

GESPIN 2019
11 - 13 September



This paper was presented at the 6th Gesture and Speech in Interaction Conference that was held at Paderborn University, Germany from September 11-13, 2019.

To cite this paper:

Ferré, F. (2019). Gesture / speech alignment in weather reports. In: Grimmering, A. (Ed.): *Proceedings of the 6th Gesture and Speech in Interaction – GESPIN 6* (pp. 34-38). Paderborn: Universitätsbibliothek Paderborn. doi: 10.17619/UNIPB/1-805

Gesture / speech alignment in weather reports

Gaëlle Ferré

Nantes Université & CNRS UMR 6310 – LLING, Nantes, France

Gaëlle.Ferre@univ-nantes.fr

Abstract

This paper presents a follow-up study of previous work conducted on pointing gestures and their alignment with speech in weather reports (Ferré & Brisson, 2015, Ferré, 2019). Yet, whereas the previous studies concentrated on the expression of viewpoint and how gestures function in association with other semiotic resources, the present study focuses in more detail on the timing relationships between the different modes in speech and the apparent absence of synchronicity in some gesture/speech constructions in French weather reports. What is proposed here is a theoretical analysis rather than a quantitative one in which it will be shown that in order to account for this apparent misalignment of modalities (a) the inclusion of other semiotic modes in the annotation scheme may be useful for the description of specific corpora like weather reports, and (b) temporal graphs that include gesture targets can offer a good representation of the temporal relationships between gesture and other domains involved in communication acts.

1. The challenge of multimodality

Multimodality implies that the meaning making process constantly involves several semiotic resources (Adami, 2017). Oral communication in face-to-face interactions, as an instance of meaning making, involves not only language, but also gesture, posture, facial expression and other bodily behaviors such as proxemics and attitudes. Spoken interactions also typically occur within a physical environment of which certain elements can be integrated in communication acts, as shown by Goodwin (1994, 2007) and Streeck (1996), and form their own semiotic system. Each semiotic system involved in communication acts has its own systemic affordances and material constraints so that what is communicable in speech may not be so easy to communicate in a visual mode (like gesture or graphic representation) and vice versa. This is the reason why Discourse Analysis should not focus on only one modality even if some modalities can be predominant in certain social practices, as in the type of media that is the object of study in the present paper.

In some ideal world, any speech act would contain at least one syntactically complete and grammatically correct clause made of words themselves formed with distinct morphemes. The clause would be bounded by clearly identifiable prosodic boundaries and would be uttered with an intonation contour that would be congruent with the speech act accomplished verbally (whether it be a statement, a question, or any other type). The syntactic clause could also be accompanied by a gesture whose onset and offset would precisely match the syntactic and prosodic boundaries. This gesture would in turn be composed of different phases that would also match the lexical or morphemic boundaries in speech. Lastly, the information conveyed by gesture and prosody would be congruent with the semantics of the clause and its constituents.

This ideal communication act is indeed found in spontaneous interactions, but as anyone who has worked before on naturally occurring interactions knows, misalignments also occur and this therefore makes the issue of the temporal alignment of information units in the different modalities and their conjoint analysis a central one in any multimodal study of video corpora. Considering this issue, the challenge of multimodal discourse analysis, i.e. the study of relationships between different modalities in discourse and the way each modality participates in meaning-making processes, consists in annotating data in linked but nevertheless different semiotic modes that do not always share the same temporal structure and in revealing the interactions between them in as systematic a treatment as possible.

2. Temporal alignment of modalities

Generally speaking, the vast majority of Intonation Phrases temporally coincide with syntactic clauses in speech (Barth-Weingarten, 2016), although this depends a lot on the degree of improvisation and informality of interactions. In this respect, weather reports are well rehearsed types of media, based on scripted material, which means that speech delivery is very fluent. This type of media also takes the form of monologues and prosody is therefore not used as a turn-management device as can be the case in dialogues, in which speakers sometimes purposely avoid to make syntactic and prosodic boundaries coincide not to lose their speech turn.

As far as gestures are concerned, it has been observed that gesture production is linked to the syntactic structure of the speech it accompanies depending on the language of the speaker: Kita and Ozyürek (2003) noted that if some information is typically given in the form of two syntactic clauses in a language, speakers tend to express this same information with two different gestures, whereas when the language enables speakers to express the information in a single syntactic clause, then speakers tend to produce a single gesture to accompany their verbal expression. In terms of discourse structure, McNeill (1992) also observed in a narrative task that speakers tend to produce one gesture per narrative clause which means that a gesture in this case participates in the expression of one idea unit.

Yet, there are differences between gestures and prosody. The major difference between the visual and vocal semiotic modes lies in the fact that whereas one cannot speak without any prosody at all, hand gestures are not required to accompany every piece of verbal information, which means that manual gestures are perhaps a bit more independent from speech than prosody. It also means that not every syntactic clause is accompanied by a gesture.

Speech and gesture may also differ in respect to their temporal structure and this has an impact on their alignment with each other. Shattuck-Hufnagel and Ren (2018) signal that studies concerned with gesture/speech synchronization present contradictory results: while some scholars found a (fairly) good alignment between gesture and speech (Loehr, 2004; Chui, 2005, for instance), others found that some gesture types tend to be produced in anticipation of speech (Schegloff, 1984; Leonard and Cummins, 2009; Ferré, 2010). Shattuck-Hufnagel and Ren (ibid.) however note that the different observations made in this respect may be explained by the fact that scholars were working on different languages and considering different gesture types or even base their observations on different gestural landmarks (gesture apex, whole stroke or even whole gesture phrase) and with different time windows. For McNeill – although the author doesn't specify how precisely he measured this figure – (2005, p. 32) “the stroke is synchronous with co-expressive speech about 90 percent of the time (...). When strokes are asynchronous, they precede the speech to which they link semantically”, i.e. their *lexical affiliate* (Kipp et al., 2007).

3. Semantic gesture/speech mismatches in weather reports

While working on pointing gestures in weather reports both in English and French, we observed a difference between the two languages in terms of gesture/speech alignment. In French, mismatches were found slightly more often than in English between some pointing gestures and the locations pointed at on the map shown in a background screen. Whereas in French weather reports, 9 % of the pointing gestures towards a location on the screen showed clear misalignment with the location mentioned in speech – and therefore fit well with the description provided by McNeill (ibid.) quoted in the previous section – the English corpus showed a lower misalignment rate of 4 %. Although the corpus is extremely limited in size and the difference may not be significant, we may still wonder if we can really talk of gesture/speech mismatches in these cases in French and why the two languages tend to function differently in this respect.

Figure 1 below presents a sequence in a French weather report, in which some gestures do not align with what is referred to in speech. As he begins a new description in (a), the forecaster mentions the ‘Val de Garonne’ and points to this particular location on the map of France that is shown in the background screen. He then initiates a second move but the Intonation Phrase shown in (b) does not contain any spatial reference. Yet, the forecaster anticipates on the next Intonation Phrase and already points to the Pyrenees. In (c) where the Pyrenees are mentioned in speech, he anticipates again in gesture on the next Intonation Phrase and moves his hand directly to the Alps so that as he finishes the word ‘Pyrénées’, his hand is now fully pointing at the Alps on the map.

He catches up in the last Intonation Phrase shown in (d) and points to the Mediterranean as he utters ‘the Alps and the Mediterranean’ packaged in a single Intonation Phrase. The misalignment between speech and gesture is so large that although the apex of each gesture is aligned with the stressed syllable of each Intonation Phrase and the gestures could then be considered as respecting the gesture-speech alignment rules observed by Loehr (2004), there is a mismatch of more than 200 ms in semantic content between what is referred to in speech and what is pointed at in the background screen for two gestures in the sequence.

In sum, the example illustrated in Figure 1 shows that whereas the first and last gestures in the sequence align their apex with the right locations in speech, the second gesture does not align with any spatial location in speech and the third one points at the Alps on the map when the Pyrenees are mentioned in speech. The last gesture produced by the forecaster aligns with one of two locations mentioned in speech. It starts as the Alps are being mentioned but its apex coincides rightly with the mention of the Mediterranean. The gestures in (b) and (c) are then clearly misaligned with their lexical affiliates and the gesture/speech constructions in these two cases seem rather ill-formed, unless one considers that the different elements or *objects* forming the construction can be analyzed in terms of their respective properties and the relationships they entertain with each other on different planes of discourse (Blache, 2004). These relationships between objects in utterances can be represented in the form of temporal graphs (Bird and Liberman, 1999).

	/de la grisaille brumeuse de nouveau/ <i>Grey mist again</i>	
(a)	/dans le Val de Garonne/ <i>in the Val de Garonne</i>	
(b)		/mais beaucoup de soleil/ <i>but very sunny</i>
(c)	/en allant vers les Pyrénées/ <i>towards the Pyrenees</i>	
(d)		/les Alpes et la Méditerranée/ <i>the Alps and the Mediterranean</i>

Figure 1. Gesture / speech temporal (mis)alignment in French (Prévisions Météo-France, 17 Nov. 2015).

4. Temporal graphs for gesture/speech constructions in weather reports

Bird and Liberman (ibid.) consider that any linguistic domain (prosody, gesture, discourse, syntax, phonology...) comprises a number of *objects* organized in a linear way on a temporal axis, so that a multimodal corpus is composed of different objects with an onset and offset time that can be represented by nodes on a timeline. A weather forecast, as said before, is a type of media based on three major semiotic resources: speech, gesture and a background screen. The aim of pointing gestures in this communication type is to establish a link between the background screen and speech content and to open up focus spaces on that screen for the audience to concentrate on (Grosz and Sidner, 1986). The example presented in the previous section can be represented as in Figure 2. (a) shows a multimodal construction in which an Intonation Phrase made of a single syntactic Prepositional Phrase includes a lexical reference to a spatial location. The phrase is accompanied

by a pointing gesture towards a congruent location on the map shown in the background screen. In (b) the syntactic phrase uttered in an Intonation Phrase is also accompanied by a point towards a location on the map, but the gesture-map construction links to the location expressed verbally in (c). The gesture that accompanies the Intonation Phrase here in turn matches a location expressed verbally in (d). This last Intonation Phrase groups two syntactic Noun Phrases that refer to two different spatial locations, but the gesture produced during the utterance of this phrase targets the last spatial location mentioned in speech.

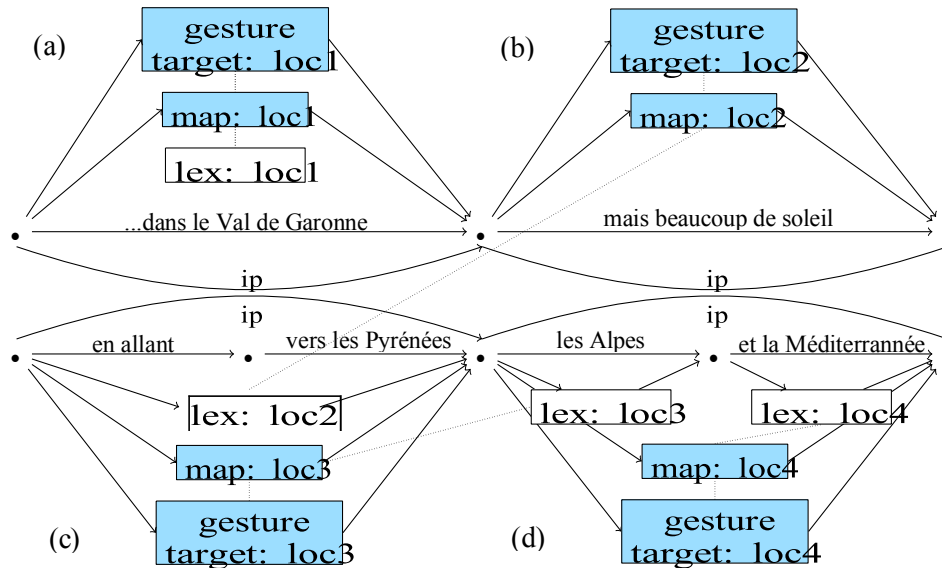


Figure 2. Graph showing dependency relations between syntax on the timeline, prosody, gesture and a visual map (ip = Intonation Phrase, lex = lexical information, loc = location).

5. Conclusion

As was shown in this paper, multimodal constructions may well be composed of objects temporally synchronized with each other as in Figure 2 (a), where the syntactic, semantic, prosodic and gestural domains are all congruent with one another and are besides perfectly coupled with the communication environment (a location on a map, for instance, in the case of weather reports). They may however also be partly synchronous with one another as in Figure 2 (b), (c) and (d): whereas (b) comprises a single syntactic phrase uttered in an Intonation Phrase, (c) and (d) both comprise two syntactic phrases packaged in single Intonation Phrases. Besides, if the gestures produced in these three constructions are nicely aligned with Intonation Phrases, their targets in (b) and (c) are not synchronized with the corresponding spatial locations in speech. This means that multimodal constructions are not always based on the semantics of speech, but rather on the way the information is packaged into prosodic units.

Lastly, although the corpus on which this theoretical paper is based is quite limited in size thus precluding any generalization, it appeared that pointing gestures were more frequently misaligned with referential spatial locations – as they tended to anticipate the lexical reference more often – in French weather reports than in English ones. This might be due to the different information structure of the two languages: whereas spoken English is very similar to written English considering word order, there is a large difference between written and spoken French in terms of information structure, with a tendency to place focused elements at the beginning of a sentence in spoken French. The semantically misaligned gestures in weather reports, that open up focus spaces on a map, may be considered to be following the information structure of oral French (which could be the reason why they tend to align with intonation rather than syntactic phrases), whereas the verbal information, based on scripted material, rather follows the information of written French which could explain the fact that pointing gestures in weather reports anticipate more frequently on speech in this language.

References

- Adami, E. (2017). Multimodality. In García, O., Flores, N., and Spotti, M., editors, *The Oxford Handbook of Language and Society*, pages 451–472. Oxford University Press, Oxford.
- Barth-Weingarten, D. (2016). *Intonation Units Revisited. Cesuras in Talk-in-Interaction*. Amsterdam, Philadelphia: John Benjamins.
- Bird, S. and Liberman, M. (1999). A Formal Framework for Linguistic Annotation. In *Technical Report MS-CIS-99-01*, pages 1–48, University of Pennsylvania.
- Blache, P. (2004). Property Grammars: A Fully Constraint-Based Theory. In Christiansen, H., Rossen Skadhauge, P., and Villadsen, J., editors, *Constraint Solving and Language Processing*, pages 1–16. Springer, Berlin.
- Chui, K. (2005). Temporal Patterning of Speech and Iconic Gestures in Conversational Discourse. *Journal of Pragmatics*, 37:871–887.
- Ferré, G. (2010). Timing Relationships between Speech and Co-Verbal Gestures in Spontaneous French. In Kipp, M., Martin, J.-C., Paggio, P., and Heylen, D., editors, *LREC: Workshop on Multimodal Corpora*, pages 86–91, Valetta, Malta. ELRA.
- Ferré, G. (2019). Time Reference in Weather Reports. The Contribution of Gesture in French and English. In Galhano, I., Galvão, E., and Cruz dos Santos, A., editors, *Recent Perspectives on Gesture and Multimodality*, pages 31–40. Cambridge Scholars Publishing Ltd, Cambridge.
- Ferré, G. and Brisson, Q. (2015). “This Area of Rain will Stick South in the Far North”. Pointing and Deixis in Weather Reports. In *Proceedings of GESPIN 4*, pages 101–106, Nantes, France.
- Goodwin, C. (1994). Professional Vision. *American Anthropologist*, 96(3):606–633.
- Goodwin, C. (2007). Environmentally Coupled Gestures. In Duncan, S., Cassell, J., and Levy, E., editors, *Gesture and the Dynamic Dimensions of Language*, pages 195–212. Amsterdam, Philadelphia: John Benjamins.
- Grosz, B. J. and Sidner, C. L. (1986). Attention, Intention, and the Structure of Discourse. *Computational Linguistics*, 12(3):175–204.
- Kipp, M., Neff, M., and Albrecht, I. (2007). An Annotation Scheme for Conversational Gestures: How to Economically Capture Timing and Form. *Language Resources and Evaluation*, 41:325–339.
- Kita, S. and Ozyürek, A. (2003). What does Cross-Linguistic Variation in Semantic Coordination of Speech and Gesture Reveal?: Evidence for an Interface Representation of Spatial Thinking and Speaking. *Journal of Memory and Language*, 48:16–32.
- Leonard, T. and Cummins, F. (2009). Temporal Alignment of Gesture and Speech. In *Proceedings of Gespin*, pages 1–6, Poznan, Poland.
- Loehr, D. P. (2004). *Gesture and Intonation*. PhD thesis, Georgetown University, Georgetown.
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. The University of Chicago Press, Chicago, London.
- McNeill, D. (2005). *Gesture and Thought*. University of Chicago Press, Chicago, London.
- Schegloff, E. A. (1984). On some Gestures’ Relation to Talk. In Maxwell Atkinson, J. and Heritage, J., editors, *Structures of Social Action. Studies in Conversation Analysis*, pages 266–296. Cambridge University Press, New York.
- Shattuck-Hufnagel, S. and Ren, A. (2018). The Prosodic Characteristics of Non-referential Co-speech Gestures in a Sample of Academic-Lecture-Style Speech. *Frontiers in Psychology*, 9:1–13.
- Streeck, J. (1996). How to Do Things with Things: Objets Trouvés and Symbolization. *Human Studies*, 19:365–384.