Johanna Kißler
Bielefeld University

The status of multimodal constructions in developing models and empirical testing - a psychological and neuro-cognitive perspective

The previous debaters have already introduced and highlighted important issues. Therefore, I will try to touch upon already articulated viewpoints in my own comments as I address the questions presented to us:

One question regards the breadth or inclusiveness of cognitive linguistics and language studies, asking **should multimodality be considered as a part of language, should gesture be a part of language, should picture media, and even artefacts seen in street communication be a part of language?**

My intuitive answer is yes to all of these. Yes in that these are widespread, related, interesting, and important communicative media and phenomena worth studying scientifically. From my own point of view, I would also like to include emotion to that list, another phenomenon important in communication that is typically treated as extra-linguistic and that is dear to my heart. In fact, it is hard to imagine any real-life language discourse devoid of emotion. So much so, that there have been heated debates as to whether emotion should not really be called communication. Obviously, allowing for such breadth brings up a definitional issue, where each field needs to decide and to a large extent also negotiate the scope of its research. In broadening language studies to include all kinds of old and new "multimodal" commu-

nication phenomena, one incurs the risk of definitional, and to some extent also methodological, vagueness and arbitrariness. What kinds of approaches, methods and tools should be used for this type of analysis?

In linguistics, I believe, a great leap forward has been made by the "Chomskyan revolution", with its focus on syntax and combinatorial rules. As great as this leap has been, providing among other things a basis for now-adays widespread automatic language processing by machines, it has also resulted in many self-limitations. Opening up the field via cognitive linguistics and perhaps also more generally communication studies therefore has borne and still bears great promise. Combining the virtues and rigor of the Chomskyan approach with the open-mindedness of today's Cognitive Linguistics, as I perceive it, and as I have been enjoying it during this conference, could greatly advance the field. In practice, this would mean specifying the combinatorial rules of multimodal communication, its syntax so to speak. A gargantuan task! To achieve it, new computational approaches, drawing on statistics, big data and machine learning techniques seem highly promising. Dylan Glynn has already talked about regression approaches, and also about classification of types and tokens which would be one basic requirement in this context. Computational linguistics and computer science approaches in general may be helpful here and may make the seemingly gargantuan task of generating a multimodal communication grammar considerably more manageable. On the commercial side, companies like Apple or Google have provided an impressive, if somewhat terrifying, proof of concept in this regard.

Classification of types and tokens as addressed by Dylan Glynn provides a good link to another topic that was brought up, namely **the issue of representation**. How, where, and in what format do we store knowledge and combinatorial rules? As a psychology student many years ago, I was overwhelmed and somewhat irritated by the concept of representation, since it seemed both highly abstract and at the same time suggested a physical reality that I could not wrap my mind around. Only much later – comparatively recently in fact, in working with computer scientists – have I come to fully appreciate the usefulness and necessity of representation for modelling purposes. As a psychology student, however, I was deeply confused by it and started to look for something seemingly more real and tangible. I hoped to find it in studying the brain.

Today, many years later, I am of course aware that neuroscience is a different level of analysis and that reduction will not always buy you a lot. Whether brain science as a reductionist approach to cognitive or communicative phenomena can provide answers will depend a lot on the questions asked. In general, we can reduce our sensory impression of, for instance, the color green to chemical phenomena on our retina and to the activity of light waves, fundamental physical phenomena. However, along this reductionist

pathway, one can easily lose track of the original question and its adequate level of analysis. After all, not all of us want to end up as astrophysicists. We may all lie awake at night and wonder about the origin of the universe, but we may do so in different ways, leading us to pursue different research questions and different career paths during the day.

So: **Are there mental representations?** I would say: not really in a fully-understood physical sense. Still, my closest approximation of representation from a biology-inspired point of view, is a view of representations as dynamic co-activations of neural systems on a millisecond scale that ultimately give rise to semantics and syntax. Here, I would fully agree with the organizers' proposal. I also believe that neuroscience has made considerable progress in understanding how this happens. This is a level of analysis that has, over the years, come close and dear to my heart, not very surprisingly after all.

Because it falls broadly in my realm of expertise, I also want to bring up the concepts of **multimodal neuronal assemblies and multimodal integration** and what these may or may not buy you regarding representation. Of course, understanding neuronal assemblies will not result in direct understanding of interactive behavior of people in a room. However, it can reveal some principles that can be useful for studying language, communication and interaction. For instance, we know that there are initially separate processing pathways within our brains and that there are dedicated areas that integrate information coming from these pathways. We can study the integration of the different pathways and see how they weigh auditory and visual input, how they weigh verbal input, how they weigh tactile input. In short, how they weigh different modalities against each other. For instance, the human brain tends to give a lot of weight to visual input over other modalities, this is a known fact. It also typically weighs pictorial input over verbal input. Or, as the saying goes "a picture is worth a thousand words". This is something that Barbara Dancyger has also touched upon. In episodic memory, there even is a "picture superiority" effect. Furthermore, in interaction and memory, there is also an "emotion superiority effect"– emotion weighs heavily on our minds and biases us in many different ways.

Still, we all come from a heavily verbal tradition. Our cultures have historically developed to be more verbal than anything else. This is for a good reason. Language allows us to communicate via a sparse, simplified code. Therefore, a word can also be worth a thousand pictures – or abstract from a thousand pictures. This is, in my view, because in the past non-verbal and non-language based representations were expensive. It was costly to produce pictures and difficult to transmit them. Until recently one could not just make an audiotape of what people said, how they said it, or video-record their facial expression while they spoke. Instead, humans used a verbal code

when they sought to conserve. Initially via oral history, and since its invention, mostly via written code, which we still rely on so much.

But these verbal and written codes are not what, I believe, we as humans had originally evolved to process. Therefore, as soon as all these multimodal artefacts have become available and are thrown at us everywhere, we are led and potentially also mislead, by their seeming veracity and the fact that they feed into our biologically older processing systems whose input we weigh over the more recently developed linguistic, and particularly the written code. Oftentimes, if we have video recordings of speech or conversation available, we give more weight to the facial expressions than to what is said and we will in my experience and opinion typically weigh prosody over the content. Therefore, yesterday it surprised me to see a presentation that argued against this. In general, I believe that there are principles from neurobiology that we should consider and draw upon to explain phenomena that other people identify in the streets or in art history, or cultural history, or nonverbal communicative behavior.

Although I perceive neurobiology as a useful and recently reasonably successful level of analysis, it clearly need not be the only one. Representations are useful concepts to think about, and they are necessities for computational analysis, where their likely neurobiological substrate is still largely disregarded.

I am afraid I do not have much to say about the status of semiotic signs and their analysis-perhaps other than that to me semiosis in itself seems both too static and too broad to capture the dynamics of ongoing communication. However, I would like to remark on a different, but in my view related issue: So far, we are lacking a good mechanistic understanding of the many levels and dynamics of cognitive representation and I am not even sure whether the concept of "mechanistic knowledge" will seem desirable to the present audience. Still, many successful branches of (natural) science view mechanistic understanding as an important goal, perhaps the ultimate goal of their research endeavors.

As with the question of multimodality, the questions of representation and level of analysis are to a considerable extent questions of definition. Within the conference, but also with this debate, I thought there were many such questions of definition. I attended a presentation yesterday that was entirely devoted to definitional considerations, and I found this was interesting, because I come from a branch of academia where theoretical issues of definition have been largely replaced by **operationalization and testability**. Operationalization and testability loom large in my scientific world, and I believe that cognitive linguistics as I experienced it in the past three days could benefit from moving in that direction, ideally without losing its own elaborate theoretical foundations.

I fully advocate that there is also the issue of induction, of generating ideas. Here, strictly experimental approaches arguably may have thrown out the baby with the bathwater. Focusing only on testability and getting rid of induction, getting rid of generation of ideas, so to speak, results in stark and possibly harmful limitations. There is no need to shut out the "real world", just walking down the street and looking around, looking at phenomena in the real world, can be an important part of induction.

Having said that, I think it is equally important to look at the status of data and decide: **Is this a datum? Is it an example?** In many cases in what I have seen in the past days, what I have seen called data, to me are examples. There is nothing wrong with these observations per se. Still, from my point of view, it is important to clarify their status and to determine to what extent an example really "transcends" to a different level – either because it points to something very common – which could be determined quantitatively or to something very powerful, a new mechanism, a new dynamic mapping between symbols, for instance. In the end, it is up to each discipline to decide what is to be treated as evidence.

On the other hand, personally, I very much regret that psychology, the discipline that I come from, in spite of being viewed by de Saussure as a natural point of departure for language studies, recently has not been interested very much in the study of naturalistic language-use or communication. Exceptions exist regarding issues like the experimental processing of pitch, faces, words, and the likes, or how these are integrated in the brain. These are all important. Still, a more macroscopic or even "holistic" look would also be desirable and nowadays the methods are there to capture and analyze larger, more realistic and therefore often multi-modal, units without losing control, without falling back to "armchair philosophy". An almost exclusive focus on micro units seems odd for a science that claims to aim for an understanding of how people behave. Psychology is the "study of human experience and behavior". Multimodal interaction clearly takes up a considerable part of human behavior. Therefore, a joint, truly interdisciplinary and multi-perspective effort may in the long run turn out to be most fruitful to understand the different aspects and levels of multi-modal communication.