ORALITÉ ET GESTUALITÉ

COMMUNICATION MULTIMODALE, INTERACTION

Actes du colloque ORAGE'98 édités par
Serge Santi,
Isabelle Guaïtella,
Christian Cavé
et Gabrielle Konopczynski

The Synchronization of Gesture and Speech in Dutch and Arrernte (an Australian Aboriginal language) : A Cross-cultural Comparison *

Jan-Peter de Ruiter and David P. Wilkins

Max Planck Institute for Psycholinguistics

Introduction

The synchronization between spontaneous iconic gestures and their accompanying speech is a phenomenon studied extensively by McNeill (1992), but also by Morrel-Samuels & Krauss (1992) and De Ruiter (1998). The implicit assumption made by these authors is that the synchronization of gesture and speech are the same for any speaking human being. That is to say, potential cross-cultural differences are presumed not to play a role in the issue of gesture / speech synchronization.

This study is aimed at testing whether this assumption is indeed correct. This is done by comparing synchronization in both Arrernte and Dutch, two languages and cultures that differ in some very important respects. First, Arrernte is a case-marking language with free phrase order, while in Dutch word order signals grammatical function. Second, Arrernte is a culture in which gesture can be assumed to play a more prominent role than in Dutch. As in other Australian communities (Haviland, 1993), gestures are oriented correctly in an absolute frame of reference and are attended to directly as part of the « utterance ». Moreover, Arrernte also has an auxiliary sign language, and the gesture space appears to be significantly larger than in Dutch. These differences might lead one to suspect that there would be differences between the groups in the synchronization of gesture and speech.

---

* Note des éditeurs : Suite à un problème de transmission, les tableaux et figures de cette contribution n'ont pu être reproduits.

The study

De Ruiter and Wilkins (in preparation) used videotaped narrations to investigate the synchronization of gesture and the affiliated speech. They created a database with timing information and detailed transcripts of each gesture. The gesture database has the following features :

4 speakers from each community — all considered « elder » members by their respective communities.

All speakers were talking about places, locations and events in their familiar, community surroundings — i.e. imparting information based on their «lived» experience of the local geographic and social space which they have inhabited for the better part of their lives.

The speakers were video-taped in a free-interview situation. A single camera was used, facing the subject from the front.

All speakers were established as dominant right-handers.

The research focuses on co-speech gestures which are spatial in nature — that is, gestures, which function to encode real orientations, locations and movements in the local geographic and social space being talked about.

We examined a minimum of 15 manual excursions from each subject.

A manual excursion is a movement from rest out to gesture and then back to rest. It may be simple or complex — that is, while the hand is out, it might make only one gesture and return, or it might make a series of discreet gestures before returning.

We treated each component gesture of a complex manual excursion as a single gesture (for purposes of quantification and comparison). So, although only examining a minimum of 15 manual excursions per person, this may equate to substantially more gestures if most of the 15 are complex.

Videotaped segments were digitized and we used Media-Tagger, a tool for transcription and analysis of digitized video fragments on Apple Macintosh, to code the gestures and to create the data-base.

For every gesture, the onset of the (preparatory) movement, the onset of the meaningful phase (roughly equivalent to the notion of stroke [McNeill, 1992]), and the offset of the meaningful phase were located in time. The retraction was not coded.

For each gesture, we isolated the actual speech affiliate of the gesture — i.e., the word or phrase with which the gesture is semantically and pragmatically linked. The precise temporal boundaries of the actual speech affiliate were determined by using the Macintosh waveform display function.

Each gesture was coded as either obligatory or non-obligatory (supportive) obligatory The utterance (contextually) requires a gesture to go along with it for it to be decodable and semantically complete. Supportive : While the utterance could stand on its own in context (with the primary proposition intact) a gesture accompanies it any way to expand on or emphasis information that is in the utterance.

Further coding of the gestures included :

– Handedness of the gesture.

– Whether a place name was used in the affiliated speech.

– Whether the affiliate contained spatial deixis (e.g. dit (this) or daar (there) in Dutch)

– Whether the affiliate contained a motion verb.

– Whether the affiliate had a closed class grammatical motion marker

For purposes of qualitative analysis, transcripts were generated from the database, having the following features : (a) the speech line gives all the speech which occurred during the gesture ; (b) curly brackets (« {xx} ») enclose the portion of speech which timed with the meaningful part of the gesture ; (c) timing of the meaningful phase is shown by timing codes added to each bracket (in multiples of 40 milliseconds) — note that the first time code shows the time in milliseconds from the initiation of the gesture ; (d) the duration of the meaningful phase is given under the dotted line which stretches between begin-time and end-time ; (e) the actual affiliate text is given underneath as is the affiliate timing ; and (e) a line over the transcribed text shows where the actual affiliate is.

For Arrernte we have a total of 100 gestures. Of these 44 are obligatory and 56 are supportive. For Dutch we have total of 80 gestures. Of these 22 are obligatory and 58 are supportive.

Some preliminary analyses

There is a significant body of research, primarily on English, which shows that gesture initiation precedes the onset of the associated affiliate. For example, for all 60 gestures considered by Morrel-Samuels and Krauss (1992), gesture onset preceded the onset of the affiliate. So we can ask, do gestures anticipate speech in our Arrernte and Dutch data ? Or to put it another way, how often does it happen that the onset of the affiliate precedes the onset of the gesture ?

Dutch has only five example where speech onset precedes gesture onset, which is approximately 6 %. Arrernte, however, has 12 gestures in which speech precedes gesture, which amounts to 12 %. Thus, 94 % of the time

Dutch speakers' gestures anticipated affiliated speech, while 88 % of the time Arrernte speakers gestures anticipated affiliated speech.

So basically, the answer seems to be yes — gesture onset precedes the affiliate. According to McNeill, gesture initiation indeed precedes the affiliate, but the meaningful phase (or stroke, as he calls it) overlaps in time with the affiliate.

To investigate this and related issues, we have adopted the six categories of timing relation from de Ruiter (1998).

Each gesture can be categorized as falling into one of these temporal relations. Similarly, each meaningful phase of the gesture can be categorized as falling into one of these six relations. Now if both Morrel-Samuels & Krauss and McNeill are correct, gestures should be of type 3 or 5. In our Dutch data 84 % of the gestures were type 3 or 5, while in the Arrernte data 80 % of the gestures were of type 3 or 5. This analysis indicates that McNeill is, statistically speaking, correct.

Of course this does not distinguish Obligatory versus Supportive Gestures — We might assume — following Levelt, Richardson and La Heij (1985), and also De Ruiter (1998) — that there will be a tighter timing between obligatory gestures and their affiliates than supportive gestures and their affiliates. We therefore computed the correlation between the temporal midpoint of the gesture (meaningful phase) and the temporal midpoint of the affiliated speech.

Thus, the data from both languages goes against the prediction — supportive gestures are more tightly timed with their affiliates than obligatory gestures. This effect is most marked for the Arrernte data.

So far, Arrernte and Dutch synchronization appear to be remarkably similar. However, there are also interesting differences between the two languages. Arrernte speakers use a much larger gesture space. This means that, on average, Arrernte gestures should have longer preparatory motions (the part of the gesture before the meaningful phase). This was indeed the case. The average duration of the preparation movement was 803 ms. for Arrernte and 559 ms. for Dutch (p < .001).

If the preparatory phase is longer for Arrernte, how will this affect the synchronization of the meaningful phase with the affiliated speech ? There are two main theoretical proposals for the temporal synchronization of gesture and speech. One is the interactive theory by McNeill (1992, in press). According to the interactive theory, gesture and speech develop together as one, implying that they share state information in order to continually align in time. According to this theory, the synchronization of Arrernte meaningful phases should be unaffected by the fact that Arrernte gestures have longer preparatory movements. The other theory by Levelt et al. (1985), later embedded in a model by De Ruiter (in press) assumes that gesture and speech are initiated in synchrony, but after initiation both gesture and speech will proceed without cross-talk between them. From this theory, the prediction would be that Arrernte gestures will be shifted « to the right » in time, relative to the affiliated speech. We can investigate this by looking at the difference between the onset of the meaningful phase and the onset of the affiliate for both languages. For Arrernte, this difference is on average −194 ms (a negative number means that speech preceded gesture) while for Dutch this is 53 ms.

Note that the amount of temporal shift between Dutch and Arrernte (53 + 194) is 247 ms, while the extra preparatory time for Arrernte gestures (803 - 559) is 244 ms. This is strong support for Levelt et al. (1985)'s claim, and for the gesture / speech model by De Ruiter (in press), for it reveals that the larger gesture space does lead to another « time alignment » of gesture and speech.

Conclusions

McNeill's claim that the initiation of gestures generally precede their affiliated speech, while the meaningful phase overlaps in time with the affiliate is clearly supported by these cross-cultural data.

The claim by Levelt et al. (1985) and De Ruiter (1998) that obligatory gestures are more tightly synchronized than supportive gestures was shown to be wrong for both Dutch and Arrernte : it is actually the other way around. Supportive gestures are more synchronized than obligatory ones. This is perhaps the consequence of the fact that for obligatory gestures it is more clear what the relationship between the gesture and the speech is, while for supportive gestures this is not always as clear. If that is the case, it is more important for supportive gestures to be aligned in time with their affiliates.

While the comparison between Arrernte and Dutch gesture / speech synchronization reveals a surprising level of similarity, there also is an interesting difference between the two languages : Arrernte gestures have longer preparatory movements. This is likely to be a consequence of the larger gesture space used in Arrernte. However, this leads to another time alignment of speech and gesture (meaningful phase) onset, which is compatible with the « ballistic » theory of gesture and speech synchronization by Levelt et al. (1985) and De Ruiter (in press).