Language & Communication 54 (2017) 73–81

Contents lists available at ScienceDirect

Language & Communication

journal homepage: www.elsevier.com/locate/langcom

Bodily constraints contributing to multimodal referentiality in humans: The contribution of a de-pigmented sclera to proto-declaratives

Juan Olvido Perea García*, Katrine Rosendal Ehlers, Kristian Tylén

Aarhus Universitet, Nordre Ringgade 1, 8000 Aarhus C, Denmark

ARTICLE INFO

Article history:
Available online 26 November 2016

Keywords:
De-pigmented sclera
Multimodality
Referentiality
Cooperation
Eye

ABSTRACT

We present the results of an empirical study that measured the contribution of a conspicuous eye-gaze (as a function of scleral de-pigmentation) of humans in conveying multimodal referentiality by combining visual and auditory cues in a naturalistic setting. We made participants interact in a cooperative task in which they had to convey referential meaning about co-presential entities. In one of the conditions, participants had no access to their interactants' eye-gaze. We interpret the results as supporting the idea that our eye morphology contributes to instantiating multimodal referentiality in cooperative tasks in peripersonal space.

1. Introduction

Purely verbal discourse without accompanying extralinguistic cues is a relatively recent cultural invention associated with the rise and expansion of writing systems, as well as a rather infrequent phenomenon in contexts of social interaction (Linell, 1982). In most of human interaction, linguistic cues typically co-occur with a rich variety of signals of a different nature, such as body orientation, facial expressions, eye-gaze, and hand gestures. This has led researchers from a range of fields to describe linguistic interaction as an inherently multimodal activity (Kendon, 2011), in which linguistic cues and other semiotic resources interact and co-evolve – on both phylogenetic and ontogenetic timescales – ratcheting on each other as they entwine in ever more complex patterns of expressive behavior.

Bearing on the assumption that gestural communication paved the way for vocalizations to convey referential meaning in our evolutionary path, most efforts have focused on addressing the interaction between hand and bodily gestures and vocalizations (Corballis, 2002). However, the link between gaze-cues and vocalizations remains largely unexplored, even though it has been considered pivotal to the ontogenetic development of linguistic abilities (Dunham et al., 1993; Mundy et al., 2007; Tomasello and Farrar, 1986; Tomasello, 2009; Baldwin, 1993). To what degree do humans actually rely on gaze-cues to jointly establish shared reference to entities in the immediate environment? As a first step towards a satisfactory answer to this question, our study sets out to test the assumption that linguistic and gaze-cues constrain each other in interaction. In this article, we present an experiment in which we measured the distinctive contribution of gaze-cues to the instantiation of referentiality. In a controlled experimental environment, we simulated the physical and interactive affordances of a naturalistic setting in which our hominin ancestors would have relied on gaze-cues as an efficient pointer – activities such as

* Corresponding author.
E-mail addresses: juan.olvido@gmail.com (J.O. Perea García), karo@cc.au.dk (K.R. Ehlers), kristian@dac.au.dk (K. Tylén).

http://dx.doi.org/10.1016/j.langcom.2016.10.007
0271-5309/

cooking, tool-making, and hunting – that is, cooperative tasks in peripersonal space. We focus especially on the affordances of the eye to indicate a gaze direction for someone to follow. This is because, among many other hypothesized functions (such as aiding in recognizing emotions: Poggi et al., 2009; or aiding in assessing the health state of potential sexual partners; Tomasello et al., 2007) the morphological properties of a de-pigmented sclera also afford using the eye-ball as a reliable spatial pointer. In co-occurrence with vocalizations, gaze allows interactants to efficiently coordinate their attention towards the relevant entities in cooperative interactions in peripersonal space. We propose that this referential function of eye-gaze, when orchestrated with vocalizations, could have been essential for our ancestors to coordinate their activity in complex tasks that either required the use of hands to manipulate objects (e.g. tool-making or cooking), or in which acoustic cues would be disfavored (e.g. hunting). Under these circumstances, gaze-cues become the only reliable visual pointer. We speculate that a general orientation towards these kinds of cooperative activities might have motivated a selective pressure for the particular characteristics of the ocular morphology of modern hominins, not dissimilar from those of other extant great apes, like Western lowland gorillas (Mayhew, 2013), or Sumatran orangutans (Perea García, 2016). This evolutionary trend in developing a conspicuous eye morphology is most undoubtedly most visible in modern humans, which led previous research to conclude that it was uniquely human (Kobayashi and Koshima, 1997, 2001).

1.1. Language as a multimodal activity

The ability to communicate linguistically is indisputably a uniquely human trait. Even though Western scholarship has entertained itself on the topic of language for centuries, it has been concerned mostly with its written form (Linell, 1982). This has led to a conception of language as an object-like *system*, rather than a social coordination process or *activity* (Love, 2004). This in spite of the fact that face-to-face conversation remains the most pervading communicative practice among humans, and critically constitute our entry into language in processes of language acquisition (Cowley, 2007; Levinson, 1983; Peräkylä, 2008; Tomasello, 2000). This bias in the Western language sciences has led to verbal signals being studied in isolation, neglecting their typical co-occurrence with other kinds of signals in the context of conversation. However, more recent developments, such as Conversation Analysis (Goodwin et al., 1987) do conceive language as only one among many communicative resources in the multimodal activity that constitutes conversation. Researchers in these and related fields thus argue in favor of considering language as essentially emerging from the interaction of different communicative semiotic resources, granting vocalizations a referential function initially on the basis of their co-occurrence with visual cues such as for instance pointing and gazing, both on phylogenetic (Altenmüller et al., 2013; Leavens et al., 2010; Levinson and Holler, 2014; Liebal et al., 2013; Partan and Marler, 1999; Taglialatela et al., 2011) and ontogenetic (see Emery, 2000 for a review) time scales.

On these views, understanding the ways in which different communicative resources interact, affecting each other's dynamics, becomes as important as understanding each resource in isolation. With their integrated message model (IMM), Bavelas and Chovil (2000) provide a useful theoretical framework to understand the interaction of cues from different modalities. Their model distinguishes between two main communicative resources in face-to-face dialogue, depending on the modality in which they are expressed: 1) Auditory Acts of Meaning, and 2) Visual Acts of Meaning (AAM, and VAM, respectively). The interaction between the basic meaning of each of these "acts" will contribute to constraining inferential processes towards a specific interpretation. A common way in which acts of meaning conveyed through different modalities interact is by pure "redundancy" – that is, the basic meaning of each act contributes to constraining inferences towards the same interpretation. For instance, a child could be smiling while saying "I'm so happy!" where both modalities indicate that he or she is happy. This contrasts with how an older individual might exclaim "I'm so happy!" with an accompanying neutral face, marking the interpretation of the IMM as not being straightforward (e.g., irony). Redundancy, however, does not mean "repetition" – rather, the authors propose that this functional overlap is only partial, and contributes to nuances in the perceived communicative intention behind the IMM.

In this study, we explore communicative strategies that exploit the functional overlap of AAM (spoken verbal utterances) and VAM (eye-gaze) in conveying referential information about entities that are immediately co-present to both interactants. This enables the establishment of "common ground" (Rączaszek-Leonardi et al., 2014) between the interactants, deemed essential for communication (see Tomasello, 2010). Note that this functional overlap is what Bavelas and Chovil (2000) described as "redundancy", since the function of both vocalizations and eye-gaze is to refer to immediately co-present entities. We compare this communicative strategy – largely overlapping in function with the linguistic category of *deictics* – and protodeclaratives as observed in human infants and extant great apes (Gómez, 1996; 2005; 2007), hypothesizing that it could a have been key for the development of critical cultural innovations and abilities (such as toolmaking, hunting, cooking) in the history of our ancestors.

1.2. The role of bodily constraints in conveying referentiality

Understanding the relationship between bodily constraints and linguistic communication (Tylén et al., 2013), as well as the development of the morphological features themselves in the history of our genus, can help us reconstruct the communicative practices of our ancestors. This can be done by inferring the effectiveness of specific communicative resources afforded by the morphological constraints of our body. For example, we assume that our hominin ancestors could point insofar as their arms were freed from the task of locomotion either by brachiating, like orangutans, or by quadrupedal locomotion, like gorillas.