

Gestural representation in the domain of animates' physical appearance

Magdalena Lis (magdalena@hum.ku.dk)

Centre for Language Technology

University of Copenhagen

Njalsgade 140, 2300 Copenhagen, Denmark

Abstract

The paper presents a pilot study on gestural representation of entities referring to animates' physical appearance. We identify representational format and form features of gestures referring to entities in this semantic domain, and patterns in their temporal overlap with speech. We furthermore integrate the results with our previous findings from the domain of eventualities.

Keywords: gesture production; semantics; iconics and deictics.

Introduction

Co-speech gestures are spontaneous body movements semantically and temporally rightly related to verbal utterances they occur with (McNeill, 1992, 2005). Those among them that refer to entities by means of similarity (iconics) or contiguity (deictics) are called representational gestures. Gestural representation is partial and a single gesture conveys only some aspects of the whole concept to be expressed. Which aspects are chosen, determines the form of the gesture. Discovering factors that influence these choices can give us insight into speakers' cognitive-semiotic processes involved in gesture production.

Poggi (2008) has suggested that one of such factors is referent's "ontological type, that is, the type of semantic entity [the referent] constitutes." She has proposed that different features of the referent are depicted in gesture depending on which type it belongs to. In our previous work, we have introduced a framework for investigating that hypothesis (Lis, 2012a), proposed a set of referents' types and subtypes relevant for gestural representation (Lis, 2012c) and tested the hypothesis in the domain of eventualities (Lis, 2012b). In our framework, the types are assigned to referents based on the semantic categorization of lexical units (LUs) used to denote the referents in speech. For that, we employ a semantic-conceptual classification of LUs in a wordnet - a linguistic taxonomy. In (Lis, 2012b), we used a wordnet for Polish, plWordNet¹, to identify different semantic subtypes among eventualities. The first subtype encompasses eventualities in which (trans)location and spatial relations are more prominent. Examples of verbs that denote such eventualities are 'run' and 'jump.' The second category consists of eventualities in which body internal motion, e.g. handling, is more salient. Examples include verbs 'throw' and 'open.' Our results showed a strong correlation between the category of the eventuality, non-verbal representation and gesture's form (Lis, 2012b). Gestural representation was coded in terms of

Viewpoint - a notion adopted from McNeill (1992, 2005). We found that gestures co-occurring with clauses containing a verb that denotes an eventuality of a subtype (trans)location and spatial relations tended to be performed from Observer Viewpoint (i.e., they are represented from an external perspective of a viewer) and be single handed with Open Hand or Index Finger handshape. For eventualities of the subtype body internal motion, the gestures were more likely to have Character Viewpoint (i.e., to be pantomimic re-enactments of an eventuality) and be bi-handed symmetrical with a Fist or Claw handshape.

The present study applies the framework to a different semantic domain, namely gestural reference to human and non-human animals' physical appearance, more specifically to entities of the type 'body' and 'garment.' That domain is of particular interest due to the fact that during gesturing, speaker's origo - point of reference - is often allocated to the speaker's own body (Fricke, 2002) creating a space to map entities onto. Moreover, a pattern has been identified for non-verbal reference to body in ASL - American Sign Language (Taub, 2001). Our study investigates whether such pattern exists for co-speech gestures as well.

The data

Corpus

The study was conducted on a part of the Polish Multimodal Corpus. It consists of audio-video recordings of subjects performing two narration tasks. The subjects describe to an addressee 1) a film scene and 2) a set of images and short videos. The data analysed in the paper come from 21 participants, all students (21-24 years old) and native speakers of Polish. Detailed information about the corpus can be found in (Lis, 2012c).

Annotation

Verbal level Speech was transcribed in Anvil (Kipp, 2004). In the annotation, we followed a framework introduced in (Lis, 2012a). In this framework, entity is assigned a semantic type according to the classification of an LU used to denote that entity in verbal utterance. For this a large-scale artificial network called wordnet is used. Wordnet constitutes a taxonomy in which LUs are grouped into sets of synonyms (synsets) and semantic and lexical relations between them are encoded (Fellbaum, 1998). A relation that constitutes a skeleton of the wordnet structure is hyponymy. It defines an LU in terms of a subclass to class. In wordnet, the non-lexical synsets in hyponymy's upper-level hierarchies define semantic domains. A domain denotes a segment of reality symbol-

¹<http://plwordnet.pwr.wroc.pl/wordnet/>

ized by a set of related LUs, which all share a common semantic property. An example of hyponymy hierarchy in the wordnet for Polish, plWordNet 2.0 (Piasecki, Szpakowicz, & Broda, 2009), is given in Figure 1.



Figure 1: PLWordNet’s hyponymy relation for the noun *oko* ‘an eye’

The focus of the present study is on entities of the semantic type ‘body’ and ‘garment.’ On the level of speech, such entities are denoted by nouns. Nouns with the following domains in plWordNet 2.0 were subsumed under the type in question:

- body, organism (e.g., *kregostup* ‘backbone’)
- body part, limb (e.g., *ucho* ‘ear’)
- element of garment (e.g., *spodnie* ‘trousers’)
- body cover (e.g., *wlosy* ‘hair’)
- accessory (e.g., *guzik* ‘button’)

Non-verbal level Co-speech, representational gestures of the hands were annotated on a separate track in Anvil. The analyses reported in this paper focus on gestures co-occurring with clauses which contain LUs denoting entities of the type ‘body’ and ‘garment.’ All annotations regard stroke, i.e., the meaning-bearing part of gesture. To account for differences in gestural representation, we coded the gestures for the techniques employed. Gestural techniques refer to a set of semi-otic devices by which hands turn into representational signs. Each of the techniques emphasizing slightly different aspects of the referent. The proposed set is inspired by Müller (1998) Modes of Representation and consists of the following techniques:

- Depicting – articulators trace an outline or sculpture a shape
- Acting – articulators mime an action
- Embodying – articulators stand as a static model of the referent itself
- Indexing – articulators point within a gesture space

To account for gestures’ physical form, we annotated: Handedness, Handshape, Location and Repetitiveness. Handedness indicates which articulator performs a gesture - right (RH), left (LH) or both hands, and in the last case - whether they are symmetric (2SH) or not (2DH). Handshape regards the configuration of palm and fingers; the values for this attribute were taken from ASL (Tennant & Nelson-Metlay, 1998).² Location refers to a space in which the gesture is performed and indicates whether it corresponds to that of the referent itself. To annotate correspondence, we used meronymy-holonymy encodings in plWordNet 2.0. Meronymy is a semantic relation of a part to whole, holonymy is its opposite. Gesture space was horizontally divided into 3 areas, each given a tag based on meronymy-holonymy encodings in plWordNet: above centre (tag: head, hairstyle), centre (tag: trunk, body cover, upper limb) and below centre (tag: lower limb). Gesture’s location was coded as corresponding to that of the referent if the gesture was executed in the space, whose tag is a holonym of the LU denoting the gesture’s referent. Additionally, gestures were coded for repetitiveness: whether a particular movement pattern occurs once (sg) or multiple times (pl) within a gesture. Each gesture was associated with its lexical affiliate in the speech transcript using MultiLink option in Anvil. Table 1 lists the attributes and values for gesture annotation.

Table 1: Attributes and corresponding values for annotation on non-verbal level

Attribute	Values
Technique	Depicting, Acting, Embodying, Indexing
Handedness	RH, LH, 2SH, 2DH
Handshape	ASL_5, ASL_C, ASL_G, ASL_O, ASL_S, complex
Location	corr., non-corr.
Repetitiveness	sg, pl

Results

Corpus analysis In our data, 514 clauses contained LUs referring to the semantic domains in question. The LUs denoted human body (e.g., *glowa* ‘head,’ *stopy* ‘feet’), body of non-human animals (e.g., *lapa* ‘paw,’ *ogon* ‘tail’) and parts of garment (e.g., *koszula* ‘shirt,’ *obroza* ‘collar’). 325 of those clauses were accompanied by a manual, representational gesture.³

We divided the clauses into two groups, depending on the semantic subtype of the eventuality denoted by their verbal constituent. Verbs were assigned subtypes following the framework described in detail in Lis (2012b). Clauses which

²Due to their similarity, the handshapes B, Open B and 5 were overlapped under ASL_5; A and S under ASL_S; Bent 5 and C under ASL_C; and D, 1 and G under ASL_G.

³Note that those nouns could still be accompanied by hand gestures of other types (non-representational) or representational gestures of other articulators (non-manual).

contain verbs denoting (trans)location and spatial relations or body internal motion belong to the first group. Examples of these clauses are 'he took off his hat,' 'the dog wags its tail.' Group two consists of clauses with state verbs 'to be' or 'to have.' Examples are 'she has a fringe,' 'there is a man in a white T-shirt.' The remaining verbs were collected under the umbrella term 'Other' for future research. 87% of the gestures which co-occurred with the clauses from the group2, represented an entity of the type body or garment. 95% of the gestures coinciding with the clauses from the group1, represented an entity of the type eventuality.

192 of the gestures represented body and garment. The most often used gestural technique was Depicting. The second technique, in terms of frequency, was Indexing. Some of the gestures were tactile. Figure 2 presents speaker's choices of techniques for those gestures.

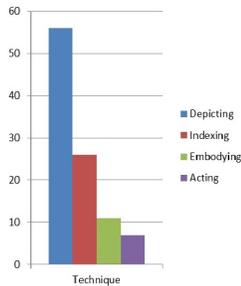


Figure 2: Techniques of gestures representing the entities of the type 'body' and 'garment'

The most frequent, i.e. Depicting Technique, gestures were mostly bi-handed and symmetrical. The most often used handshape was ASL_5. The majority of those gestures were performed at the location corresponding to the location of the referent, as annotated using meronymy-hyponymy encodings in plWordNet 2.0. Almost all were single. The results regarding the physical form of Depicting gestures are given in Figure 3.

133 gestures represented eventualities. Almost all of them co-occurred with the clauses in which nouns denoting body or garment were accompanied by verbs denoting eventualities of the subtype (trans)location and spatial relations or body internal motion. The gestures accompanying clauses with body internal motion verb were more likely to be performed with Acting technique, while the clauses with verbs denoting (trans)location and spatial relations tended to be accompanied by Depicting gestures ($p < .001$, $\chi = 70.42$, $df = 1$).

Overall, there was no statistically significant difference between the two narration tasks with regard to any of the analyses reported above.

Temporal alignment We also analysed the temporal relation between speech and gesture. The analysis was carried out on narrations of the film scene (128 gestures). The LUs were tagged for their Part of Speech (POS) according to the tagset proposed by Petrov, Das, and McDonald (2012). We

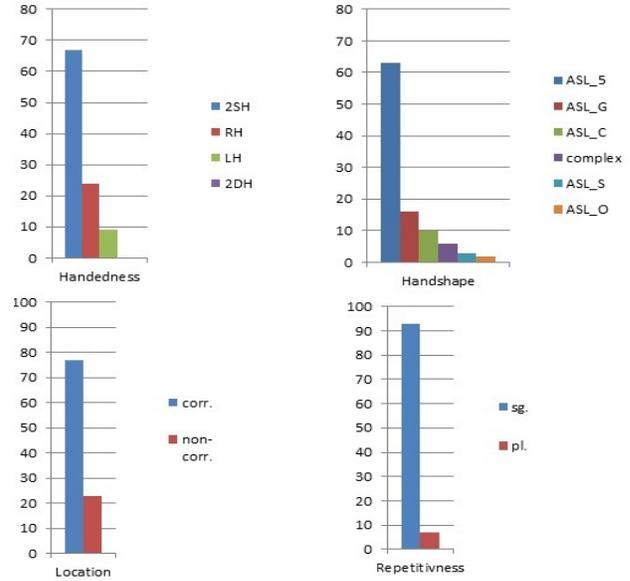


Figure 3: Features and attributes for Depicting Technique gestures

tested with which POS the gesture stroke synchronizes. We found that for the category1 clauses (i.e., clauses with a verb of the type translocation spatial relations or body internal motion), the strokes were more likely to overlap with VERBs, while for the category2 clauses (i.e., clauses containing state verbs like 'to be' or 'to have'), the strokes were likely to synchronize with NPs (ADJ or NOUN) ($p < .001$, $\chi = 57.8$, $df = 1$).

Discussion and future work

Our results show speakers' tendencies in gestural representation of entities of the semantic type 'body' and 'garment.' Location seems to be the most distinguishing feature for the non-verbal representation in that domain: the majority of the gestures were performed in the corresponding gesture space. Pointing gestures were however not very frequent. Depicting Technique gestures executed in the relevant gesture space were used most often. Speakers tended, thus, to represent information not only about referent's location but also aspects of referent's shape. Furthermore, the most frequently used technique had a consistent physical form in terms of Handedness, Handshape and Repetitiveness. The gestures were mostly non-repeated and performed with two symmetrical hands and an open handshape, indicating an approximate shape or an extension of the entity in space rather than a detailed outline.

The semiotic devices used by the gesturers in our study largely correspond to those used by signers. Taub (2001) discusses representation of body parts as a special set of patterns in ASL. The signers' articulators represent human body parts of the same type (hands stand for hands) or animal's body parts corresponding to signer's articulators by visual or functional similarity (hands stand for human feet or animal

forepaws). Those forms are often used productively to encode location of the referent. Most frequently, a combined iconic device is employed: path-for-shape iconicity along with the mapping of (the body of) the referent on the gesturer's body. This corresponds to the gesturers' use of Depicting Technique in the corresponding location, as observed in our data. The fact that the same representational formats are chosen by gesturers and signers to represent entities from the domain of physical appearance may suggest similarities in the online conceptualization of those entities. Furthermore, while representation of body and garment in co-speech gesture is not conventionalized, in sign languages it is. Gestural representation of entities in this domain may, thus, serve as an example of iconically-motivated representation gradually becoming conventionalized.

Moreover, the results provide further evidence for our previous findings. In (Lis, 2012b), we found a strong correlation between eventuality semantic subtype and gestural representation. Also in the present study, eventualities of the subtype body internal motion tended to be pantomimed in gesture (Acting Technique), while gestures representing paths (Depicting Technique) were more likely for (trans)location and spatial relations eventualities.

Furthermore, our results show that semantic and syntactic structure of a clause may provide cues as to which entity will be depicted in the concurrent gesture. More specifically, the findings suggest that the type of entity represented non-verbally may vary depending on the category of the verb used in speech. If LUs denoting the entities of the subtype animate's physical appearance co-occur in a clause with LUs denoting an eventuality from the domain of body internal motion or (trans)location and spatial relations, the eventuality may be more prevalent for determining the form of the gesture co-occurring with that clause. Furthermore, in such cases the stroke of the gesture tends to synchronize with the verb itself. If LUs denoting the entities of the subtype animate's physical appearance co-occur in a clause with state verbs 'to be' or 'to have,' those entities may be more decisive for the form of the gesture. In that case, the stroke is likely to coincide with noun or adjective. The corpus analysis reported in this paper and the results of our previous research (Lis, 2012b) may provide a basis for defining preference rules for gesture production in the combined domains of body and garment and eventualities.

However, the restricted character of our data needs to be acknowledged. The analysis encompasses 325 gestures from a specific type of corpus. Our hypotheses need to be tested on a bigger sample and different types of interactions. No differences in gestural representation were found between the two narration tasks in our corpus, although one of them regarded a set of unrelated images and led to elliptic descriptions, while the other was a fully fledged story. Both were, however, narrations. It remains to be investigated whether there are differences in gestural reference to body and garment between discourse types (e.g., dialogue, physician-patient interaction),

individuals, speakers of different languages and different cultures (e.g., cultures in which articulators other than hands are prominent in indexical functions). Furthermore, not only clause level but the level of a whole discourse needs to be taken into account.

In future research we plan to investigate discourse structure (in terms of theme and rheme) as a potential factor influencing non-verbal representation of entities discussed in the present paper. Moreover, we aim to extend the methodology adopted in the paper to gestural representation of other types of entities (manipulable vs. non-manipulable objects). We also plan to take more holistic view on iconic and deictic gestures by analysing gestures of articulators other than hands.

Acknowledgments

This research has received funding from the European Commission's 7th Framework Program under grant agreement no. 238405 (CLARA). Special thanks to Costanza Navarretta and Maciej Karpiński for their comments.

References

- Fellbaum, C. (1998). *Wordnet: An electronic lexical database*. Cambridge, MA: MIT Press.
- Fricke, E. (2002). Origo, pointing, and speech: The impact of co-speech gestures on linguistic deixis theory. *Gesture*(2), 207–226.
- Kipp, M. (2004). *Gesture generation by imitation - from human behavior to computer character animation*. Florida: Boca Raton.
- Lis, M. (2012a). Annotation scheme for multimodal communication: Employing plwordnet 1.5. In *Proceedings of the focomc'12*.
- Lis, M. (2012b). Influencing gestural representation of eventualities: Insights from ontology. In *Proceedings of icmi'12*.
- Lis, M. (2012c). Polish multimodal corpus - a collection of referential gestures. In *Proceedings of lrec'12*.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
- Müller, C. (1998). *Redebegleitende gesten. kulturgeschichte. theorie. sprachvergleich* (Vol. 1). Berlin: Verlag Spitz.
- Petrov, S., Das, D., & McDonald, R. (2012). A universal part-of-speech tagset. In *Proceedings of lrec'12*.
- Piasecki, M., Szpakowicz, S., & Broda, B. (2009). *A Wordnet from the Ground Up*. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej.
- Poggi, I. (2008). Iconicity in different types of gestures. *Gesture*, 8(1), 45-61.
- Taub, S. (2001). *Language from the body: Iconicity and metaphor in american sign language*. Cambridge: Cambridge University Press.
- Tennant, B. M., R., & Nelson-Metlay, V. (1998). *The american sign language handshape dictionary*. Washington DC: Gallaudet University Press.