

Stand-up Gestures – Annotation for Communication Management

Kristiina Jokinen

Department of Speech Sciences
University of Helsinki

Kristiina.Jokinen@helsinki.fi

Minna Vanhasalo

Department of Finnish Language
University of Tampere

Minna.Vanhasalo@uta.fi

Abstract

This paper deals with the analysis and annotation of gestures which we call stand-up gestures. These gestures distinguish themselves from the flow of verbal information exchange by their regulating and coordinating function in communication, independent from the spoken content. The name also bears reference to stand-up comedy where these gestures occur as part of the normal repertoire of successful performance. Besides analysing the functions of stand-up gestures, the paper also discusses their annotation using the MUMIN annotation scheme and proposes extensions to the scheme in terms of a meta-discursive context level.

1 Introduction

In order to maintain smooth communication, dialogue participants need to pay attention to subtle gesturing by the partner. Gestures seem to have several important functions in communication, ranging from the actual content level contributions (iconic gestures) to the coordination of communication (own communication management and interaction management, see Allwood, 2002; Allwood et al., 2007), and to the giving of rhythm to spoken utterances (McNeill, 2005). Gesture studies have thus been important in sociolinguistics, intercultural communication and behavioral studies, so as to have a better understanding of how human communication takes place. For instance in second language learning, it is important to understand how gestures are used in communication: the students need to learn to observe the relevant communicative signals and to produce suitable gestures themselves. Gestures are also important for computer animations and interaction technology in order to allow more natural interactions with a computer. Besides ECAs (Cassel et al. 2003), recently also robotic companions have developed so that they can recognize gestures and thus become engaged

with multimodal communication (Bennewitz et al., 2007). New application areas are also various game and educational toys that would allow especially autistic or disabled children to enjoy and be empowered by the new technology.

This paper deals with the analysis and annotation of certain kinds of gestures which have a regulating and coordinating function in dialogues. They distinguish themselves from the flow of verbal information exchange in that they not only accompany or complement the spoken content but rather function as independent means for communication management. They are related to interactive gestures (Bavelas & Chovil, 2000) and gestures on meta-discursive levels (Kendon, 2004). We call them stand-up gestures, as they typically single out one word or phrase from the utterance as important making the expression to stand up from the flow of speech, and since they are typical to the normal repertoire of successful stand-up comedy performance.

The paper is structured as follows. Section 2 presents the data and provides two examples of stand-up gestures. Section 3 discusses the MUMIN annotation scheme and its suitability for annotating stand-up gestures. Section 4 discusses multifunctionality of gestures, and Section 5 provides an extension to annotation schemes in terms of extended contexts. Section 6 draws conclusions and points to further research topics.

2 Stand-up Gestures

2.1 Pointing in repairing

In the first example there are four people playing a North-Finnish game called *tuppi* (similar to bridge). Players play as partners, one pair against the other pair, and the main rule is that each player must, if possible, play a card of the suit led. A player with no card of the suit led may play any card, which is called *sakata* (in the example there is a past tense 1. person form of *sakata* – *sakkasin*). When a player has to *sakata*

he usually just plays the least useful card he has. However, at times a player can put forth a really good card and use *sakata* as an opportunity to give a signal to the partner of a desired suit to be played next, this is called *merkkisakkuu*. The players are aware that the meaning of the choice of a card in certain circumstances can be either a neutral *sakata* or a marked *merkkisakkuu*, but they are strictly not allowed to express explicitly which one of the two moves they make when playing the other suit. Most typically the *sakata* situations occur when one's partner, the co-player, is forced to play a different suit on one's card. These situations are also easiest for the players to notice and interpret correctly, because they represent highly conventionalized practice. However, the players also follow closely what cards the other pair puts forth when they have to *sakata*, because these can be – although very rarely are – a *merkkisakkuu*. The better the players can read the game and distinguish between *sakata* and *merkkisakkuu*, the better players they are.

Figure 1 shows the relevant stand-up gesture that occurs when the players discuss the game they have just finished. M and his partner T have severely lost, and M had given an explanation for their losing (line 1): he had misread a neutral *sakata* as a marked *merkkisakkuu*. This is misunderstood by A who asks for clarification – the other initiated other repair – on line 3 with the question *which spade*. Soon after asking this A makes a self initiated self repair with the turn *oh the one that I:: did sakata* with an accompanying pointing gesture, Index Finger Extended. The gesture is on the elongated pronoun *I::*, i.e. it points out the most important word of the sentence, and of the repair sequence.

What happens in the dialogue is that A first misinterprets M's reference of *merkkisakkuu* to be some spade played by M's co-player, but then understands that M actually means the spade she herself had *sakata*. A's understanding is evident when looking at her self correction together with the gesture. The stand-up gesture points out the most important word from the utterance, i.e. the correction of misunderstanding, and is made towards M. The reason of the original misunderstanding can be spelled out explicitly as follows: “Which spade played by your co-player? Oh you mean the spade that I (and not your co-player) have to *sakata*.” The misunderstanding and its solution is conveyed by the accompanying gesture which is synchronized with the relevant word of the repair indicating which part of the misunderstanding is the repairable part.



Figure 1. "oh the one that I did sakata"

- 1 M: mä luulin et se pata olis ollu
 2 me(h)rkkisakkuuh hh
I thought that the spade would have been a merkkisakkuu
 3 T: hehe
 4 A: \$m(h)ikä patah\$.hhhh ai se
 5 minkä [mä:] sakkaasin
\$which spade\$.hhh oh the one that [I::] did sakata
**[LH Extended index finger flicks to M
 hand rotating from palm lateral position to palm
 up position. Arm rests on the table whole time.]**

Furthermore, with her repair, A also shows that she trusts that all the players have common ground and general knowledge about playing the game: she does not explicitly explain why she had the kind of misunderstanding she had. Her repair *the one I did sakata* is interpreted correctly by all the parties with (and in) a flick of a hand.

2.2 Pointing in Managing Information

Our other example is from a situation where two young women chat over a lunch. One of them is telling stories about a janitor and snow plowing, and the relevant stand-up gesture occurs between two story-telling episodes. The first story goes as follows: “Our janitor has started plowing snow again, and as you remember from last winter, he used to plow snow at an inconvenient time at 5 am., the snow tractor made utterly annoying noise, and the job took 2 hours to finish even if the yard to be plowed is really small”. The second story concerns the janitor doing snow plowing again this year. However, to justify the story about the same janitor doing the same task, the narrator gives a piece of new information that explains why the follow-up story is new and its telling worthwhile. One of the complaints of the snow plowing last year was that the janitor started to work too early. The new information is that the janitor now starts later, and this fact had allowed the narrator to observe the janitor's working in a more detailed way: she now knows why the snow plowing takes so much time. The new information is accompanied by a pointing gesture

that singles out the newsworthy content (Figure 2), and in the follow-up story the narrator gives an account for the lengthy snow plowing.



Figure 2. "started a bit later than before"

1 N: .hhh me \$j(h)ust niinku-\$ nyt se on
 2 alottanu vähä [**myöhemmin**] s*i*tä niinku,
\$We just like-\$ Now he has started a bit [later] than before like,
[RH index finger stretched, slightly crooked, palm towards oneself, points quickly straight to the interlocutor twice]

When the narrator marks the word *later* with the gesture, not only does she mark the word as important in relation to the content of the story already told, but it also refers to the parts to come: the new detail in the shared information expounds the premises for understanding the conditions of the new story to come. The narrator has been able to watch the snow plowing exactly because it has been done at a reasonable time in the morning when she has been awake. This is not explicitly said in words, but conveyed together with the gesture in the given context.

3 Gesture Annotations

For the applications and research mentioned in Section 1, it is important that the collection and analysis of large multimodal corpora are available and accompanied with rich annotations comprising of verbal and non-verbal phenomena. For instance the AMI corpus (Carletta, 2006) is a large video corpus of meetings and spontaneous interactions and it is accompanied with annotations that also deal with multimodal aspects of communication. Several other video corpora and annotation schemes have been developed as part of projects or individual efforts, see e.g. Martin et al. (2007), and the examples in this paper.

As part of our analysis of stand-up gestures, we have used the MUMIN annotation scheme (Allwood et al., 2007) which is intended as a general instrument for the study of hand gestures, facial displays and body posture in interpersonal communication. The annotation scheme contains categories to describe the form and dynamics of

communicative elements as well as their function in managing feedback, turn-taking, and sequencing. The distinctive feature in the scheme is the use of semiotic categories to encode elements as semiotic signs: *Indexical Deictic*, *Indexical Non-deictic*, *Iconic*, and *Symbolic*.

Considering the analysis of stand-up gestures, their description is distributed among the categories for interaction and communication management. The MUMIN scheme provides annotation categories for their form (hand shape, orientation, location, direction of the movement) and functioning in the information structure (opening, continuing or closing of topics; emphasis), turn management (opening, holding, yielding, etc.), and sequencing (opening, continuing or closing speech act sequences). This is useful when interpreting communicative signs via a dynamic process where the combination of characteristic features determines the sign's interpretation, i.e. gesture signs are not fixed categories but form a continuum along the defined features. By defining elementary features and modelling their combinations it is possible to construct a flexible framework in which similarities and interpretations of various communicative gestures can be compared and studied. From the computational view-point, this supports modelling and experimentation with various classification and clustering algorithms.

Since gestures are multifunctional and multidimensional, this many-to-many nature needs to be incorporated in the annotation. However, the analysis of stand-up gestures also seems to require understanding of the linguistic, pragmatic and social contexts in which they occur, and how the different contexts affect the layering of more than one meaning and function on a gesture. We will return to the different contexts in Section 5, but will first look at the example gestures and how their form, meaning and functions are motivated by the lexical affiliate, parts of speech and common ground between the participants.

4 Multifunctional gestures

4.1 Local meaning and function

Kendon (2004) has identified different gesture families, e.g. Open Hand Prone and Open Hand Supine families. Based on his observations he suggests that each gesture family has its own *semantic theme*. Gestures in Open Hand Prone family express in general stopping or halting of an action (own or other), whereas those in Open Hand Supine family express general offering and

giving of ideas and concepts. According to Kendon the Index Finger Extended is yet another gesture family which, however, has not been thoroughly identified nor classified. The main semantic theme of the Index Finger Extended family seems to be the same as that of the Open Hand families with one distinction: the gestures in this family are precise and explicit. Our analyses of the two stand-up gestures support this distinction. The gestures explicitly single out the important word of the utterance, i.e. the one that refers to what has been repaired in the previous misunderstanding or is the relevant new content in the story telling episode. The exact hand shape, index finger extended is motivated by the communicative needs on the utterance level (to point out a particular expression from the speech), while the orientation of the palm is motivated by the needs of communication management (halt conversation, offer information).

4.2 Communication management

In example 2.1, A's index finger is oriented horizontally and the hand rotates from a palm down position to a palm up position, thus offering the repair to the interlocutor. This resembles the Palm Open Supine family's semantic theme of giving and offering. In example 2.2 the index finger is obliquely horizontal, but the palm is facing the speaker (i.e. back of the hand is towards the listener) and the finger points straight to the interlocutor. The narrator halted telling the second story for a moment in order to give some new information with respect to the given information (i.e. the annoying snow plowing starts later in the mornings this year). This resembles the semantic theme of stopping and halting of the Open Hand Prone (vertical) family.

Allwood (2002) talks about own communication management and interaction management, referring to the aspects of communication that concern meta-level control of the interaction and can include such functions as repairs, initiations of topics, direction of the focus of attention, etc. In example 2.1 the gesture in conjunction with the repair of one's own speech belongs to the own communication management plane. Furthermore, the orientation of the gesture, palm up, is sensitive to the local negotiation of context. The speaker knows that she has made a correct repair of the person who used the spade for *sakata*, and with the orientation of her gesture she signifies her understanding and hands the understanding of a successful repair over to the interlocutor. The gesture in example 2.2, however, manages

the structuring of information. The palm orientation of the pointing gesture shows that the speaker is temporarily halting the flow of storytelling but not halting it altogether. She is not merely offering a new piece of information but rather stopping the storytelling in order to give the particular piece of information that motivates the later story. The palm orientation away from the listener cuts the interlocutor's opportunity to take the floor during the stop.

Gestures are often related directly to the information flow of the dialogue. However, stand-up gestures require that the speaker is aware of the means to coordinate the conversational situation and to focus the partner's mind on some particular aspect in the exchanged information or to prepare the partner to have the right stance in order to interpret the message in the intended way. Stand-up gestures often occur in everyday contexts (as in our examples) where the speaker controls a story telling situation and indicates the start of a new topic, a repair, or otherwise important new information. They also often indicate the speaker's dominance over the floor, since the speaker can thus coordinate the flow of information, turn-taking, and interpretation of the presented ideas. The speaker's role as the initiator of a topic also allows her to control the topic management, to continue or close the chosen topic. This kind of control can be especially seen if the speaker has a dominant role in the activity (e.g. chairing a meeting), and in storytelling situations and stand-up comedies where the gestures are frequently used to manage the flow of information and lead the story towards its punch-line.

Instead of getting their meaning from the content of the verbal flow of information, stand-up gestures indicate to the partner non-verbally how the conversation is to be understood and divided into communicatively important segments. They are distinguished from the normal flow of information so as to catch the partner's attention, and by so doing they also control the dialogue flow.

4.3 Social Interaction

With the notion of catchment McNeill (2005) refers to the social-interactive nature of all gestures: gestures have an active role in creating, shifting and updating the common ground between the interlocutors. Catchment is used only in the context of cohesives, i.e. similar kind of gestures that keep recurring in the dialogue. We propose, however, that not only repetitive cohesives, but single gestures (the stand-up gestures) can create and indicate the common ground be-

tween interlocutors. In other words, catchments can be seen as part of the constant ongoing negotiation of context in conversations – yet another level of context in addition to the local utterance and communication management levels.

Pointing straight to the interlocutor is usually considered insulting unless the social relationship is such that this is acceptable, in which case pointing can act as a bonding strategy. For instance, in example 2.2 the narrator recognizes, that the previously given information (the first story) is already part of their shared knowledge, so she starts the storytelling with summoning *as you remember from last winter*. By making the pointing gesture straight to the interlocutor, the narrator also seems to want to gain heightened attention of the interlocutor: it is at this point that the truly new information begins which the interlocutor has not heard before. The narrator has thus taken their social relationship into account and acknowledged their long shared history of similar discussions: the gesture points out that the follow-up story will update their shared knowledge of the janitor and snowplowing, and therefore asks for intensive attention.

The last interesting observation is that the gesture in 2.1 is made in the periphery, whereas the gesture in 2.2 is made in a more central place. This can be accounted for with the help of the notion of common ground. Bavelas & Gewing (2004) showed that interlocutors use less explicit, smaller and peripheral gestures when reference is made to the common ground, and when the reference is not to the common ground, the gestures become larger, more central and explicit. In example 2.1 the speaker is handing over information that is self explanatory for all the parties because it is based on their common ground: shared knowledge of the game conventions. The gesture is thus rather small and peripheral. In 2.2, however, the narrator updates the common ground as she is about to move from the first story (given information) to the follow-up story (brand new information), and the gesture is consequently larger and more central. The place of the stand-up gesture can thus be said to be motivated by the social interactive level where the notion of common ground explains the choice between the periphery and central place.

5 Stand-up Gestures and Context

As shown above, interpretation of the gesture is related to the context in which the gesture occurs. The context influences the form and func-

tion of the gesture, and depending on the closeness of the interlocutors' relationship, also the gesture's acceptability and interpretation. Concerning interactive situations, we especially like to emphasise the communicative context in terms of activity types and the speakers' roles (cf. Levinson, 1992; Allwood, 2002). Activity types impose constraints on acceptable contributions in a given communicative context, and roles set up strong expectations on the appropriate behaviour and how contributions should be interpreted.

Often, however, gestures have different relations to their context, or the relation of the gesture to its context is not explicitly spelled out: the gestures are multi-contextual. Figure 3 depicts the five different context levels that we consider important when analyzing gestures.

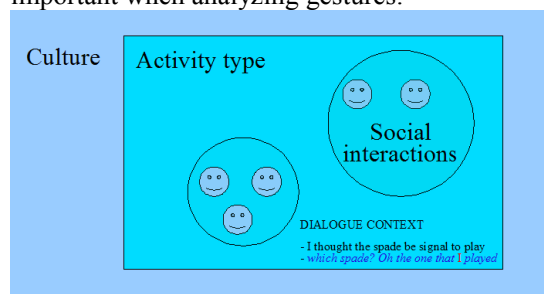


Figure 3 Contexts that influence the form, meaning and function of a gesture.

The most local context is the representational context of the gesture, which consists of a lexical affiliate. For instance, a beat can give emphasis on a word highlighting it, and an iconic gesture can express semantic features of the referent by similarity or homomorphism. A stand-up gesture also singles out the most important word of the utterance and thus resembles beats, but rather than being repetitive and rhythmical as beats, a stand-up gesture is a single “stand-alone” gesture. The next level context is the dialogue context. Gestures operating on this level deal with the relationship between speech segments, sequencing and structuring of information, managing contributions and turn taking (“what I said previously”, “the next point”, new vs. given information, repairs). The third level context deals with social interactions. Gestures on this level denote the relationship between interlocutors and the common ground between them. The fourth context level concerns the activity type that the speakers are engaged in (ranging from everyday chatting to task-oriented discussions, from informal events to formal performances). For instance, pointing a finger to the listener of a story or to the audience of a comedy act asks for heightened attention to the shift in focus: the ges-

ture indicates that there is a transition from an old story to a new one, or that the punch line is coming. The largest context is the cultural context that concerns social norms and relationships, i.e. culturally conditioned behaviour patterns that limit the appropriateness and interpretation of gestures. Emblems are typical examples of gestures on this level.

The five contexts interact with each other and the larger contexts usually affect the more specific ones. Each context also influences the form and function of the gestures in various degrees. We propose that the contexts be taken into account in the MUMIN annotation scheme. They can be included as a special annotation feature “Context” with five values (lexical, segmental, social, activity, culture) or via a more sophisticated linking system based on the gesture’s multifunctionality and multidimensionality.

As discussed in Chapter 3, the hierarchical feature-based annotation seems reasonable compared with a simple gesture categorisation, especially when thinking of the continuum that different gestures make with respect to their form and function in general. However, as always with annotations, an important yet open issue is how much detail will be sufficient in the annotation scheme without getting too deep into the micro-analysis of gestures and lose useful generalisations. On one hand we have views about highly organised interactions where no phenomenon is too small to be considered meaningful (cf. Goodwin, 1984). On the other hand, there are practical goals and needs for developing models for interactive systems for which a certain level of generality, frequency, and categorisation is desirable and necessary. Gesture families as suggested by Kendon (2004) seem useful in this respect.

6 Conclusions

We have discussed the form and function of stand-up gestures on the basis of corpus examples. The gestures are important in coordinating interaction on meta-discursive level: constructing common ground and regulating information flow so that the verbal activity is not disrupted. The speakers need to learn how to distinguish communicatively meaningful gestures from those that do not matter, and also to provide a correct interpretation for them. It is through this kind of gestural communication that the speakers construct mutual knowledge and create social bonds.

We have also proposed five contextual levels in which the gestures can be interpreted: linguis-

tic, dialogue, social interaction, activity type and cultural context. For various applications and further modelling (e.g. gesture lexicons for ECAs and in human communication studies), the contexts need to be included in the annotation scheme, so as to be able to describe gestures on a meta-discursive level where they can be related to the whole dialogue and the dialogue partners.

References

- Allwood, J. 2002. Bodily Communication - Dimensions of Expression and Content. In Granström, B., House, D., Karlsson, I. (Eds.) *Multimodality in Language and Speech Systems*. Kluwer. pp. 7-26.
- Allwood, J., Cerrato, L., Jokinen, K., Navarretta, C., Paggio, P. 2007. The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. In Martin, J. et al. (Eds.), pp. 273–287
- Bavelas, J., Chovil, N., 2000 Visible acts of meaning. An Integrated Message Model of Language in Face-to-Face Dialogue. *Journal of Language and Social Psychology* 19(2), 163–194.
- Bavelas, J., Gerwing, J. 2004 Linguistic influences on gesture’s form. *Gesture* 4(2), 157–195.
- Bennewitz, M., Faber, F., Joho, D., Behnke, S. 2007. Fritz - A Humanoid Communication Robot. Procs of the 16th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).
- Carletta, J. (2006) Announcing the AMI Meeting Corpus. *The ELRA Newsletter* 11(1), 3–5.
- Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (Eds.) 2003. *Embodied Conversational Agents*. MIT Press, Cambridge, MA
- Goodwin, C. 1984 Story structure and organization of participation. In Atkison, J.M., Heritage, J. (Eds), *Structures of Social Action. Studies in Conversation analysis. Studies in Social Emotion and Interaction*. Cambridge University Press. pp. 225–246.
- Kendon, A. 2004. *Gesture: Visible action as utterance*. Cambridge.
- Levinson S. 1992 Activity types and language. In Drew P., Heritage J. (Eds.) *Talk at work*. Cambridge University Press. pp. 67–107.
- Martin, J.C., Paggio, P., Kuenlein, P., Stiefelhagen, R., Pianesi, F. (Eds), Multimodal corpora for modelling human multimodal behaviour. *Special issue of the International Journal of Language Resources and Evaluation*, 41(3–4).
- McNeill, D. 2005. *Gesture and Thought*. University of Chicago Press, Chicago and London.