

Barbara Dancygier
The University of British Columbia

THE STATUS OF MULTIMODAL CONSTRUCTIONS IN DEVELOPING MODELS AND EMPIRICAL TESTING -A MULTIMODAL SEMIOTIC PERSPECTIVE

The primary focus of this conference has been multimodality. The organizers also proposed some questions for discussion, helping us think about how the recent focus on multimodal aspects of communication relates to or changes our approach to language.

Like other debaters, I would like to tweak the questions a bit, to be clear about where my expertise and my convictions lie. For example, we were asked to comment about the existence of mental representations. Not being an expert in this area, I am cautious about sounding too confident, but I would still argue that some form of representation of concepts needs to be there somewhere, so that we can reach some stable information about elements of reality. The way I typically think about such issues is by referring to frames, in the Fillmorean sense – knowledge structures we use and access all the time; importantly, accessing an aspect of the frame (through an action such as reaching for the phone) equals accessing the entire frame (understanding the phone to be a communication device, acknowledging that phone communication happens across different locations and without face-to-face contact, etc.).

I also think consideration of multimodality adds some useful angles to the concept. Frames can be accessed via language, but also via embodied be-

havior, images, and aspects of sensory perception. Like language, an image evokes a frame, even when it is not a complete representation – so a child might draw a yellow blob with lines sticking out of it, and we do recognize it as the sun. In fact, sounds and onomatopoeic expressions are equally efficient in evoking concepts, which is why children are happy to call a train ‘a choo-choo’. Yes, such representations are schematic and partly conventional, and they can be further described as iconic or indexical, but they connect to some shared idea or concept – the frame. And they make the connection without relying on language only, even if a linguistic item is available.

We were also asked to comment whether representations are mono- or multi-modal. I am convinced that we have access to concepts that come with multiple aspects of representation – linguistic, but also visual, auditory, tactile, or related to a motor program – in other words, concepts are represented multimodally. This naturally means that frame structures are multifaceted and this complexity of representation is why frame metonymy is such a common trope – you do not need to represent a full frame in communication, but you may access the frame structure by being given a token, and the tokens can be of various kinds. Looking at various multimodal artifacts which rely at least partially on visual representation, such as ads or street art, I see a broad range of creative forms which evoke a frame visually, by showing an aspect of it (for example, McDonald’s ‘golden arches’ for the food served and business model used). But such artifacts also distort various aspects of the frame, in various modalities. They change size, color, shape or texture, to evoke familiar frames and yet communicate something new. They change form enough to draw attention to it, but not too much, so that the viewer or listener can still recognize it and construct meaning starting from a familiar frame. Importantly, those multimodal representations are so salient and stable that manipulating them does not cause lack of comprehension. Rather, what happens is construction of new meanings out of the existing ones.

Let me briefly clarify what I mean by frames being ‘stable’. The invitation to the debate suggests that representations need not be understood as stable patterns but as dynamic co-activation of neuronal assemblies. From my perspective, as a linguist rather than a neuroscientist, I view stability as the existence of frames which can be reliably shared by various speakers over a certain time span, even if the information in the frame changes. To return to the simple example of the ‘phone’ frame, contemporary phones are not rotary and do not have separate receivers – the material form and the attendant motor programs have changed. And yet, when one considers phone icons, they usually look like rotary phones or like receivers (check the on/off icon on your mobile phone, which typically looks like a receiver). So what is stable and what is dynamic, given how quickly the technology

changes? Much of the frame, such as the idea of remote communication, remains stable, and while the actual material or visual aspects change, the 'left-over' iconicity remains stable.

Construction of meaning thus requires that frame structure is largely stable, but dynamically used, and that at least some visual and enactive aspects are reliable. This is true about a number of communicative forms, but especially true about the forms which metonymically appeal to aspects of experience. One interesting example is color. For example, most Lego ads, however creative, will maintain the primary Lego colors; also, a McDonald's ad will somehow use bright red and bright yellow – both use color to evoke rich frames of the products the companies sell (toy plastic blocks a child can use to build things, or fast food). However, when Lego advertises the artistic potential of playing with the blocks, they show a picture imitating the Mona Lisa, made of Lego bricks, but in colors that Lego does not offer – closely resembling the tones of the Leonardo painting. So in one case, color evokes the familiar Lego set, in the other, it evokes the artistic potential. In both cases rich frames are evoked, with one aspect of the representation (color) highlighted in one case and tweaked in the other.

To continue with the Lego example, the various modalities of the frame have to work together to create meaning. Constructing things out of plastic blocks requires a mental representation of what the final product will look like, but it also requires a sense of how manipulation of smaller structures contributes to the emergence of larger structures. In a sense, the experience correlates with meaning construction – well defined building blocks can be brought together through embodied experience into a new, richer concept, complete with material, tactile, visual, and other multimodal features. The underlying multimodal nature of concepts is necessary for meaning to emerge.

Another question we were asked is about the connection between signs as discussed in semiotics and linguistic forms. The example of a map is very useful here – it is a visual form, and yet makes navigating space possible, without using language. I think the question is tapping into further aspects of multimodality, especially visual representation. Before I respond, I'd like to clarify my assumptions.

There is a bit of a theoretical and terminological overlap that is often taken for granted, but requires clarification. For a semiotician, language, as a symbolic system, exists naturally in the broader context of various signs, including icons and indexes. The difference between language and other signs matters in this context. But it seems to me that the need to make such distinctions is less obvious if we consider the role of conceptualization, and the scope of the terms used. Recently, multimodality has become a term that covers a lot of ground, as it refers to many aspects of cognition and communication: there is 'multimodality in interaction', that is, gesture,

eye-gaze, etc., and there is interaction between language and other semiotic channels, such as image or sound; the latter trend is particularly interesting to semioticians, and that's where the issue of whether semiotic or cognitive linguistic interpretations are similar or different is the most pronounced. The way the question has been formulated (especially via the map example) seems to suggest that the division between multimodal embodied interaction and multisensory artifacts is worth maintaining. I actually disagree. The reason why any such division has to be misleading is conceptualization – the crucial aspect of any form of expression. Cognitive linguistics gives priority to conceptual structures, over specific forms of expression, and approaching both 'multimodalities' from that perspective obviates the need for ad hoc distinctions. Multimodality can be approached from the perspective of communicative load-sharing, and the various embodied or semiotic modalities can interact with language in a number of complex ways.

I will return here to the examples I discussed in my talk, which I described as 'street deixis'. In my neighborhood, a flower shop has recently closed down, leaving a large sign in the window: "Thank you, Kitsilano!". Language is used here in a rather standard form of giving thanks, but that's the only typical aspect of this utterance. The addressee, ostensibly the whole neighborhood, is any inhabitant, any passer-by, any person walking the street. In fact, walking is not enough, the passer-by has to stop, use eye-gaze to focus on the window (rather than another person), its message, and the sign. The 'speaker' here is the business, the flower shop, which is still, in the mind of the passer-by, metonymically associated with the particular location, while in fact no longer occupying it. And so on, and so forth. Deviations from ordinary speech acts of giving thanks are numerous here, and so profound that they should invalidate the act altogether, and yet all works fine. I would argue that it does because all the multiple channels of communication, all the multimodalities involved, linguistic, deictic, visual, and embodied, work together on the basis of familiar concepts. It is quite clear how we need stable frames and conceptualizations to interact with such a sign, especially those pertaining to ordinary language interaction and the pragmatics of speech acts, but also to doing business in a neighborhood, moving along the street, turning to look at a shop window, etc. Conceptualization holds this communicative event together (though only for the brief moment of someone stopping and reading), while the nature of specific expressive modalities is adjusted to the situation. Meaning construction mechanisms take over, regardless of the modalities used. What is necessary is the mental representations in the mind of the passer-by, who will be transformed for a brief moment from just a person walking by to a participant in a communicative act. Interaction with the modalities is what holds such acts together, regardless of their semiotic nature.

So, to conclude, the distinctions between linguistics and semiotics are only important if we look at the form of expression, but they disappear when we look at the paths of conceptualization. Conceptualization patterns are the great equalizer here. And we need to add that cognitive linguistics has developed an impressive toolkit to describe conceptualization, so it can now apply it to multimodal artifacts. In fact, as I argued in my talk, language use adjusts to the context of multimodal artifacts, and so even if we focus on language alone, we can learn much from multimodal usage. Some observations would simply escape us if we insisted on dividing lines based on form.

Besides, talking about signs seems to assume a clear relationship between form and function. But we have to consider that, to give just one example, the increased use of images, whether iconic or symbolic, is communicatively a relatively new thing, and new forms are emerging all the time. Are these innovations signs? Or are they simply emerging conventions which are subject to change as they evolve to fit new communicative needs by using new available means? Internet memes are a good example here. New ones are emerging all the time, and old ones are being adjusted and changed (for example, it used to be one image macro and text, but more and more often the image macro is having other images superimposed or photoshopped onto it). It seems to be more productive to focus on conventional usage and its emerging constructional features than to search for semiotically satisfactory definitions. As I said earlier, cognitive linguistic tools are very well suited to capturing the emergence of new forms and meanings, and correlations between them.

The next question asked us to reflect on whether language is independent of other modalities. I think the comments I made earlier make it pretty clear that my answer is “No, absolutely not.” Language is immersed in all other modalities, in embodied interaction, in visual, auditory and tactile perceptions. And, in action. This is because language participates not only in embodied cognition, but also in enactive cognition – and these areas of cognition pertain to how we rely on our bodies to communicate things, but also to do things.

The next question that matters a lot here, and which has been raised, is methodological – if language is so deeply multimodal, how are we going to test our hypotheses? How do we gather data? How do we process data when the data is “everything”? We have developed various methods of enquiry, but they are not so open-ended as to cover all the aspects of multimodality that I mentioned. Representative examples can be gathered, but what tests can we apply to them? Given that the answers are not easy to propose, what should we do? I can just feel the dreaded ghost of ‘introspection’ looming over us. Well, I’m actually not afraid, because I feel that ‘introspection’ is nothing more than a spectre some of us resurrected from the olden days of

armchair linguistics. What we do when we gather examples to support hypotheses, even if we cannot provide sufficiently rich corpora, is we analyze. As long as our examples are attested and as long as there are sufficiently many of them that we can recognize patterns, we are being analytical, not introspective. Perhaps with time all forms of multimodal communication will lend themselves to corpus and quantitative methods. The technology is out there and it will support our efforts. But meanwhile, we have to go on looking for patterns, ideas, and hypotheses. There are presumably many ways to understand multimodal behavior and communication, and we have to try them all. We cannot limit our enquiry to things that the methodology we ourselves have designed allows us to do. The methodology will catch up, if we give it a chance.

Thank you!