognomic qualities consequently blurs the formal distinction between signifier and signified: "the 'arbitrariness' of linguistic forms seems to us to be unwarranted" (Werner & Kaplan 1963, p. 16). This is particularly salient in the case of gesture, since "the material moment of bodily posture and motion, and the moment of meaning, are an indissoluble unity, i.e., a gesture cannot be significantly analyzed into a bodily displacement and a meaning arbitrarily related to it" (Werner 1978a; 1978b, p. 424). G-M&B seem to share this

view, as they reveal when they approvingly cite Waugh's (2000) indictment of arbitrariness in a text that calls for the reconciliation of form and meaning.

But how are we to harmonize the call for unity that pervades the target article with the authors' claim that it is necessary to distinguish between imagistic gestures and categorical sign (or speech) for the sake of predicting learning potential? While there doubtless is enormous practical value in their experimental insight (Goldin-Meadow et al. 2012), it seems an insufficient ground on which to proclaim a univocal imagistic/gesture, categorical/sign mapping, and particularly so in a field that is just coming of age. Imagistic and categorical are not separate positions across a schism but rather the fuzzy endpoints in a continuum. The brain's responses to uni- and crossmodal mismatches are fundamentally the same (Cornejo et al. 2009; Kelly et al. 2004). As the physiognomic nature of words makes clear, imagistic properties are also to be found in linguistic formats. The imagistic-categorical distinction is not an empirical claim but an axiom that assumes that meanings are products instead of processes. It is a tenacious residuum of the inherited metaphor of speech as written language that persists in gesture studies.

Building a single proposition from imagistic and categorical components

Kathryn Davidson

Department of Linguistics, Harvard University, Cambridge, MA 02138.
kathryndavidson@fas.harvard.edu
http://scholar.harvard.edu/kathryndavidson

Abstract: Bimodal bilingual language provides further evidence for the viewpoint advocated by Goldin-Meadow & Brentari (G-M&B) that sign, speech, and gesture work together to create a single proposition, illustrating the potential in each set of articulators for both imagistic and categorical components. Recent advances in formal semantics provide a framework for incorporating both imagistic and categorical components into a single compositional system.

Goldin-Meadow & Brentari (G-M&B) highlight the relationship between imagistic properties and categorical properties in communication, focusing on how they emerge from and interact in speech, sign, and gesture. Indeed, it does seem clear that the appropriate comparison in forthcoming research on this topic should be the combinations of (a) speech and gesture with (b) sign and gesture, given increasingly sophisticated theoretical and experimental tools able to distinguish gesture and sign, and because imagistic gestures and categorical properties of speech and sign form complementary aspects of the communicative signal.

One important piece of evidence in favor of this view comes from a combination of sign and speech that was not discussed in the target article: sign-speech ("bimodal") bilingualism. Communication between individuals who are native users of both a signed and spoken language frequently involves natural "code blended" utterances (simultaneous signed and spoken analogs of unimodal bilingual "code switches") that exhibit aspects of the grammar of both languages. Studying code blends can provide unique insight into the ways that combinatorial ("linguistic") and imagistic ("gestural") signals can combine through oral and manual articulators, because in bimodal bilingual utterances each set of articulators has the potential to carry a full linguistic signal.

The flexible relationship between language and gesture is perhaps most clearly highlighted in code blends involving sign language classifier predicates, which are signs that involve a categorical/linguistic *handshape* that bears a grammatical/linguistic relation to the sentence's subject pronounced with a *movement and location* that have imagistic properties of gesture (see sect. 4 of the target article). When combined in code blends with a spoken language, these classifier predicates can either serve the place of a typical co-speech gesture (when the overall structure is based on the spoken language), or they can serve as a main predicate (when the overall structure is based on the sign language) with an accompanying gesture in the spoken language.

Consider example 1 below: in that example, English is the language providing the structure of the utterance and the mouth produces most of the words, while the hands merely provide an accompanying gesture but one that includes categorical, linguistic components from ASL (the classifier handshape for legs using the "V" handshape, CL-V). In another kind of blend, ASL provides the dominant underlying structure in example 2, and (whispered) English provides a sound effect – a verbal gesture of sorts – to accompany classifier constructions (contrast this with the English word *golf* that accompanies the nonclassifiers sign GOLF).

(1) *English speech*: And my mom's you know walking down.
  ASL sign: CL-V(walking down stairs)
                 (Emmorey et al. 2008)

(2) *ASL Sign:* GOLF CL-1(path of ball going up) BALL CL-1 (path of ball going up)
  English Whisper: golf (sound-effect) ball (soundeffect)
  "In golf the ball goes high up, the ball goes like this … "
                 (Petroj et al. 2014)

Both (1) and (2) are examples of co-opting articulators typically used for combinatorial information in each language, now for imagistic, gestural purposes. Both blends using classifiers support the view that the oral and manual modalities are each capable of providing either imagistic or combinatorial components; that the manual modality is sometimes considered to be primarily gestural is a result of the traditional focus only on spoken language.

Bimodal bilingual code blends also support a second claim from the target article: that multimodal utterances convey a single proposition. Although unimodal (sign or speech only) bilingualism leaves open the question whether with, for example, two mouths, a bilingual could or would simultaneously produce two separate propositions in two languages, in the case of bimodal bilinguals the answer is evident: Despite separate articulators, many studies of bimodal bilingualism have shown that the two channels combine to produce a single proposition (de Quadros et al. 2015). It is crucial, then, to understand how a compositional semantic system should handle all of the components of such a proposition, both the imagistic and discrete, in sign and/or in speech.

It is unfortunate that in the target article the authors discuss the difficulties of accounting for imagistic components of sign (and speech) at various levels of linguistic analysis: the phonological, morphological, and syntactic, but have no dedicated discussion about meaning (semantics). However, very recent theoretical linguistic advances within formal semantics and pragmatics have provided tools to address precisely this question of how to incorporate both the gestural and linguistic aspects of meaning in many areas of semantics, including binding, scales, anaphoric reference, speech reports, and sign language classifier predicates.

Classifier predicates have especially been the focus of one account that directly compares the sign and gestural components of classifiers with the combination of speech and co-speech gesture, in an implementation of one of the primary suggestions of the target article. This is accomplished within a formal framework by modeling the gestural component as a *demonstration*