Dylan Glynn
The University of Paris 8 Vincennes - Saint Denis

The status of multimodal constructions in developing models and empirical testing - a methodological perspective

I'm not really a specialist in multimodality, even if I'm extremely interested in it. I have doctoral students who work on it, but I wouldn't call myself an expert by any means. If I were to consider myself an expert at all, it would be in methodology for linguistics and I expect my contribution will be from a methodological perspective. Professor Fabiszak kindly sent us four questions and last night, and before the dinner, I sat down and tried to think my way through them. I will try to answer the questions but in reverse order since I felt the first questions were the more difficult and important and the last questions were less challenging.

The last question is more or less: *Is language independent from other communicative systems such as gesture?* I hope this is a reasonable summary of the question. As a cognitive linguist, the answer is obviously and emphatically *no*. In fact, I feel the question would best be turned around: why would language be independent from other communicative systems? If we accept Occam's Razor (the principle of simplicity in scientific explanation), one would need an empirical reason (an observational reason) why one would want to posit the existence of a theoretical construct such as an independent faculty of language. Now, I'm not saying there aren't good reasons to posit this. There is research beyond my areas of expertise, in aphasia for example, that

https://doi.org/10.32058/LAMICUS-2019-008

may warrant such a proposal. However, clearly the onus should be on those who believe that we need such an understanding of language to argue for it. Therefore, the question should rather be, are there arguments for understanding language as an independent communicative system?

I do wish to add a few caveats to my initial categorical response. Firstly, this position does *not* entail that various structures in semiosis, syntax, lexis, facial expressions and so forth, are not highly specialised. However, highly specialised is different from being independent simply because highly specialised does not preclude interdependent structure with other cognitive capacities.

Secondly, I can think of various reasons for analytically *treating* language as an independent system, even if it is *understood* theoretically as part of a broader set of communicative competences and structures. The most obvious reason is the complexity of the system. In science, it is widely believed that it is *better to say more about less than say less about more*, a maxim with which I could not agree more. However, given the fact that previous approaches to language (that have treated it as an independent system) have yet to adequately account for it in those terms, despite the daunting complexity, the Cognitive Linguistic 'holistic' approach may well be necessary. As scientists, we need to limit our object of study but not at the expense of being able to explain it. Of course, that in turn places the onus on those who seek to explain language as part of a holistic semiotic-communicative system to demonstrate how that can be done rigorously, i.e. adhering to scientific principles such as falsifiability. It is precisely this challenge that I believe is the most important challenge for Cognitive Linguistics or any functional theory of language which holds that an explanatorily adequate approach cannot afford to delimit its research to language as an *independent phenomenon*.

To summarise my response, I can say that *no*, language should not be *theoretically understood* as an independent system unless you can give me a good reason otherwise. Moreover, it should not be *empirically treated* as such despite the fact that this renders our object of study exponentially more complex.

I believe question three can be summarised as: *Are linguistic semiotic signs useful ways of describing linguistic behaviour?* However, I am concerned that I have misunderstood this question. In light of my concern, I'll do my best to answer both the literal question and what may be implied by that question. To begin with the literal interpretation of the question, let me turn to something I discuss when teaching. Each year, I ask my first-year students to think of something that they cannot name. Of course, they are stumped and are never able to think of something that cannot be named. We conclude that to think of something, to conceptualize something, one categorises it and by doing so one creates or, indeed usually, assigns a predetermined form-meaning pair, a semiotic sign. In our classes, we call this

cognising. It is from this point that we introduce the theory of "*Cognitive* Linguistics" – language understood as experience-based conceptualisation.

In class, to underline the extent that we cognise our experienced world, we consider two examples. Firstly, we take the example of a young child when he or she sneezes. I recount a story, where when I first moved to Hungary, I was surprised that the children would make the sound *Haptsi* when sneezing. I observed that even children in the relatively early stages of language acquisition would "say" *Haptsi*. No one in my classes "says" *Haptsi* when they sneeze. They say something like *Achoo*, as in English. Similarly, a second example is the verbal response to accidently burning oneself cooking in the kitchen, where a French speaker will "say" *Aaii!*, an English speaker will "say" *Ow!*. In these instances, we are observing behavioural responses, not lexemes in any traditional sense of the term. Yet we see that they are, in fact, form–meaning pairs. In other words, they are linguistic signs. In the microseconds between experiencing pain or the desire to sneeze and uttering the verbal response associated with those experiences, we cognise the experience and ascribe a learnt semiotic sign to it. From this, we conclude that it is very difficult for us to think or behave without cognising.

Given that I take this response to the literal question as self-evident, I expect the implicature in the question was about objectivity and the inherent circularity that results from using the sign-system of language to describe language. Perhaps, the implicature is, therefore, that we should use logic or mathematics instead of signs to describe communicative events. If the question implies the use of logic, as in truth-based reason, then I have yet to see an adequate description of linguistic behaviour using such reasoning. Given that this is a Cognitive Linguistics conference, I find it unlikely that this was the intention of the question. However, the convenors know that I specialise in quantitative methods, especially the application of quantitative methods to the slippery world of qualitative data and linguistic behaviour. Therefore, perhaps the question implies the use of mathematics.

In response to this, I would say that statistics, or mathematics applied to describing the probability of events in the world, is also predicated on signs. In Fig. 1, you see the mathematical equation that is linear regression, a widely–used statistical method in linguistics.

Figure 1. Linear Regression

$$y = k + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

Obviously, this is a mathematical formula, written using symbols. In Figure 1, the y signifies the dependent variable, the k signifies the intercept and the

β signifies the predictor. There exists a well-worn debate in mathematics as to whether formulae are discovered or invented. Does the mathematician discover something in the universe and ascribe signs to it or are the formulae creations that represent (signify) how humans can explain the universe? Such debates, however, do not concern the use of applied statistics. In the formula presented here, the symbols are form-meaning pairs. In this instance they represent variables and, in any given application of this formula, these variables are directly or indirectly quantifiable phenomena – things in the world.

In summary, other than perhaps logic which has been systematically shown to produce inadequate descriptions when applied to *parole* / performance, I can think of no way we could describe linguistic behaviour without using signs. Perhaps a more fruitful question would be: *How can we describe linguistic behaviour in a way that produces falsifiable results*? Taking this further, is it possible to use quantitative methods to help determine descriptive and predictive accuracy of those descriptions? Arguably, these questions are more pertinent than whether the use of linguistic signs to describe linguistic behaviour is appropriate or "objective".

Let us turn to the first two questions, which were, in my opinion, the most important.

I found that question two was full of many implications and entailments, and so I broke it down into what I thought were the three principal sub-questions:

> *Are multimodal interactions understood as mental representations*?
>
> *Are these representations mono- or multimodal*?
>
> *Are these representations best understood as stable patterns on neural assemblies*?

Firstly, *can we understand multimodal interactions as mental representations*? My initial response is: What are mental representations? How are they operationalised? Are we speaking about Chomsky's trees or Langacker's pictograms? Do we mean instantiations of schematic relations? I know that as a linguist, representations are constructs we use to help us talk about an object of study that is not observable. I think perhaps this last point is crucial. So leaving aside a meta-discussion on whether representations are scientific or the empirical question as to whether they exist as part of an individual's linguistic competence, representations can be understood as analytical tools. I hope we can all agree that we don't have pictures in the brain and that image schemata and grammatical pictograms don't exist anymore than X-bar patterns exist. If we do all accept that point, then representation is understood firstly as an analytical tool. In order to appreciate whether representations are applicable to *multimodal interactions*,

I would prefer to think about representations operationalised as *generalisations across usage*. It should be evident that I am here evoking Langacker (1987) and Hopper's (1987) Usage-Based Model where structure emerges from usage just as a path emerges across a frequently traversed field or forest. If one accepts this operationalisation of representation, then of course generalisations can emerge from *multimodal interactions* and of course they can be described in terms of representations. Whether speakers possess these generalisations and employ them in semiosis is an empirical question. Personally, I find Dąbrowska's (2004, 2008) position and evidence in this regard the most convincing. This position can be summarised as: sometimes yes and sometimes no because it can vary from speaker to speaker and even from time to time for a given speaker.

As for the next question – *Representations, whether they are monomodal or multimodal*. We are cognitive linguists, everybody here, so we treat language structure in a non-modular holistic way. That is about as fundamental as one can get to the theory of Cognitive Linguistics. I expect the text-book response to this question would be: if, for the individual (and, by extension, the speech community), the extra-verbal sign (an emoji, a gesture, a raised eyebrow *etc.*), is an *ad hoc* or infrequent addition to a type of verbal component in communication (a morpheme, lexeme or construction), then it is probably treated as context information that we use in processing, combined with encyclopaedic semantics to help understand the communicative event. In such situations, one would speak of mono-modal representations (or generalisations). However, if, for the individual (and, by extension, the speech community), the extra-verbal sign is frequently or systematically associated with a type of verbal communication, then it will become entrenched as such. In these situations, one would speak of multi-modal representations (generalisations). As a cognitive linguist, I typically employ the term *construction* to refer to the *Gestalt* of any complex for form-meaning pair, mono-modal or multi-modal. I am confident that this would be unproblematic for any linguist that agrees with the tenets of the cognitive paradigm. Prosodic marking, facial expressions or even memes, indeed any form-meaning pair, is a potential component of a construction. Where construction is understood as a *representation* of a complex form-meaning pair constituting a part of a speaker's linguistic competence.

Finally, the last sub-question concerned the *co-activation of neural assemblies versus stable patterns*. As an empirical quantitative linguist, my immediate response is: to the best of my knowledge, our understanding of synaptic connectivity, plasticity and organization (neural assemblies) still offers no data that facilitate the description of conceptual structure – representations. For this reason, I will side with stable patterns. Indeed, the notions of patterns and stability are central to the usage-based model of language. With

stability, we are talking about the relative consensus between speakers on the language system at a given place and time. Put in other terms, this is the degree of conventionalisation in language. This can be operationalised in terms of the predictive accuracy derived from relative frequency or relative salience or, ideally, a combination of the two (cf. Arppe *et al*. 2010 for discussion on this point). Pattern is a notion that we use to capture both the idea that the structures we identify are not discrete and that they are multidimensional (complex) in nature. Put simply, instead of rules, which are by definition discrete and which tend to lead people to think about structures independently from one another, we speak of patterns. Patterns, even complex patterns, can of course be operationalised in terms of probabilities making the notion the perfect tool for speaking about language structure from a usage-based perspective.

This response deserves a simple caveat. I do not wish to suggest that the study of neural phenomena cannot inform linguistics. At some stage in the future, I believe it may be crucial in our efforts to fully explain conceptual structure. Nevertheless, from my perspective with a limited understanding of the field, there remains a large gap between the study of neural structure and anything that we would call conceptual structure.

The first question is perhaps the most difficult, but perhaps also the most pertinent: *Multimodality is not only a feature of spoken communication. Written language and other non-spoken modes are multimodal. What does this mean for our understanding, investigation and modelling of language*?

As somebody specializing in methodology, it was the last bit that caught my attention: what does this mean for investigating and modelling language? If we believe that language has a multimodal potential, which I assume we all do as cognitive linguists, then we must extend our model to include it. Given that we believe that in order to explain language it must be treated holistically but also that the language faculty is part of a general cognitive capacity, this much should be reasonably obvious in Cognitive Linguistics. Nevertheless, with respect to this last point, I do have some opinions that may be more contentious and therefore more useful to discuss.

Let us begin by considering emojis. Language is a systematic, structured, semiotic system, but it is not closed nor is it static. Via the iconic symbols that made up early writing systems, the use of images in language dates back to the origin of writing. Similarly, we all know that using an exclamation mark or a question mark in written text or an emoji in an SMS is highly structured, is part of the grammar of language. Interestingly, one must assume emojis and punctuation originated primarily as a means for expressing information canonically encoded by prosodic patterns and facial expressions (and perhaps some other gestures). If there were ever doubts that gesture and prosody were beyond the realm of grammar, such an observa-

tion should put such doubts to rest. Given this, why not simply extend this principle to memes or any other semiotic production, simplex or complex? What I believe is at stake here is the question of the systematicity of the form–meaning pairing (what cognitive linguists would call entrenchment and functional linguists would call automatisation) versus *ad hoc* form-meaning pairs produced in context and dependent on that context.

In other words, the problem should not be whether language is multimodal or whether emojis have entered language as a structured part of that language, but determining which of these non-verbal types belong to speakers' competences and which do not. The problem is rendered all the more difficult because the distinction between entrenched and *ad hoc* signs is non-discrete, varied and dynamic, between individuals but also for the individual. Let us consider this problem more closely.

Although we speak of languages, assuming the usage-based model, we are, in effect, always making generalizations across individuals in a speech community. Variation between individuals is an inherent part of the system and the reason we speak in terms of patterns not rules. However, it is also possible that, for a given individual, a sign (form-meaning pair) may be entrenched at one stage of his or her life, yet at another stage, it may not be entrenched (cf. Dąbrowska 2008). Moreover, the notion of entrenchment, even for the individual at a given moment in time is typically understood as a matter of degree. This amount of systemic instability when compounded by the fact that individual types (signs) in non-verbal communication can be extremely difficult to isolate as well as the fact that we lack transcription traditions for most non-verbal types, makes the issue of scientifically integrating extra-verbal communicative structure, such as emojis, into language study a nontrivial methodological question. A few examples should make this point clear.

Anybody who has received an SMS from their grandmother will probably have found that she sounds a little stiff, as if something unspoken is amiss. The communication event is failing because one is expecting visual representations of certain interpersonal things that we do with emojis and punctuation marks. That's multimodal grammar at work. The grammatical acceptability of the SMS from one's grandmother is low because she omitted the little smiley at the end or a wink after her sarcastic comment. One would expect that modelling this, just like we model the pragmatics or inter-personal semantics of lexis, should be straightforward.

The problem is a result of a lack of stability. Although phonetic and even phonological variation is part of verbal communication, relatively speaking, both the form and the entrenchment of the signs are stable. This is not the case for emojis. Between different computing platforms and in different contexts from the old school smiley your primary school teacher added to her marks in pen, via the keyboard representations for that venerable tradition

(such as :) or :P) to the animated gifs in modern chat, the formal variation is immense (cf. Tabacaru & Van Der Mark 2019). The analytical tools we have developed for language handle this degree of formal variation poorly (cf. Glynn forthc. for a discussion how we may overcome this problem). Furthermore, in the case of emojis at least, variation in the degree of entrenchment is substantial. The case of one's grandmother writing an SMS is an example in hand. For many people who do not regularly use computer-mediated communication, an emoji may be simply an *ad hoc* sign, processed actively using the full gamut of cognitive competences at the speaker's disposal. It must also be remembered that this dynamic complexity in stability of form and degree of conventionality impacts upon language's two most complex phenomena: semasiological variation (polysemy associated with the form) and the combinatory effects upon that semasiological variation (grammatical constructions). Lastly, of course all this complexity and variation needs to be sensitive to stylistic and other sociolinguistic effects. None of this is to say that we should not attempt to integrate such communicative elements into our models, but it is to say that it is an extremely difficult task.

Of course these methodological hurdles are not restricted to emojis. Actual facial expressions for which the emojis stand as symbolic graphs produce different challenges. Unlike emojis, facial expression is so systemically integrated and preconscious that, sufferers of autism aside, we can be confident that their role in communication is firmly entrenched in all speakers' competences. Whether it is recognising irony or mirativity or many other basic functions of language, facial expressions are a frequent and essential part of language structure. In this regard, facial expressions are more like prosody than emojis. Just like prosody, the identification of types and then, in turn, the tokens (q.v. Johanna Kißler uses the term "datums") of those types is an extremely difficult endeavour. With an emoji, the identification of type and token is relatively straightforward, where what constitutes a given facial expression or prosodic pattern (type) and then identifying occurrences (tokens) of that facial expression or prosodic pattern systematically is a fundamental problem for research in gesture and prosody.

When we extend these questions to complex novel signs such as memes (q.v. Barbara Dancygier) we are faced with both the kind of complexity and instability we observe with emojis as well as the issues of type and token identification we face in the study of gesture and prosody. Is a given meme (type) widely entrenched in a language system? How do we deal with the many variations of that type – is it one type with many tokens or many different types? Extrapolating for the discussion above should be self-evident.

Therefore, at least within cognitive linguistic circles, the question should not be *if* we should integrate multimodality in our descriptive, predictive and explanatory models of language, but *how* we can collect the data to attempt

this. For me, this is primarily a methodological challenge. We have a long and venerable tradition of identifying phonological, morphological and lexical types and tokens just as we have a reasonably well-developed tradition of identifying constructions, but multimodal types and tokens are still new. Put simply, until we have established analytical tools for identifying an occurrence of the "throw gesture" as opposed to a non-semiotic physical response to some random stimulus, I believe we need to focus on the methodological challenges. For memes and emojis, we need to address issues of variation and stability, operationalised in terms of frequency (corpora) and salience (experimentation). For prosody and for gesture, we need to operationalise the notion of type to permit the systematic identification of tokens of that type. Without reliable ways of collecting data, we will ultimately not be able to integrate these communicative dimensions into our endeavours to explain language. These issues are non-trivial, but given our understanding of language, I believe it is essential that we confront them.

References

Arppe, Antti, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert, M. and Arne Zeschel 2010: Cognitive Corpus Linguistics: five points of debate on current theory and methodology. *Corpora* 5, 1-27.

Dąbrowska, Ewa 2004: *Language, Mind and Brain: Some psychological and neurological constraints on theories of grammar.* Edinburgh: Edinburgh University Press.

Dąbrowska, Ewa 2008; The later development of an early-emerging system: The curious case of the Polish genitive. *Linguistics* 46, 629-650.

Glynn, Dylan Forthcoming: Emergent Categories: Quantifying semantic distinctiveness and similarity in usage. *Contrast and Analogy in Language: Perspectives from cognitive linguistics*. In: Marcin Grygiel, Barbara Lewandowska-Tomaszczyk (eds.) Amsterdam: John Benjamins.

Hopper, Paul 1987: Emergent grammar. In: *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*. 139-157.

Langacker, Ronald 1987: *Foundations of Cognitive Grammar,* Vol. 1. *Theoretical prerequisites*. Stanford: Stanford University Press.

Tabacaru, Sabina, Sheena van der Mark 2019: Crosslinguistic perspectives on the use and meaning of emoji in Asia and Europe. The Fifteenth International Cognitive Linguistics Conference, 06-11 August 2019, Nishinomiya, Japan.