The temporal dynamics of alignment in multimodal interaction
Bert Oben, Geert Brône, and Kurt Feyaerts
bert.oben@arts.kuleuven.be

The tendency towards convergence between speakers in interactional discourse is a widely studied phenomenon in different disciplines and for different semiotic modalities, including speech, gesture and posture. Recent work in (psycho)linguistics and cognitive psychology has focused on the role of imitative behaviour - alternatively referred to as alignment (Pickering & Garrod, 2004), resonance (Du Bois, 2011) or conceptual pacts (Brennan & Clark, 1996) - in establishing successful communication. Interactional discourse requires speakers and their utterances to be geared to one another in multiple ways so as to facilitate meaning negotiation, and this process requires alignment at different levels of (linguistic) representation. In the majority of cognitive studies dealing with imitative behaviour, the perspective has been largely monodimensional and restricted to minimal contexts. The focus is generally on one semiotic channel or on one linguistic level rather than on alignment as a clustered phenomenon with features co-occurring simultaneously on different levels (e.g. Branigan et al., 2007, on syntactic alignment, Kimbara, 2006, on gestural mimicry, De Fornel, 1992, on postural echoing, and many others). The restriction to minimal contexts, such as pairs of utterances rather than longer sequences of discourse, has led to a relative disregard for the discursive emergence and persistence of convergence across speakers in interaction. In this paper, we zoom in on the discursive development of interactive synchronisation as an online and gradual process with multimodal alignment sequences that emerge, persist and die out in the interaction. In order to arrive at such a fine-grained multimodal picture, we conducted a corpus study using the InSight Interaction Corpus (Brône & Oben, 2012). This corpus consists of video recordings of both targeted and free-range dyadic interactions, with multiple camera perspectives providing a full view of the dialogue partners’ nonverbal behaviour, including hand gestures, facial expressions and body posture (see Figure 1 for a screenshot of the recording set-up). The use of head-mounted scene cameras and eye-trackers provides a unique “speaker-internal” perspective on the conversation, with detailed production information (scene camera and sound) and indices of cognitive processing (eye movements for gaze analysis) for both participants. In order to arrive at a full-fledged account of alignment sequences in ongoing interaction, we first looked at distributional patterns within each of the semiotic channels, and the role of different levels of representation. We address questions such as “does inter-speaker alignment on one level increase in the course of an ongoing interaction?” and “does alignment persist across the boundaries of individual sequences (e.g. specific thematic units) as part of a larger interaction?”. For the first question, we singled out lexical and gestural means of object representation in a set of collaborative tasks in the corpus (where subjects were asked to describe spatial scenes projected on a screen). The data reveal a steady increase in cross-speaker convergence (or a decrease of variation) as interactions unfold, both in lexical choice and gestural means of representation. This effect persists across the boundaries of individual tasks in the interactions of the corpus. A second step in the analysis deals with potential differences in the distribution and temporal build-up between different semiotic channels. To what extent does the emergence of strong cross-speaker alignment follow a similar or different temporal path in different modalities? In order to answer this question on the temporal dynamics of multimodal alignment, we compared the time course of establishing the lexical and gestural routines described in the first step above, and plotted their co-occurrence in time. More specifically, by time-aligning the instances of lexical and gestural cross-speaker convergence, we obtained a fine-grained picture of their multimodal discursive development. The results of this analysis reveal both high degrees of
interaction/overlap between channels (what we label “clustered alignment”) and some significant differences in durability or persistence.

References