Verbal or visual? How information is distributed across speech and gesture in spatial dialog

Kirsten Bergmann, Stefan Kopp
Artificial Intelligence Group
University of Bielefeld
P.O. 100131, 33501 Bielefeld, Germany
{kbergman, skopp}@techfak.uni-bielefeld.de

Abstract

In spatial dialog like in direction giving humans make frequent use of speech-accompanying gestures. Some gestures convey largely the same information as speech while others complement speech. This paper reports a study on how speakers distribute meaning across speech and gesture, and depending on what factors. Utterance meaning and the wider dialog context were tested by statistically analyzing a corpus of direction-giving dialogs. Problems of speech production (as indicated by discourse markers and disfluencies), the communicative goals, and the information status were found to be influential, while feedback signals by the addressee do not have any influence.

1 Introduction

In spatial dialog like in direction giving, humans make frequent use of speech-accompanying gestures. By ”gesture” we mean expressive movements of the hands and arms while speaking. According to (McNeill, 2005) there are four semiotic dimensions within these gestures, as there are iconicity, metaphoricity, deixis and temporal highlighting (beats). Iconic features of gestures present visual information about concrete referents, while metaphoric features refer in the same way to abstract referents. Deictic features point to concrete or abstract referents within the external space, and beats are small and fast movements that structure utterances. One often finds several of these features mixed in the same gesture. This paper focuses on gestures that have their major dimensionality in iconicity and deixis, and we present a study that investigates how information is distributed across these gestures and their concomitant speech.

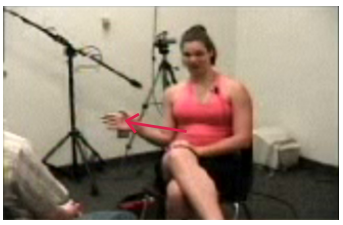

Figure 1: Gesture accompanying the utterance ”take a right” as an example for gestural redundancy.

Gestures are temporally coordinated with speech as well as closely related to the content of the verbal utterance they accompany (McNeill, 1992). The semantic synchrony of both modalities can be thought of as a continuum of co-expressivity, with gestures encoding completely the same aspects of meaning as speech on one extreme. Although both modalities express information in their specific way, we refer to this as *redundancy*. Figure 1 gives an example for redundant meaning in speech and gesture. The utterance ”take a right” contains an action (”take”) and a direction (”right”). Both aspects are expressed as well by the accompanying dynamic gesture made to the right. That is, these two features are communicated redundantly by speech and gesture. At the opposite extreme of the continuum there are gestures encoding aspects that are not uttered verbally, in other words these gestures *complement* speech. In figure 2 an illustrating example for complementarity is given. The direction giver talks about an entrance and visualizes the entrance by gesture. The major content conveyed by speech is the existence and function of an entity, namely

brandial 2006: Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue, pages 90–97.

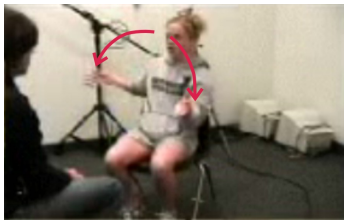

Figure 2: Gesture accompanying the utterance "it's the entrance" as an example for gestural complementarity.

being the entrance. Without the accompanying gesture the recipient's mental representation of the entrance could take different shapes, but the gesture visualizes the arch-shaped architecture of this specific entrance. So the specification of the entrance's shape is a complementary feature of the speech-accompanying gesture. Interestingly, there seems to be a 50:50 distribution of redundant and complementary gestures (Cassell et al., 2000; Cassell and Prevost, 1996), and even the blind distribute semantic components across the modalities (Iverson and Goldin-Meadow, 1998). The question is when people gesture at all, how they distribute information across speech and gesture. What are the influencing factors? So far, research has not been able to give any satisfying answers on this. McNeill (1992) contends that representational gestures are more likely to be used for newsworthy concepts. Cassell and Prevost (1996) analyzed manner-of-motion verbs and accompanying gestures using semantic features to distinguish between redundant and complementary gestures. They found rhematic information with a focus marking newness or contrast resulting mainly in complementary gestures, while thematic information with a focus marking contrast is accompanied mainly by redundant gestures. Yan (2000) studied gestures from a house description experiment using semantic features to classify redundant and complementary gestures. He developed a hierarchy of rules that managed to predict 60% of the gestures. His major findings are that the introduction of single/multiple object(s) is accompanied by complementary gestures, while redundant gestures are used to localize objects. Bavelas et al. (2002) report findings suggesting that gestures are used to compensate for problems of verbal encodability. Kita and Özyürek (2003) found cross-linguistic variations in iconic gestures, indicating that gestures are shaped simultaneously both by spatial properties of the referents and the way the spoken language packages information. Furthermore, Melinger and Levelt (2004) found first direct evidence that the decision to gesture influences decisions about what is explicitly mentioned in speech or is omitted.

The aim of this study is to find factors that can explain the observed occurrence of redundant and complementary gestures. For this purpose we have included both, meaning itself and the wider dialog context in the analysis. A level for comparing the semantics of speech and gesture has to be established firstly. The following steps aim at problems of verbal encodability as well as different kinds of feedback that signal understanding or non-understanding. Moreover, the particular communicative goals, as Denis (1997) identified them in route directions, as well as the information status might have an impact on the co-expressivity of speech and gesture as well.

2 Method

Our corpus analysis takes place within the scope of a study done at the Northwestern University in Chicago (Kopp et al., 2004). In the following, this study is described briefly, supplemented by a description of the annotation scheme developed for our purpose.

2.1 Participants

28 undergraduates (11 males and 17 females) participated in the experiment as direction givers. All of them were native speakers of English. They got the task to describe a route across Northwestern University's Campus to another person they thought was unfamiliar with the campus.

2.2 Materials

Ten different routes existed, each of them starting at the building where the experiment took place, and connecting five locations on the campus.

2.3 Procedure

Each direction giver got a list of ten routes and was asked to sort out those ones she/he did not feel comfortable to give directions for. Among the remaining routes one was selected randomly. In order to guarantee comparable conditions, the participant was instructed to make her-/himself familiar with the route by walking it. Afterwards she/he was seated face-to-face with the direction follower. They were instructed to make sure that

the direction follower understood the directions and would be able to find the way on her/his own. Audio- and videotapes were taken of each dialog. For the videotape, four synchronized camera views were recorded.

2.4 Coding

Some annotation has been done in the scope of other studies (Kopp et al., 2004; Kopp, 2005). This includes the transcription of the direction giver's words and the segmentation of the occuring iconic and deictic gestures. Moreover, the gesture morphology has been annotated, that is *hand shape*, *hand orientation* and *hand location*. The latter includes shape and extent of the trajectory, which is used to judge the gesture's semantics in the following.

In the scope of our own corpus analysis a total of 1508 gestures out of 10 different dialogs were annotated by two coders using the tool *Praat*[1] to transcribe the words of the direction follower, and the multimodal annotation tool *Anvil* (Kipp, 2004). The following levels of annotation have been added to the corpus: (1a) speech semantics, (1b) gesture semantics (2) problems of verbal encodability, (3) dialog acts, (4) communicative goals, and (5) information status.

2.4.1 Semantics of speech and gesture

The central annotation levels are gesture and speech semantics. In a first step, the lexical affiliate of each gesture, i.e. the word(s) deemed to correspond most closely to a gesture in meaning, has been determined (Schegloff, 1984). For each utterance one or more semantic features (SFs) were annotated both for the gesture and its lexical affiliate. Judging the semantics of speech and gesture is not an easy task. Because of their underspecifity gestures can not be interpreted without looking at their verbal context. Therefore, the risk of circularity is given, when gesture and speech semantics are overhastily equated (McNeill, 2005). To devoid this circularity, we first determined the idea unit underlying the multimodal utterance. Based on this, we judged the semantic contributions of both modalities. An example of the procedure of semantic interpretation is given in figure 3. The utterance refers to "Cook Hall", and the underlying idea unit encloses information about the appearance and location of this entity. The information

[1] http://www.praat.org

Figure 3: Gesture accompanying the utterance "you're gonna see a big building to your right" with additional information about the referent in the form of map and photo.

that the referent is a building with a rectangular front emerges from the photo, while a look at the map reveals that "Cook Hall" is on the right side of the route being described. With this additional information one can fasten down the distribution of information: Verbally the direction giver introduces an entity as "big building" which is to the "right". According to this, we assign the following SFs to the verbal utterance: ENTITY, SIZE and RELATIVE POSITION. The accompanying gesture visualizes the shape of the building. Additional information about the relatively large extent of the gesture is adopted from gesture morphology. Thus we annotate the SF categories SHAPE and SIZE for this gesture, not relative position since the gesture is made in front of the speaker and not to his right.

For the overall set of SFs, semantic categories developed by Jackendoff (1983) have been modified depending on the domain of spatial discourse. The following categories adequately cover the semantics of both speech and gesture in our corpus, given with the rules used to annotate speech semantics:

- ENTITY: Streets, paths, buildings, signs etc.
- RELATIVE POSITION: Prepositions characterize information about the spatial position of entities, e.g. "on your left" or "behind the parking lot".
- ACTION: Information about actions, verbally conveyed by motion verbs like "walk", "go", "head", "follow" etc.
- DIRECTION: Directional information concerning actions is realized verbally with adverbs like "left/right" or

”north/south/west/east”.

- PATH: There are three variants of paths: (1) bounded paths, characterized by prepositions like ”from” or ”to”, (2) paths along a reference object, characterized either by verbs like ”pass” or by prepositions like ”along”, ”through” or ”around”, and (3) paths running relative to a reference object, characterized by verbs like ”follow” or by prepositions like ”on”.

- SHAPE: Words like ”circular” or ”zig-zaggy” are annotated as shape.

- SIZE: Adjectives like ”huge” or ”small” are coded as size.

- AMOUNT: An amount of entities can be verbalized by numerals or by words like ”several” or multiple”.

- PROPERTY: Other properties of entities, except size and shape.

Concerning the meaning of gestures the same categories are used. The first decision to be made is applied to the dynamics of each gesture. A gesture can be either dynamic or static. Dynamic gestures include a trajectory between starting point and target point, while static gestures only consist of a posture at a target position. In the latter case either RELATIVE POSITION, SIZE or AMOUNT are taken into consideration. Typically, positioning gestures are done with one hand, while sizes are visualized with both hands, but in case of doubt the (verbal) context is decisive. If two entities are localized, AMOUNT is annotated additionally. For dynamic gestures there is a wider range of possibilities. In a first step one has to distinguish gestures referring to actions and gestures referring to entities. For the latter ones the SFs SHAPE, SIZE, AMOUNT and PROPERTY are considered. Supportive for the coder is a look at the gesture morphology where gesture shapes may be found (Sowa, 2006). If the gesture conveys a SHAPE, typically the trajectory or the inner sides of the hands form it. SIZE can be found in a dynamic gesture as well, because sometimes a ”scaling” movement refers to the size of entities. Moreover, the morphology clearly contains information about the extent. Typically, AMOUNT is assigned to a dynamic gesture if it refers to more than two entities. In these cases RELATIVE POSITION is annotated as well. PROPERTY is used if any properties of entities except the above ones are visualized, e.g. smoke out of a chimney. If the gesture refers to an action, we annotated the SF ACTION in either case. In addition, either DIRECTION or PATH are conveyed. Directional gestures are pointing gestures, visualizing the direction of an action, while paths are visualized with a ”sweeping” movement of the hands. Sometimes the SHAPE of the path is depicted additionally.

2.4.2 Verbal encodability problems

We coded two different characteristics for problems of verbal encodability: discourse markers and disfluencies. Both kinds of characteristics have been coded for their occurrence (either within the particular gesture’s lexical affiliate, or directly before it). A special case of discourse markers are hedges, which are defined as ”words whose job it is to make things more or less fuzzy” by Lakoff (1972). ”kind of”, ”sort of”, ”somehow”, ”like” etc. are considered to be *more fuzzy hedges*. Disfluencies reflect production problems coming along with spontaneous speech. According to Shriberg (1999) the following features are coded as disfluencies: (1) filled pauses (”uh”, ”um”), (2) repetitions (”the the”), (3) repairs (”that’s called Cook Buil- Cook Hall”), and (4) false starts (”and then you gonna may- once you get to the end of the building”).

2.4.3 Dialog acts

Following the annotation scheme DAMSL (Dialog Act Markup on Several Layers) by Allen and Core (1997), we analyzed how the co-expressivity of speech and gesture is influenced by (non-)understanding signals. In DAMSL, forward looking functions state how an utterance constrains the future actions or beliefs of the hearer, and affects the discourse. We used the utterance tags *Statement*, *Influencing-addressee-future-action* and *Info-request*. Backward looking functions indicate how the current utterance relates to the previous dialog. We coded the utterance tags *Acknowledge* (”okay”, ”aha” etc.), *Repeat-Rephrase*, and *Completion* as understandings signals, and *Answer* referring to the forward looking info-requests.

2.4.4 Communicative goals

In terms of the communicative function of a dialog act, according to Denis (1997) two major components can be identified, as there are actions/instructions and striking points along the route, so-called landmarks. Based on this, Denis develops several categories of communicative goals that can be distinguished in route directions. Our segmentation of these categories in the corpus is based on the preceding annotation of forward looking functions. Utterances tagged as *Statement* or *Influencing-addressee-future-action* were assigned to the following categories:

- *Reorientation*: Instruction to change the orientation, e.g. ”turn right”
- *Locomotion*: Instruction aimed to reduce the distance between the actual position and the destination, e.g. ”go straight on”
- *Action+Landmark*: Instruction combining action and landmark, e.g. ”cross X”, ”turn left at X”, ”go past X”
- *Landmark*: Reference to landmark without localization or further description
- *Landmark with spatial orientation*: Localization of a landmark, e.g. ”there’s a road in front of you”
- *Landmark description*: Non-locating description of landmarks, e.g. ”it’s a big pink colored building”

2.4.5 Information status

Finally, we coded the information status for each SF using the following states: *new* for SFs introduced in the dialog, *evoked* for SFs already given verbally and *evoked by gesture* for SFs already given only by gesture.

In general, annotation-based corpora depend on subjective judgements of the coders, and reliability of these judgements is mandatory. We reached a mean Kappa value of κ=0.774 (SD=0.101) indicating substantial agreement among the two coders on a test set of about 20% of the corpus. Especially judging speech and gesture semantics with a set of categories is always difficult and approximative, and one could imagine more or less categories. We established our category set iteratively in order to adequately cover the relevant meaning, while at the same time ensuring reliability.

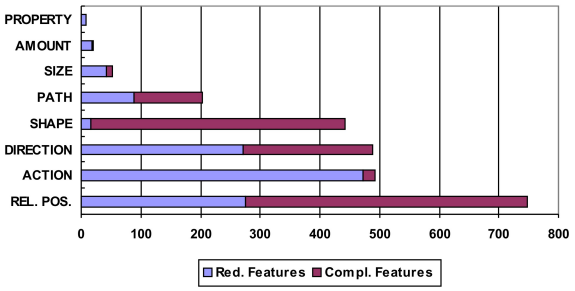

Figure 4: Distribution of the different kinds of redundant/complementary SFs.

3 Results

In the course of judging the gesture semantics, each gesture got assigned between one and five SFs: 51,1% of the gestures have one SF, 31,8% of them have two SFs, and 17,1% have three or more SFs. Among these SFs, 48.63% are redundant while 51.38% are complementary to the accompanying speech. This distribution supports earlier findings by Yan (2000) and Cassell and Prevost (1996) on a level of semantic features. In terms of gesture-wise consideration, one finds 31.7% of the gestures being completely redundant, that is they do not have any complementary SFs. Another 38.9% of the gestures do not have any redundant SFs and therefore are exclusively complementary. Finally 29.7% of the gestures have both redundant and complementary parts. Figure 4 summarizes the number of times that different types of SFs occur in gestures.

The first analysis of the corporal data concerns problems of verbal encoding that become apparent in discourse markers and disfluencies. If there are any discourse markers in speech, there is a significantly higher proportion of complementary SFs in the accompanying gestures (χ^2=13.625, df=2, p=0.001). In addition, the frequency of redundant SFs is decreased in these cases (χ^2=24.279, df=2, p$<$0.001). Concerning redundancy the same findings hold for disfluencies. Gestures accompanying disfluent utterances also have a significantly lower proportion of redundant SFs (χ^2=6.813, df=2, p=0.033), while there is no correlation of disfluencies and complementarity (χ^2=2.128, df=2, p=0.345). Compared to the overall temporal occurrence of gestures in our corpus, gestures accompanying discourse markers or disfluencies oc-

cur more frequently before their lexical affiliate (e.g. ”there is a [little... like...] kind of an alley”).

Further analysis has been done concerning the influence of the direction follower’s feedback on the distribution of information across the modalities. This feedback manifests either in interposed questions or in understanding signals. Regarding interposed questions there is neither a significant relationship between the resulting answers of the direction giver and redundancy in gesturing (χ^2=3.272, df=2, p=0.195), nor is there any influence of backward-looking utterances and complementarity (χ^2=1.604, df=2, p=0.448). Regarding positive feedback of the direction follower, the time passed since the last understanding signal may be relevant for judging the influence of these signals on co-expressivity of speech and gesture. The following time-intervalls have been tested: 0-4.99s, 5.00-9.99s, 10.00-19.99s, 20.00-29.99s and >30.00s. Across all intervall lengths, we did not find any significant influence of utterances tagged as *Acknowledge*, *Repeat-Rephrase* or *Completion* on the number of redundant SFs of speech-accompanying gestures (χ^2=7.079, df=8, p=0.528), nor on the frequency of complementary SFs (χ^2=8.325, df=8, p=0.402).

Furthermore, we analyzed the influence of communicative goals on the frequency of gesturing in general. The majority of annotated communicative goals is accompanied by exactly one gesture (76.3%), while 10.9% do not have any accompanying gestures, and 12.9% are accompanied by two or more gestures. Nevertheless this distribution depends on the kind of communicative goal (see figure 5). Descriptions of actions without any reference to landmarks (*Reorientation*, *Locomotion*) as well as utterances of the category *Landmark* do have one accompanying gesture in the majority of cases. *Landmark descriptions* are more often uttered without gesturing, while *landmarks with spatial orientation* tend to go with two or more gestures.

In addition, we tested the influence of communicative goals on the co-expressivity of speech and gesture inference-statistically.

- *Reorientation*: If the direction giver instructs the direction follower to change the direction, the accompanying gestures are characterized by a significantly higher proportion of redundant SFs (χ^2=227.998, df=2, p<0.001). ACTION and DIRECTION are the types of SFs found in these gestures. The number of complementary SFs is decreased in case of reorientations (χ^2=46.578, df=2, p<0.001). PATH, DIRECTION and RELATIVE POSITION are the kinds of SFs that are used complementarily.

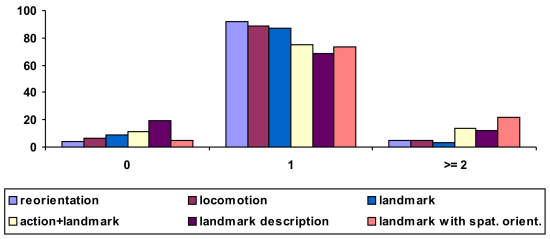

Figure 5: Frequency of gestures per communicative goal.

- *Locomotion*: Concerning the number of redundant SFs in speech-accompanying gestures locomotions are similar to reorientations. The number of redundant SFs in gestures accompanying utterances tagged as *Locomotion* is significantly higher than expected (χ^2=54.303, df=2, p<0.001). Again, ACTION and DIRECTION are found to be used most frequently. Regarding the influence of locomotions on the number of complementary SFs in speech-accompanying gestures, there is no significant relationship between those two variables (χ^2=2.029, df=2, p=0.363).

- *Action+Landmark*: Concerning the redundancy in gestures accompanying utterances of the kind *Action+Direction*, two or more redundant SFs occur more often than expected (χ^2=98.904, df=2, p<0.001). They are usually of the kinds ACTION and DIRECTION as in the case of locomotions and reorientations, but also of the kinds PATH and RELATIVE POSITION. The proportion of complementary SFs is increased in this category (χ^2=26.179, df=2, p<0.001). Especially one complementary SF is used relatively often.

- *Landmark*: In this category the frequency of redundant SFs is significantly decreased, (χ^2=106.632, df=2, p<0.001), while the number of complementary SFs is higher than expected (χ^2=46.423, df=2, p<0.001). RELATIVE POSITION and SHAPE are found to occur most frequently in the gestures when the

direction giver mentions a landmark.

- *Landmark description*: The same findings as for landmarks hold for landmark descriptions. The proportion of redundant SFs in accompanying gestures is lower than expected (χ^2=88.432, df=2, p=0.001) and the proportion of complementary SFs is higher than expected (χ^2=33.582, df=2, p<0.001). Moreover, the SFs RELATIVE POSITION and SHAPE are also the ones used most often.

- *Landmark with spatial orientation*: In the case of landmarks with spatial orientation there is a large number of gestures with RELATIVE POSITION as the only redundant SF (χ^2=110.852, df=2, p<0.001). Gestures with more than one redundant SF occur rarely. Concurrently, the frequency of complementary SFs is lower than expected (χ^2=79.427, df=2, p<0.001). If there are any complementary SFs they are of the kind RELATIVE POSITION or SHAPE.

To sum up, one may say that actions are described with speech-accompanying gestures that have more redundant SFs, while the proportion of redundant SFs is decreased when conveying information about landmarks. Concerning complementarity there are more such SFs than expected in gestures that belong to the categories *Landmark*, *Landmark description* and *Action+Landmark*. Less complementary SFs can be observed when referring to landmarks with spatial orientation.

Concerning the influence of the information status of the SFs, the only found correlation exists for the category ENTITY. The redundancy of gestures accompanying the introduction of entities is decreased, while utterances referring to evoked entities are accompanied by gestures with a higher proportion of redundant SFs than expected (χ^2=13.012, df=2, p=0.001). Moreover, the frequency of complementary SFs is slightly increased in case of new entities, while evoked entities are accompanied by gestures with less complementary SFs (χ^2=4.480, df=2, p=0.106).

4 Discussion

To our knowledge, this study is the first one analyzing the influence of dialog context and communicative goals on the distribution of information across speech and gesture. Our analysis of the direction giving dialogs reveals three major factors influencing the co-expressivity of speech and gesture, while others were found not to do so. First, problems concerning verbal encoding have an effect on the distribution of meaning, leading to more complementary and less redundant SFs in gestures. This goes together with the results of Bavelas et al. (2002) who found more non-redundant gestures when people had to describe pictures that were hard to encode. It seems as if people compensate for such verbal problems by adding complementary information to gestures. Second, the co-expressivity of speech and gesture is influenced by communicative goals. Instructions are accompanied by gestures with more redundant SFs, while gestures referring to landmarks are characterized by more complementary SFs. When giving directions, instructions are really important for the direction follower to find her/his way, especially reorientations and actions referring to landmarks. For this reason it would make sense to convey this information redundantly. However, Beattie and Shovelton (2006) recently found, that speakers tend to convey salient information gesturally. One could argue that at least in case of reorientations it would be difficult to have complementary SFs beyond ACTION and DIRECTION in gesture, but in fact we found gestures with SHAPE, PATH or RELATIVE POSITION as complementary SFs. Nevertheless, the number of complementary SFs is decreased significantly and information about actions, directions and sometimes paths is conveyed redundantly instead. Concerning the larger number of complementary SFs when referring to landmarks, one should think of the particular strengths and weaknesses of both modalities. Shapes and positions can often be easier visualized with hands and arms, than uttered verbally. In these cases the risk going along with complementary meaning in gesture, that is being overlooked by the dialog partner, is accepted. In the category *Landmark with spatial orientation* the localization is conveyed by speech, and in consequence there are less complementary SFs in the accompanying gestures. In the same sense one can interpret the found relationship between communicative goals and the use of gestures in route directions in general. Actions and landmarks have one accompanying gesture in the majority of cases. Descriptions of landmarks are not necessarily accompanied by any gesture.

Within landmark descriptions there may be contents, e.g. colors, that can not even be visualized. In contrast, gestures occur more often when entities are set in relation to one another, as in *Landmarks with spatial orientation*.

Third, the introduction of entities goes along with slightly reduced redundancy and increased complementarity in gesturing. So findings of Yan (2000) are supported tendentiously, but the influence can only be observed for entities, not for other kinds of SFs.

Finally, no influence could be found for feedback signals of the dialog partner, but there are at least two aspects relativising these results. First, the direction followers were not really unfamiliar with the campus. So their interposed questions do not reflect real understanding problems. In fact, the questions were of the kind ”what is the color of the building?” or ”how long does it take to get from here to there?”. Second, and even more important is the fact that only verbal signals of understanding have been annotated. Because of the video quality it was not possible to code nonverbal signals of feedback, although there is no doubt that such signals like head movements or facial mimics are equally good for signaling understanding or non-understanding.

References

James Allen and Mark Core. 1997. Draft of DAMSL: Dialogue Act Markup in Several Layers.

Janet Bavelas, Christine Kenwood, Trudy Johnson, and Bruce Philips. 2002. An Experimental Study of When and How Speakers Use Gestures to Communicate. *Gesture*, 2:1:1–17.

Geoffrey Beattie and Heather Shovelton. 2006. When size really matters. *Gesture*, 6:1:63–84.

Justine Cassell and Scott Prevost. 1996. Distribution of Semantic Features Across Speech and Gesture by Humans and Computers. In *Proceedings of the Workshop on Integration of Gesture in Language and Speech*.

Justine Cassell, Matthew Stone, and Hao Yan. 2000. Coordination and Context-dependence in the Generation of Embodied Conversation. In *First International Conference on Natural Language Generation*.

Michel Denis. 1997. The Description of Routes: A Cognitive Approach to the Production of Spatial Discourse. *Current Psychology of Cognition*, 16:409–458.

Jana M. Iverson and Susan Goldin-Meadow. 1998. Why People Gesture When They Speak. *Nature*, 396:228.

Ray Jackendoff. 1983. *Semantics and Cognition*. MIT Press: Cambridge, MA.

Michael Kipp. 2004. *Gesture Generation by Imitation - From Human Behavior to Computer Character Animation*. Boca Raton: Florida.

Sotaro Kita and Asli Özyürek. 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.

Stefan Kopp, Paul Tepper, and Justine Cassell. 2004. Towards Integrated Microplanning of Language and Iconic Gesture for Multimodal Output. In *Proceedings of the 6th International Conference on Multimodal Interfaces*, pages 97–104, New York, NY, USA. ACM Press.

Stefan Kopp. 2005. The Spatial Specificity of Iconic Gestures. In Klaus Opwis and Iris-Katharina Penner, editors, *Proceedings of KogWis05. The German Cognitive Science Conference*, pages 112–117. Basel: Schwabe.

George Lakoff. 1972. Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. In P.M. Perantean, J.N. Levi, and G.C. Phares, editors, *Papers from the 8th Regional Meeting: Chicago Linguistics Society*, pages 183–228.

David McNeill. 1992. *Hand and Mind - What Gestures Reveal about Thought*. University of Chicago Press: Chicago.

David McNeill. 2005. *Gesture and Thought*. University of Chicago Press: Chicago.

Alissa Melinger and Willem J.M. Levelt. 2004. Gesture and the Communicative Intention of the Speaker. *Gesture*, 4:119–141.

Emanuel A. Schegloff. 1984. On some gestures’ relation to talk. In J. M. Atkinson and J. Heritage, editors, *Structures of Social Action*, pages 266–298. Cambridge University Press.

Elisabeth Shriberg. 1999. Phonetic Consequences of Speech Disfluency. In *Proceedings of the International Congress of Phonetic Sciences*, volume 1, pages 619–622.

Timo Sowa. 2006. *Understanding Coverbal Iconic Gestures in Shape Descriptions*. Akademische Verlagsgesellschaft Aka: Berlin.

Hao Yan. 2000. Paired Speech and Gesture Generation in Embodied Conversational Agents. Master’s thesis, MIT, School of Architecture and Planning.