Annotation of space and manner/path configuration in bilinguals’ speech and manual gestures

Federica Cavicchio, Amanda Brown, Reyhan Furman, Shanley Allen, Asli Özyürek, Tomoko Ishizuka and Sotaro Kita

CIMeC Unitn, Syracuse University, University of Alberta, University of Kaiserslauten, Ratboud University of Nijmegen, Tama University, University of Warwick

E-mail: author1@xxx.yy, author2@zzz.uuu.edu, author3@hhh.com

Abstract (10-point Times New Roman bold, centred)

Different languages and cultures use gestures differently. The goal of this paper is describing the coding scheme used to annotate a corpus of English/Italian bilinguals and English and Italian monolinguals describing a set of stimuli designed to elicit the description of manner and path events and the corresponding gestures ("the Tomato Man stimuli"). The first question we investigated was the relationship between clause structure for motion event expressions and gestural representation of the same event. From the seminal work of Kita and Özyürek (2003), many studies have investigated manner and path in the verbalization of motion events and the co-produced manual gestures in different languages. Following Talmy's (1985) typology, English allows verbal constructions to conflate complex meaning within a single clause as path can be expressed as a "satellite" to the verb. That is, manner and path can be expressed in within a single clause (e.g. roll down). On the other hand, Italian is more restricted in situations in which manner of motion verbs can occur with path phrases. That is, manner and path are often realized by two verbs (scende rotolando, i.e. it goes down rolling). We describe the annotation scheme we used to code speech and gesture in English/Italian bilinguals and monolinguals. Three annotators codified 403 tokens corresponding to the gesture stroke phase. We coded gestural and verbal expressions of manner and path and gesture space.

1. Introduction

Different languages and cultures use gestures differently. The goal of this paper is describing the coding scheme used to annotate a corpus of English/Italian bilinguals and English and Italian monolinguals describing a set of stimuli designed to elicit the description of manner and path events and the corresponding gestures ("the Tomato Man stimuli", Özyürek et al., 2001). Researchers have been interested in whether bilinguals transfer gestures from one language to another, that is, if some linguistic or spatial aspects of gestures that are linked to a certain language are transferred to the other language while speaking (see Nicoladis, 2007 for a review). Gesture transfer (or lack thereof) will give an insight on how gesture and language are linked in production. This project investigated whether gestural expressions of manner and path, gesture frequency and gesture space transfer from the first language to the second language. Thanks to this coding scheme, we addressed two research questions on the matter of bilingualism. The first question investigated the relationship between clause structure for motion event expressions and gestural representation of the same event. From the seminal work of Kita and Özyürek (2003) many studies have investigated manner and path in the verbalization of motion events and the co-produced manual gestures in different languages. Following Talmy’s typology (1985), English allows verbal constructions to conflate complex meaning within a single clause as path can be expressed as a "satellite" to the verb. That is, manner and path can be expressed within a single clause (e.g. roll down). On the other hand, Italian is more restricted in situations in which manner of motion verbs can occur with path phrases. Manner and path are often realized by two verbs (scende rotolando, i.e. it goes down rolling). Nevertheless, single-clause constructions such as "rotola giù/su" (rolls down/up) can be used by Italian native speakers. Following Kita and Özyürek (2003), it is expected that, regardless of the language, single-clause verbal constructions will be accompanied by conflate gestures, combining the information about manner and path in one movement, whereas two-clause verbal constructions will be accompanied by 2 gestures, one expressing manner and one expressing path. The second question we investigated was how the four groups of speakers differ in gesture frequency and gesture space. For example, Italian is reported as a high gesture frequency language (Barzini, 1964, Kendon, 1992, 1995), as opposed to (British) English, described as a low gesture frequency language (Graham and Argyle, 1975). Another gesture parameter that varies across cultures is gesture size: bigger in Mediterranean cultures than in northern European cultures. Since the seminal study of Efron (1941/1972) comparing Jews and Italian immigrants’ gestures, we know that in different cultures gestures differ in how they are performed in the space. In particular, Efron observed that Italian immigrants’ gestures were spatially expansive, moving the entire arm from the shoulder joint, and tended to occupy the lateral (transversal) plane. More recently, Müller (1998) compared the gesture space of native Spanish and German speakers involved in a naturalistic conversation task with a language matching confederate. She found that Spanish speakers produced more gestures in the space above their shoulder than German speakers. Gesture size is an interesting variable to consider for gesture transfer in bilinguals. In the following, we describe the annotation scheme we used to code speech and gesture in English/Italian bilinguals and monolinguals. Three annotators codified 403 tokens corresponding to the gesture stroke phase. We coded gestural and verbal expressions of manner and path and gesture space.

2. Corpus description

Data were collected from two monolingual control groups so that we can properly address the questions whether
bilinguals’ gestures are different from monolinguals’ gestures and/or whether parameters of gesture production transfer from a language to another. The two monolingual control groups of English and Italian speakers were matched with the bilinguals for gender, age and education background. We focused on highly proficient Italian/English early bilinguals (i.e., they learned both languages before age 6) who had a very similar fluency in both languages. Bilinguals and monolinguals described the exact same stimuli in each language to a confederate language matching speaker. Stimuli consisted of 10 single-scene cartoons depicting actions performed by red tomato and green triangle. Participants were required to describe each cartoon as accurately as possible to a language matched monolingual speaker. 20 English native speakers, 20 Italian native speakers and 20 English/Italian bilinguals were recorded while describing to a matching language listener the ten Tomato man cartoons. Bilinguals described the stimuli twice, once in English and once in Italian, to two different native speakers. Monolinguals described the stimuli twice in their native language to two different native speakers.

4.1 Transcription
A native speaker of Italian and two native speakers of English transcribed the descriptions. Disfluencies, repetitions and laughter were transcribed with special fonts. The transcriptions were checked for accuracy by a fluent speaker of Italian and English. All the transcriptions were done in Elan 4.3.3 to ensure a correct time alignment with coverbal gestures. In this study we focused on the stroke phase of each gesture performed by the speakers, as defined by Kita, van Gijn, & van der Hulst (1998). Gesture strokes were transcribed and aligned with speech.

4.2 Coding scheme
The coding scheme was implemented in Elan 4.3.3. Annotators found the speech transcription and the gesture stroke already marked and aligned. The coding scheme for expressions of manner and path was adapted from the Coding Manual: NSF: Crosslinguistic motion event project (2004), which was developed from a coding scheme for Kita and Özyürek (2003) and used in subsequent studies (Allen et al., 2007; Kita et al., 2007; Özyürek, et al., 2005, 2008). The current coding scheme for verbal description was adapted from the Coding Manual: NSF Crosslinguistic motion event project (2004). With respect to the original manual, we added gesture space annotation to capture the difference, if any, in gesture salience between languages (English and Italian) and language groups (bilingual or monolingual).

Manner and Path verbal production ("verb type"): All the cartoons had three main action events. The annotators coded the verbal production corresponding to the target event of each cartoon. For example, they did not code the verbal typology regarding the initial event (e.g., The triangle pushed the tomato) or the final event (e.g., Tomato bumped into the tree) but only the verbal typology of the target event (e.g., Tomato rolled down the hill). There were four categories: IV, 2V, VP and VM. The speakers may describe the target event with 1 verb (e.g., it rolls up the water- coded as IV); 2 verbs (e.g., it ascends rolling to the shore, 2V) or can describe only the path (e.g., it rose up to the shore, VP) or only the manner (e.g., it rolls to the shore, VM).

Manner and Path gesture production ("gesture type"): We coded all the gestures that overlap with speech that refers to the "target event". We coded the spatial pattern for gestures expressing manner and path into seven types: M, P, C, MC, U, J. The key question is whether the gesture encodes manner (M), path (P) or Manner and Path together (Conflate, C). Sometimes a gesture combines more than one type, expressing for example both manner and conflation (MC). For example, when describing the green triangle jumping around a tree, participants might gesture the jump event (Manner) followed by a conflate gesture (jump around the tree). In particular, for jumping, one jump (up and down) or one and half (up and down and up) will be coded as a separate category, J, as it is unclear whether it should be C or M. If a single jump/rotation in one location is followed by a clear C gesture, then it is coded MC (see the first example in the second row "atypical examples" in fig. 1). Gestures that cannot be classified into M, P, C, MC or J were coded as unclear (U). Note that in order for a gesture to be coded as expressing manner, the gesture must have one full rotation; otherwise, it is coded as unclear (U). Typical and atypical examples of gestures are reported in fig. 1, left panel.

Gesture space annotation ("gesture salience"): Coders annotated the space areas where the gesture stroke took place. Gesture saliency was coded for the target gesture performed during description of the target event (e.g. rolls up). To code saliency we followed McNeill, who divided the gesture space into sectors using a system of concentric squares (McNeill, 1992, p. 89-see fig.1, panel on the right). Our annotation coding scheme reflects this notation dividing the gesture space in 2 sectors: “centre” and “periphery” expressed respectively with not salient and salient.

3. Coding scheme reliability experiment
To ensure the reliability of the adopted coding scheme, a subset of the corpus was annotated by three independent coders. For verbal description of motion events, 91 target events were coded. For gesture depiction of motion events and gesture salience the annotators rated 156 gestures. Annotators had the following training before starting their work. They were trained in a one-to-one session going through the manual with one of the authors. In the same session they were also trained on how to use ELAN. This first session lasted about 2 hours. After that annotators worked by themselves on three files. Once they completed the files, they met the author for the second time and individually went through their files with her. The annotators finalized the three files in this session. After that, the annotators went through more files alone. A Kappa statistics (Siegel and Castellan, 1998) was computed on the resulting annotated files.
With regards verb type, the Kappa score was 0.81 ($p<.001$); for gesture type, Kappa was 0.78 ($p<.001$), and for gesture salience, Kappa had the highest score, 0.89 ($p<.001$).

All the features of our coding scheme had Kappa scores above 0.75. The coding scheme is therefore highly reliable. For verb type all the coding scheme features reached a Kappa above 0.75. For gesture typology, one feature had a mild Kappa score and a low $p$ value (Unclear: $K = 0.67$; $p < .001$). It is also worth noting that annotators agreed more on the labelling of Manner and Path gestures (Manner Kappa=0.9, $p<.001$; Path Kappa=0.9, $p<.001$) whereas Jump and Manner + Conflate gestures had intermediate scores (Kappa=0.8, $p<0.001$ for Jump and Kappa=0.79 $p<0.001$ for MC). Finally, the score for Conflate gestures was fairly high (Kappa=0.73, $p<0.001$). With regards gesture salience, both categories had a high Kappa score (Kappa=0.85, $p<.001$ for salient; Kappa=0.9, Kappa<.001).

**4. Conclusion**

Despite the increasing interests in gestures, there are still not many annotation coding schemes shared and by the multimodal corpora community (a notable exception includes Lausberg and Sloetjes, 2009). In this work we illustrated the annotation coding scheme adopted to investigate whether bilinguals change their gestures when switching from a language to the other. The issue has been addressed focusing on verbal and gestural expression of motion verbs (manner, path or conflation in speech and gesture) and on gesture salience. The proposed coding scheme for typology has been adopted from Coding Manual: NSF Crosslinguistic Motion Event Project (2004), whereas for gesture salience it has been applied for the first time, based on McNeill (1992). This coding scheme focuses on both gesture content and form. This is because we wanted to test both the form (gesture space and shape) and the gesture function (manner and path description) with regards to speech. With this report we make available our coding scheme to the community, hoping to contribute to the investigation of gesture/speech interaction.

![Figure 1](image-url)
5. References

Coding manual: NFS Crosslinguistic Motion Event Project. (2004). Nijmegen, the Netherlands: Max-Planck-Institute for Psycholinguistics. (Allen et al., 2007; Kita et al., 2007; Özyürek et al., 2005, 2008)


