A micro-analytic investigation of how iconic gestures and speech represent core semantic features in talk

JUDITH HOLLER and GEOFFREY BEATTIE

Introduction

This study investigates the communicational role that iconic hand gestures (see McNeill 1985) play in everyday talk. When we talk about iconic gestures, we are referring to very special sorts of hand movements, which most people produce unconsciously, and which seem to convey meaning associated with the content of spontaneous speech. Iconic gestures must be distinguished from all other types of hand movements, like *beats* (or *batonics*, see Efron 1941, Ekman and Friesen 1969), which are quick, simple movements of the hand, which tend to follow the prosodic patterns of speech. Beats serve the purpose of stressing those parts of speech that a speaker considers to be important; thus beats differ from iconic gestures by being related to the pragmatic rather than the semantic content of speech. *Deictics*, on the other hand, are pointing gestures referring to either actually present or imagined objects or locations. In deictics, the shape of the hand does not, unlike with iconic gestures, convey semantic information itself, but together with the accompanying speech it becomes clear how a speaker is referring to his or her imagined or real surrounding. A fourth class of gestures, *emblems*, are gestures readily understandable in the absence of speech, as they are gestural signs with a strict, but culturally defined, verbal translation. For instance, the very common sign for 'excellence' in many Western European countries is represented by the index finger and the thumb forming a circular shape while the hand is swung away from the mouth accompanied by a kiss. Last of all, there are sign languages, which are different from iconic gestures, as they represent a whole gestural code, in which gestures substitute for speech and in which they are governed by conventional rules concerning meaning and grammar. In this paper, we are concerned solely with iconic gestures and their possible communicational significance in everyday conversation.

The theory of iconic gestures communicating semantic information

In 1985 McNeill proposed an important theory concerning the communicational function of iconic gestures and how these combine with the verbal mode of communication. He described speech and iconic gestures as being two forms of communication, which differ in the way that they communicate information as well as in the kinds of information they convey. The linguistic system is constrained by standardized rules of grammar as well as by the morphemes out of which it is constructed. This requires speech to be of a segmented-linear kind, as it consists of linear chains of individual words governed by syntactic structures thereby acquiring meaning. Iconic gestures, on the other hand, are global-synthetic 'in that the whole is not composed out of separately meaningful parts. Rather, the parts gain meaning because of the meaning of the whole' (McNeill 1992: 20). Moreover, they are noncombinatoric, as a certain meaning is not depicted in two separate gestures performed one after another, but rather the meaning is integrated into one single gesture. Thus, an event might be represented by one simple gesture, whereas several units of speech are necessary to describe the equivalent event.

To illustrate the difference between global-synthetic gestures and linear-segmented speech, McNeill (1992: 20) provided the following example (see appendix for transcription conventions):

'[and he's trying to run ahead of it]'

[hand moves forward at chin level while fingers wiggle]

The verbal information here consists of ten different morphemes, conveyed sequentially, whereas the gestural information consists of the hand performing two different movements simultaneously. As the meanings of the individual words have to be interpreted in the context of the whole sentence, the information of the part can change when the utterance is finally complete, whereas the gestural meaning is conveyed immediately by the wiggling movement, which can be interpreted as a running movement at the same time that the moving hand is seen as representing the running character, with the fingers here representing the moving legs. The main feature of the gesture is that both facets of communication are performed at the same time.

Comparing the kinds of information transmitted by iconic gestures and speech, McNeill suggested that iconic gestures can contribute information, which is congruent with the information contained in speech in some particular cases, whereas in other cases the information in the gesture can be seen as additional to the information conveyed by the speech. The

following two examples (extracted from McNeill 1992: 106) may help to clarify this difference:

'he tries going [up the inside] of the drainpipe'

[hand rises up with the index finger extended, depicting the character rising up and possibly the interiority of the pipe]

Here, the iconic gesture depicts mainly 'upward movement', which may seem redundant, as the gesture looks as if it only conveys information also present in speech. In discussing this example, McNeill said that, 'the gestural and the linguistic channels exhibit(ed) the same meaning' (McNeill 1985: 352). The next example shows a slightly different case (extracted from McNeill 1992: 14):

'and she [chases him out again]'

[hand appears to swing an object through the air]

This example shows the case of an iconic gesture contributing additional information, which is clearly not included in the accompanying utterance. The verbal information describes the mode of action and the recurrent nature of the action, but not the means, i.e., the reason one character is running away from the other is because it is being chased by a character wielding an umbrella. The linguistic description alone allows for considerable opportunity for a listener's personal associations and interpretation to be added, as it leaves a number of details of the action undefined. Iconic gestures in general display the imagistic side of scenes or objects, and thus the iconic gesture adds here some of the specifying details that the speech itself does not provide. Since the gesture provides not only information about the fact that an instrument is involved (umbrella), but also about how it is being used (swung from left to right), the iconic gesture illustrates in some detail *how* the action is actually being accomplished.

   Thus, comparing the semantic information transmitted by the gestures in the first two examples above, it can be seen that the contribution of iconic gestures to communication over and above the actual speech itself can vary considerably. McNeill concluded from this and related observations that iconic gestures and the verbal utterance that they accompany may be partially overlapping in terms of the semantic information they convey, but that when the gestures exhibit information not available in speech, then they are not merely co-expressive but rather they are complementary. Hence, McNeill argued that iconic gestures can provide additional insight into speakers' thoughts and their specific

cognitive representations of the original scene, and that sometimes an interlocutor is able to get the full information about a speaker's mental representation only by receiving information from both the iconic gesture and the speech simultaneously.

McNeill (1985) argued that the gestural and the linguistic system must be closely intertwined in that the cognitive processes underlying the generation of iconic gestures and speech must be sufficiently close that they share a computational stage. He drew attention to the fact that speech and iconic gestures seem to show parallel patterns of developmental processes in young children, since the iconic gestures generated by children who, in terms of Piaget's theoretical framework, have not yet reached certain 'cognitive stages' seem to mismatch with the accompanying verbal utterances. Furthermore, McNeill maintained that neurological damage seems to have parallel effects on the generation of gestures and speech. For example, aphasics of the Wernicke type, who have difficulties in combining single utterances into a semantically coherent whole, seem to have difficulty in generating iconic gestures, whereas aphasics suffering from the Broca type show a marked disturbance in producing speech with a correct grammatical structure and they seem to display a disturbance in the generation of beats. Beats, as discussed earlier, are hand movements with no semantic content but a close linkage to grammatical structure, and that serve a variety of pragmatic functions (for a more detailed discussion, see McNeill 1985: 353–366). McNeill also maintained that the close relationship between gestures and speech is characterized by a number of other general characteristics. He argued that gestures occur exclusively together with speech, and that cases in which they do not are very rare. Furthermore, McNeill stated that iconic gestures and speech are temporally and semantically synchronous in the sense that iconic gestures occur simultaneously with those parts of speech that refer to the same semantic idea as the gesture itself does, and that 'iconic … gestures almost never cross clause boundaries' (McNeill 1985: 360-361).

In summary, McNeill has developed the theoretical idea that iconic gesture and speech form a unitary, integrated system such that an interlocutor must receive information from both sources in order to obtain the full picture that any given speaker has in mind. However, he has also argued that iconic gestures and speech differ in the way that they represent semantic information as well as in the kinds of information that they provide. McNeill's theory provides a unique and highly valuable framework for the analysis of the communicational role of iconic gestures, but his theory fails to answer the question to what extent the information represented by both modes of communication

is generally redundant with respect to each other or not. McNeill's analyses also have not systematically considered how the representation of semantic information is split between the gestural and the linguistic channels.

There is perhaps one important point to make at this juncture. As the basis for his analyses, and throughout all of his subsequent theoretical arguments, McNeill uses clausal units as the fundamental speech units, clauses that are extracted from more complex segments of speech, but which are those parts of these complex segments of speech that are themselves accompanied by an iconic gesture. However, one might say that there is something a little curious about this. McNeill does, after all, stress that speech is linearly segmented in nature with a number of sequential dependencies, and that speech, unlike iconic gesture, takes time to unfold in talk in order to get its meaning across. Thus, an exclusive focus on the immediate syntactic clause might seem a little premature here, because critical semantic information is likely to be communicated in neighboring clauses. We will return to this important point subsequently.

Experimental investigations testing the communicational role of iconic gestures

Beattie and Shovelton conducted several studies in which they investigated empirically the communicational effect of iconic gestures by testing the amount of information respondents gleaned from iconic gestures (Beattie and Shovelton 1999a, 1999b, 2002). In all of these studies, they used questionnaires or interviews with decoders as the primary method of measuring the amount of information respondents received from iconic gestures. This represented a rather innovative approach to research on gestures, as previous researchers had focussed their analyses on how speech and gestures are generated, but had not tested the possible communicative effects on decoders. The decoders in the Beattie and Shovelton studies were interviewed after either having seen videoclips of single iconic gestures, which were extracted from cartoon narratives from other participants, accompanied by the corresponding clausal units of speech (video condition) or after only having heard the speech of the corresponding clausal units without seeing the gesture (audio-only condition). In two studies (1999b, 2002), Beattie and Shovelton introduced a third condition, in which the decoders saw the gestures from the video-clips played without the sound (vision-only condition). The decoder interviews either consisted of predetermined

questions referring to individual semantic categories, such as 'objects identified', 'action of objects', 'shape', 'size', 'movement', 'direction of movement/speed', and 'relative position' (1999a, 1999b, study 1 of Beattie and Shovelton 2002), or they involved the more open-ended method of simply asking the decoders to write down *any* information they had received from the gesture or the speech or from both simultaneously (study 2 of Beattie and Shovelton 2002). Also, in one study, the respondents had to answer the questions in a yes/no format (1999a); in others they had to write down the answers in a free format (1999b, 2002).

All of the above studies showed consistently that there was an increased overall accuracy in the answers when the decoders saw the iconic gestures in addition to hearing the speech (video) compared to only hearing the speech (audio-only). Moreover, the iconic gestures displayed on their own (vision-only) seemed to have communicated a considerable amount of information. However, with regard to the different semantic categories, the amount of information transmitted by iconic gestures on their own was statistically significant only with regard to the category 'relative position' in all the studies (1999a, 1999b, 2002), as well as 'relative size' in some but not all the studies (1999a, 1999b, study 1 of Beattie and Shovelton 2002).

In addition to investigating the communication of semantic information by iconic gestures, Beattie and Shovelton (2002) also considered another important aspect of McNeill's theory (1992), namely the concept of the viewpoint of the gesture. McNeill (1992) had differentiated between iconic gestures, which express a 'character viewpoint' (C-VPT) from those which express an 'observer viewpoint' (O-VPT). A C-VPT gesture 'incorporates the speaker's body into the gesture space, and the speaker's hands represent the hands (paws, etc.) of the character' (McNeill 1992: 119), whereas an O-VPT gesture 'excludes the speaker's body from the gesture space, and his hands play the part of the character as a whole' (McNeill 1992: 119). With reference to the examples above, this means that the iconic gesture accompanying the verbal utterance 'and he's trying to run ahead of it' would be classified as an O-VPT gesture, since here the narrator's hand represents the running character and his fingers the character's legs. The iconic gesture accompanying the clause 'and she chases him out again', on the other hand, would be classified as a C-VPT gesture, because the narrator imitates the character swinging the umbrella. McNeill had attempted to connect these different types of iconic gestures to the narrative distance that they express; he had argued that the different viewpoints in gestures serve as a narrative strategy in the sense that a speaker can alter the narrative distance by being more or less

'remote from the fictive world' (McNeill 1992: 191) with C-VPT gestures diminishing the narrative distance and O-VPT gestures increasing it. Beattie and Shovelton (2002), however, suggested that the different gestural viewpoints might pragmatically be even more important than this in that they might not only function as regulators of narrative distance, but they might also differ in the amount and perhaps even the type of semantic information they communicate. They investigated this by categorizing iconic gestures in their decoder studies according to McNeill's definition of gestural viewpoints and found that, although O-VPT gestures did communicate semantic information to some extent, C-VPT gestures were significantly better than O-VPT gestures at conveying semantic information overall, and in particular at conveying positional information (36.4 mean percentage accuracy for C-VPT gestures for positional information versus 5.3 mean percentage accuracy for O-VPT gestures). Beattie and Shovelton therefore concluded that gesture viewpoint directly mediates the communicative power of the iconic gesture, and that C-VPT gestures seem to be of higher communicative power than O-VPT gestures.

In summary, the results of Beattie and Shovelton's experimental studies suggested, in line with McNeill's theory, that iconic gestures do communicate a considerable amount of semantic information in talk. Moreover, Beattie and Shovelton argued that the results demonstrate more than this, namely that iconic gestures reveal critical information that is poorly transmitted in the speech, particularly about properties of the objects depicted. For example, they found that gestures revealed information about the size of an object, when the hand has been used pantomimically to show how a small object is being held. They also found that decoders picked up information about relative position, for example about the position of the nose relative to the hand when a gesture demonstrated how the nose is being squeezed, and that character viewpoint gestures communicated this type of semantic information especially well. Although these semantic aspects, relative size and relative position, were the only ones that decoders received a statistically significant amount of information about, Beattie and Shovelton also found that information about other semantic aspects was conveyed to a certain extent by the iconic gestures displayed on their own. This information concerned the number of objects involved, the identity of objects, as well as about the speed and movement of the action. Thus, they concluded, 'that McNeill (1985) had seriously underestimated the amount and the nature of information conveyed by these complex and often elaborate iconic gestures that accompany speech' (1999b: 455).

This experimental approach of Beattie and Shovelton was based very closely on McNeill's theoretical framework, and their studies tended to focus on the clausal unit, which the gesture was seen to accompany. In fact, only in their first study (1999a) did they include some extracts of speech that were longer than individual clauses (see, for example, Beattie and Shovelton 1999a: 15). Beattie and Shovelton viewed this, perhaps rightly, as a methodological shortcoming in studies of this kind, and they argued that 'varying sizes of extracts could possibly explain why some semantic categories were significant rather than others' (Beattie and Shovelton 1999b: 444). As a consequence, they used exclusively clausal units in all of their subsequent studies (1999b, 2002). Thus, although there seems to be strong evidence that iconic gestures do communicate semantic information in the absence of speech, the question remains to what extent the findings would have been different from this, if the studies had used larger linguistic units specifically designed to carry the full range of semantic information in a scene or event description.

The present study

In order to introduce the core considerations of the present study, a number of important theoretical and methodological issues need to be addressed. The first issue is the social context from which we extract our linguistic and our gestural material. McNeill (1985), and subsequently Beattie and Shovelton in their experimental studies, have used participants narrating cartoon stories. However, the social context employed here has always been of a particular kind, notably a narrator telling a story in the presence of a confederate, who quite explicitly does not engage in conversational interaction with the narrator. A study by Beattie and Aboudan (1994) showed that the frequencies of iconic gestures changed quite dramatically in different social contexts. Among others, they compared the frequencies of iconic gestures generated in a monologue condition to those generated in a dialogue condition, and found that there was a significant increase in gesture production in association with the dialogue condition, i.e., where the confederate interacted with the narrator of the story. Thus, one could present the argument that any analysis of gestures, which is not partly based on conversational material, is simply too restrictive, given that conversation is the natural forum for everyday talk, and in addition given that this is the social context in which iconic gestures are the most common. Thus, in the present study we use conversational material as our primary focus.

Second, although the findings by Beattie and Shovelton (1999a, 1999b, 2002) can be seen as probably the most striking evidence supporting McNeill's communication theory (1985), it has to be noted that, as Beattie and Shovelton themselves stated, there is 'as yet, no theoretical explanation as to why only these particular semantic features [relative position and relative size] should be significant' (Beattie and Shovelton 1999a: 28). In the present study, we aim to tackle this question directly. One hypothesis might simply be that the gestural and the linguistic systems are indeed closely interlinked and that the representation of semantic information is partitioned between both of these communicational channels, such that the semantic features 'relative position' and 'relative size' are represented gesturally very well but not at all well linguistically. We consider this hypothesis in the present study by analyzing in detail how the two semantic features that emerged in these studies as the most significant in terms of gestural communication are partitioned between the linguistic and the gestural channels. In pursuit of this question, we apply a more micro-analytic approach to our data than previously used, a micro-analytic approach that involves analyzing semantic information in a very detailed manner by using categories that are sensitive to capture important semantic differences. The categories 'relative position' and 'relative size', as applied by Beattie and Shovelton, seem rather broad in the sense that they may mask important differences between certain semantic aspects. For example, Beattie and Shovelton (1999a, 1999b, 2002) asked respondents if they received any information about 'the position of object(s) relative to anything else'. It could be argued, however, that the category 'relative position' as such incorporates a range of quite different semantic aspects that may all relate to some kind of positional information, but in crucially different ways. For example, if we ask someone to reconstruct a scene that he or she will be told about, which includes spatial features, this person may need information about, for example, the position of entity 'a' relative to entity 'b', the position of entity 'b' relative to entity 'c', as well as about the position of entity 'a' relative to entity 'c' to be able to do so. Since a question asking simply for information about 'relative position' is not likely to be answered by decoders in such an explicit way that they will state in any real detail which potential aspects of this category they received information about, we decided that such aspects should be considered as different semantic sub-categories. Therefore, we aim to distinguish these different semantic aspects of the same basic category, in the first instance, by splitting the category 'relative position' into four sub-categories, as well as by distinguishing size information according to which kinds of entities it referred to (see Table 1).

Table 1.   *Overview of the semantic categories applied in the present study*

Relative position of
1.   Agent-Object
2.   Agent-Instrument
3.   Object-Instrument
4.   Object-Surrounding space

Relative size of
5.   Agent
6.   Object
7.   Instrument

Third, we use a fundamentally different approach to that employed by Beattie and Shovelton. In this study, we analyze the data from a traditional linguistic perspective in the sense that we focus on the semantic information that is *encoded* in the gestural and the linguistic channels of communication. Of course, it could be argued that if one is interested in the communicational role of iconic gestures, then decoder studies, like those of Beattie and Shovelton, are the most appropriate method for determining what kind of semantic information is communicated by iconic gestures, but on the other hand, decoder studies represent a method that clearly has a number of inherent limitations and disadvantages. For example, one has to use relatively broad semantic categories, since otherwise the tasks would be too time consuming and thus might exceed the decoders' attention capacity or it might lead to a decrease of motivation on the part of the participants. Thus, for example, when Beattie and Shovelton questioned their decoders about relative position ('What is the position of the object(s) relative to anything else?'), they intended to tap with this single question information about the position of objects relative to each other, about the location of an action, about the orientation of objects, and about contact (see Beattie and Shovelton 1999b: 447). The consequence might here be that it is relatively difficult to differentiate which of these semantic aspects the decoders actually received information about, since decoders only gave one general answer to this question. Nevertheless, Beattie and Shovelton's studies revealed highly interesting findings concerning the communicational function of iconic gestures, since they hint at those semantic aspects that seem to be significantly well communicated by iconic gestures. Thus, Beattie and Shovelton's findings represent the basis for our analysis. More precisely, we chose to tie these earlier findings in with an investigation that uses a somewhat different methodological approach, and which will thus

provide a different analytic view of the same issue. In other words, we hope to throw light on the communicational role that iconic gestures play in talk from a somewhat different angle to see if the existing view of this complex issue is perhaps too narrow.

Fourth, this study attempts to recognize explicitly the fact that the linguistic representation of semantic information happens sequentially over successive clauses, whereas imagistic gestural representations are more or less instantaneous, co-occurring with only a small part of speech, namely the clausal unit. The fact that McNeill's and almost all of Beattie and Shovelton's analyses (an exception being their 1999a study, see above) were based on clausal units rather than more complex linguistic segments might therefore have had considerable implications for their conclusions concerning the amount and kinds of semantic information that iconic gestures and speech represent. This is because they may have ignored important semantic information that might have been repre- sented *elsewhere* in the verbal utterance rather than in the immediate clausal unit. As we have already discussed, Beattie and Shovelton (1999b) did explicitly wonder whether the size of the extracts of speech might potentially have had an impact on the communicative effectiveness of gestures, but size of extract was not *systematically* manipulated in any of their studies. In the present study, we use semantic and specifically ideational rather than clausal units as the basis for our analysis, and it will be interesting to determine if this particular theoretical decision will have an effect on the pattern of our results.

Our last point is that Beattie and Shovelton found that a crucial mediator for the communicative power of iconic gestures was the view- point of the gesture. However, the only explanation, which they offer to explain their observation that C-VPT gestures were significantly better at conveying positional information than O-VPT gestures is that the majority of their C-VPT gestures were direct reenactments of the action that the narrators discussed. We attempt here to gain more insight into how the communication of semantic information is influenced by differences in gestural viewpoint by offering a micro-analytic analysis of how C-VPT gestures and O-VPT gestures represent both position and size information.

Method

Participants

Thirty-eight students of the University of Manchester took part in the experiment.

Stimuli

The stimulus material consisted of two 'Tom and Jerry' cartoon stories (issue 30, 2000). Both cartoon stories contained minimal linguistic information (three speech bubbles covering two short sentences, some single verbs and various onomatopoeic expressions [e.g. 'grrr' or 'ooomph']). The stimulus material, both cartoons 1 and 2, represented just the last two pages of longer stories ('Park life', pp. 22–29; 'Unwanted visitors', pp. 31-35). Even though the cartoon extracts did not comprise the entire story, these still allowed participants to develop a sense of the overall narrative, and no participant appeared to display any difficulty in telling the stories, even though they started in effect at the midpoint of the story. 'Tom and Jerry' cartoons typically consist of a chain of events involving two (or sometimes three) characters interacting, and in narrating that part of the story provided to them, the narrator is likely to focus on single events and a number of directly successive actions.

Procedure

The 38 participants took part in the study in individual trials. Each participant was led into an observation room by one of the experimenters (JH) and was asked to sit down in an armchair. The experimenter sat down on a chair in front of the participant slightly to his or her right so that the experimenter could reach an overhead projector placed to the right of the participant's seat. The cartoon stories were projected onto the wall in front of the participant. Due to the arrangement of seats, the experimenter did not face the wall that the cartoon stories were projected on; this was intended to create a situation as similar as possible to a face-to-face conversation. The participants were video-recorded with an unobtrusive video camera and informed about the presence of the camera before the experiment began in line with the ethical guidelines of the British Psychological Society (BPS), who state that video recordings of participants in research can only be made 'with the expressed agreement of those being recorded both to the recording being made and to the subsequent conditions of access to it' (2000: para. 4.5). They were not told that the specific focus of the research was iconic gestures, but that it was to do with communication generally. Then, the experimenter asked the participants to narrate the cartoon stories in as much detail as possible, imagining the experimenter to be a conversational partner, who had never seen the pictures beforehand and did not know anything about the stories. During the narration, the experimenter

interacted with the respondent by asking questions about the story, referring to the kinds of actions depicted and the ways in which they were being carried out, and there were also questions asked about the cartoon characters' feelings and motives. This was designed to produce a situation more like everyday conversational talk. Experimental evidence suggests that iconic gestures are much more common in more conversational situations (see Beattie and Aboudan 1994) than in those situations employed more frequently in this type of research, which normally feature a confederate who does not interact with the participants (see, for example, Beattie and Shovelton 1999a and 1999b). After narrating the cartoon stories, the participants were thanked and fully debriefed and were offered the opportunity to not have their recorded material analysed.

Analysis

Six 'semantic events', which different narrators frequently referred to linguistically, and which were also accompanied by iconic gestures, were chosen to form the basis of the present analysis.

Event 1:   Jerry the mouse is holding a spike, with which he is about to stab Tom the cat, who is standing in front of him while the mouse faces his back.

Event 2:   This event refers to the past action of this stabbing with the focus on the recipient, i.e., Tom having been stabbed by Jerry (passive syntax/past tense).

Event 3:   Tom is being pinned to the ground by a spike with which he has been stabbed, and which is now stuck in his tail.

Event 4:   A park keeper has lifted Tom up to carry him away, grabbing him with his hand by the scruff of the neck.

Event 5:   A dog is holding Tom in front of him in the air by grabbing him around the throat, strangling him.

Event 6:   The dog has lifted Jerry up by the tail and is dangling the mouse with the extended arm in front of his body.

Of these six semantic events, four events were extracted from narrations of story 1 and two from narrations of story 2. Events 1–3 were represented in consecutive pictures from story 1 and referred to an unfolding time series of events involving Tom, Jerry, and the spike. Event 4 was also extracted from cartoon 1, but was not immediately adjacent to the events already described. Events 5 and 6 were represented in nonconsecutive pictures from cartoon 2.

The *gestural data* analyzed in the present study consisted of 58 iconic gestures generated by 27 different participants. The gestures all referred to the same six semantic events from both cartoons. Considering the number of gestures analyzed compared to the number of participants who generated these gestures, it is obvious that sometimes iconic gestures that stemmed from the same participant were included in the analysis; however, the iconic gestures that referred to the individual semantic events were all generated by different participants. Also, if a narrator referred to the same semantic event more than once and the speech was accompanied by an iconic gesture, only the first occurrence of the gesture was analyzed.

The *segments of speech* analyzed were only those showing close temporal relationship to the corresponding iconic gesture, i.e., the semantic unit (operationalized as a sentence or as a small group of highly related sentences related to the same semantic idea; see Butterworth 1975) containing the lexical affiliate. The term 'lexical affiliate' refers to the lexical component that relates to the iconic gesture 'either semantically (i.e., the shape that the gesture describes depicts a/the "meaning" or referent of a word) or by word class (e.g., "locatives")' (Schegloff 1984: 275). The reason for the focus on sentential and semantic units rather than clausal units (cf. McNeill 1992, Beattie and Shovelton 1999b, 2002) is because we are concerned here with the relative contribution of speech and gesture to the communication of semantic meaning, and because ideational units (Butterworth 1975) are primarily semantic rather than syntactic units of information. Those studies that have focused exclusively on clausal units may have systematically underestimated the role of verbal information generally in semantic communication. However, it must be noted that verbal information not directly contained in the sentence accompanied by the iconic gesture is only considered when the information occurred in the sentence directly preceding or following the one containing the lexical affiliate. In addition, the sentence must not be accompanied by a new iconic gesture, and furthermore the successive sentences must be describing the same underlying semantic idea. The example below meets these criteria:

'his tail's been pinned to the ground by this spike, **so he's obviously not been stabbed in the bottom, it was in the tail, by the spike, wasn't in the bottom.** *He's been stabbed in the tail, which is [now sort of fixed into the ground]'.*

[the hand is held at the height of the lap in the middle of the narrator's gesture space, the fingers are curled in, only the index

finger is extended and the hand moves up and down several
times on a small scale with the tip of the index finger ahead, so
that it looks as if the index finger would poke into something]

Here the two sentences clearly refer to the same underlying semantic idea,
and the parts of the utterance printed in italic directly describe the
respective semantic event and these consist of very similar information.
The parts of speech not printed in italic represent a self-correction of a
previously uttered sentence referring to a different picture, and therefore
this information cannot be seen as relevant for the scoring (as it logically
would not make sense to score verbal information, which the speaker
himself in this moment judges as invalid information). Thus, the first part
and the second part of the utterance describing the event are almost equal
concerning their morphology, apart from the fact that the first part
includes additional information about the instrument (spike) and a verb
expressing slightly different information ('to pin down' includes
information about the direction of a movement, whereas 'to be fixed to
something' is a stative verb). As it is likely that, if the first part of the
verbal information would not have been uttered, the speaker might have
specified the instrument in the second part of the utterance instead, and
might also have chosen a different verb, it seems plausible to consider
these two aspects of information when scoring the verbal information.

The following example represents a somewhat different case:

> **'Jerry's being held by his tail by the dog … ehm … I can't tell
> where he's holding him … ehm … he's holding him over
> something, but I can't tell what it is. And the dog is laughing
> looking at Tom** *[while he's dangling Jerry]*. **And Jerry's
> looking very scared now'.**

> [right hand is held in front of the narrator, fingertips are
> pointing to the ground, the tip of the index finger is touching the
> thumb imitating a pincer-like grip]

Here, compared to the part of speech printed in italic (the part that has
been considered for analysis), the first sentence contains additional
information about the part of the body from which the mouse is being
dangled (tail). Yet, this information has not been taken into account, as
the two sentences are clearly separated by two other important aspects,
which the speaker clearly focuses on, namely the exact position of Jerry in
the surrounding space and the relationship between the dog and Tom.
After that, the speaker picks up again the former description of the

respective action (the dog dangling — or holding — Jerry). Thus, in comparison to the preceding example, it seems that here both parts of speech referring to the same 'semantic event' are separated by units of thought describing clearly different semantic aspects. Therefore, the verbal information of the first sentences has not been considered in our analysis.

The examples above show clearly that a semantic idea can be described linguistically extending over significantly larger and more complex segments of speech than just isolated clausal units.

Scoring the information

Instead of one broad semantic category that summarizes all different kinds of positional information, the semantic category 'relative position' has been split up into four subcategories: 1) 'agent-object', 2) 'agent-instrument', 3) 'object-instrument' and 4) 'object-surrounding space'. Concerning 'relative size', we differentiated between information about 5) the size of an agent, 6) the size of an object, and 7) the size of an instrument. In the course of our analyses, each iconic gesture and each segment of speech accompanying a gesture was considered, if it represented information referring to any of the semantic categories. Continuing on from this, the descriptive data was quantified by scoring the semantic information using a binary code. This means that if a gesture or a segment of speech provided clear information about one of the types of positional information or about the size of an entity, the respective semantic category was scored with 1, whereas a score of 0 was given for the cases in which gesturally or linguistically, from an analyst's point of view, there was no information provided concerning any of these semantic categories (see Table 2).

However, in some cases, an iconic gesture or a segment of speech was relatively ambiguous or less clear concerning the representation of

Table 2. *Hypothetical example of the gestural and the linguistic data generated by an individual narrator and transformed into a binary code*

	Information about	Gesture	Speech
Relative position	Agent-Object	1	0
	Agent-Instrument	1	0
	Object-Instrument	0	1
	Object-Space	0	1
Relative size	Agent	0	0
	Object	0	1
	Instrument	0	0

semantic information. This concerned principally the semantic categories 'relative position: agent-object', 'relative position: object-space' and 'relative size: agent'. For example, the majority of narrators referring gesturally to event 5 represented 'full' information about the relative position of the agent (the dog) and the object (Tom) relative to each other by holding the arm up in front of them and by extending it at the same time. However, in one case a narrator held the arm in front of the body, but close to the chest instead of extending it. Thus, the gestural information about how agent and object are positioned relative to each other, as suggested by the cartoon, was partly there, but clearly in a less 'full' form compared to those narrators who showed that Jerry is being held in front of the agent *and* at a distance. A number of different linguistic representations referring to this event were also considered, since there were some verbal utterances that provided 'full' information about the relative position of agent and object by mentioning that 'the dog is extending his arm *and* holding Tom up in the air', whereas some narrators did not provide in a linguistic form the information that the dog extends his arm. Referring to the category 'relative position: object space', some speakers described event 3 linguistically, for example, by saying that 'Tom's tail is pinned *down*', whereas other narrators referred to the same event by saying that 'Tom's tail is pinned *to the ground*'. The latter verbal description is clearly more precise, since it provides the information that Tom must be on the ground, whereas the first verbal description leaves more opportunity for interpretation concerning the kind of surface which Tom is pinned down to, where this surface is within the surrounding space, and thus where Tom is positioned relative to the space that surrounds him (he might, for example, be sitting on a table, instead of on the ground). Concerning the size of the agent, there were three cases in which this semantic aspect was verbally less clearly defined compared to verbal utterances generated by other narrators. The majority of narrators referred to the agent of event 4 as 'a man', 'a park keeper', 'a park ranger' etc., which provides the information that the agent is of the size of a relatively grown-up human, whereas other narrators referred to the agent merely as 'a hand', 'a human', or 'a person', and thus did not define whether the agent was a child or a grown-up. Thus, they were less clear about the size of the agent. Such cases, in which there was unclear information provided, and which thus have been considered to leave considerable opportunity for interpretation, have been scored with 0.5, but have not been included in the present analysis.

The inter-observer reliabilities have been tested for each semantic category for both gesture and speech. The percentage agreements ranged from 83.3 for 'relative position: agent-object' (gesture) and

'relative position: object-space' (speech) to 100 for 'relative position: agent-instrument' (gesture), 'relative size: agent' (gesture), 'relative position: agent-object' (speech), 'relative position: agent-instrument' (speech), 'relative position: object-instrument' (speech), 'relative size: agent' (speech), and 'relative size: object' (speech). Cohen's Kappa was calculated for the scoring of information in the gesture and speech codes by summing the data points from the different semantic categories, which resulted in K = .83 for both gesture and speech (the calculation of individual Kappa values was not strictly possible for more specific semantic categories due to the small number of data points involved). In the following analysis, these scoring values 0 or 1 have each been summarized across all iconic gestures, as well as across all segments of speech (see Tables 3, 4, and 5 of the results section), for each semantic category. Thus, we have generated a large data set consisting of binary coded semantic information about position and size, which represented the basis for testing statistically how the two channels of communication (speech and gesture) represent these semantic features in talk.

Results

For the statistical analysis in this study, we applied Chi-square tests where appropriate and the G-statistic, which is the log-likelihood ratio goodness-of-fit test, and which operates on the logarithmic transformation of the underlying frequencies, when the assumptions of the Chi-square test were violated, because of small N's.

Table 3. *Observed frequencies for those examples scoring 0 or 1 for the semantic categories 'relative position: agent-object' and 'relative position: agent-instrument'*

Relative position: A-O

	Informational value	
	0	**1**
Gesture	12	42
Speech	57	0

Relative position: A-I

	Informational value	
	0	**1**
Gesture	5	22
Speech	27	0

Table 4. *Observed frequencies for those examples scoring 0 or 1 for the semantic categories 'relative position: object-instrument' and 'relative position: object-space'*

Relative position: O-I

	Informational value	
	0	**1**
Gesture	26	1
Speech	13	14

Relative position: O-S

	Informational value	
	0	**1**
Gesture	58	0
Speech	46	6

Table 5. *Observed frequencies for those examples scoring 0 or 1 for the semantic categories 'relative size' -agent, -object, -instrument*

Relative size: agent

	Informational value	
	0	**1**
Gesture	58	0
Speech	23	32

Relative size: object

	Informational value	
	0	**1**
Gesture	58	0
Speech	0	58

Relative size: instrument

	Informational value	
	0	**1**
Gesture	26	1
Speech	10	17

Information about relative position

The first semantic categories to be considered here are 'relative position: agent-object' and 'relative position: agent-instrument'. Table 3 provides an overview of the relative frequency of examples scoring 0 or 1 for just two of the four categories of 'relative position', namely 'agent-object' (A-O) and 'agent-instrument' (A-I). The figures show that information about the relative position of agent and object, as well as about agent and instrument, was throughout our sample encoded solely by iconic gestures, and not at all be speech. The statistical analysis reveals that iconic gestures are significantly better than speech at conveying both of these categories of semantic information ([A-O] $\chi^2 = 71.54$, d.f. = 1, p < .001; [A-I] $\chi^2 = 37.13$, p < .001).

The following examples, stemming from different narrators and from a variety of the six 'semantic events', illustrate how semantic information about relative position: 'A-O' and about relative position: 'A-I' was represented in the two respective modes of communication. These examples can be seen as representative for the majority of gesture-speech compounds analyzed in this study.

Example 1 (A-O and A-I):

'the mouse is holding some pointy thing, and he's [gonna sort of spike him]'

[both hands rise in front of the narrator to about stomach height, they clench into fists, the left hand is held behind the right hand and they are pushing forward (2x)]

Considering only the linguistic utterance, there is no information provided about either the relative position of agent and object, or about the relative position of agent and instrument. The mouse could be about to spike an object that is positioned in front of him as well as above him, behind him, or at almost any other position. The mouse could likewise be holding the instrument at almost any imaginable position with reference to his body. On the contrary, if we consider only the gesture, which demonstrates someone holding a thin object with two hands pushing frontally forward, the information is provided that the object, which the agent is pushing into with the instrument, is positioned right in front of the mouse. Furthermore, the gesture shows that the mouse is holding the instrument in front of the body at about stomach height.

Example 2 (A-O):

'Bulldog's got him sort of [… lifted him up in the air]'

[left hand rises in front of the narrator to the height of the face,
the arm is extended, the fingers are slightly curled in and the
thumb is extended, but they do not touch each other so that
there is a round space created within the hand]

In this example, the narrator transfers himself into the role of the dog by
demonstrating that the dog is standing there extending his arm frontally,
with which he is holding Tom the cat up in the air. As in the example
above, the speech on its own leaves undefined where the object is
positioned in relation to the agent. An interlocutor, who would receive
only the verbal information, could only guess the relative position of
agent and object, since, for example, the dog might be holding the cat up
in front of him, behind him, at the right side of his body, etc. However, by
imitating gesturally the dog's posture and by illustrating the dog's grip
around the cat's neck, clear positional information concerning agent and
object is provided by the iconic gesture. The gesture on its own shows that
the cat is positioned right in front of the dog, and with reference to the
dog's body that the cat's face is being held at about the same height as the
dog's face.
   Example 3 (A-O):

'Jerry's being picked up by the dog, by [Jerry's tail]'

[left hand rises in front of the narrator to about chest height,
the back of the hand is up and the fingers are performing a
pincer-like grip, index finger and thumb are touching each other]

Here, again, we see that the utterance does not specify where the object is
positioned relative to the agent, since the dog could in practical terms hold
his arm in an almost infinite range of different directions, which would
change the position of the dog and Jerry relative to each other. Only the
gestural representation of the 'semantic event' provides the information
that the dog is holding the mouse in front of his body at about the height
of his chest.
   However, when we focus on the other two subcategories of 'relative
position', which refer to the relative position of object and instrument
(O-I) and of the object in its surrounding space (O-S), we find that a very
different pattern emerges here compared to the findings with the
categories 'relative position: A-O' and 'relative position: A-I'.

Table 4 provides an overview of the relative frequency of examples scoring 0 or 1 for the other two categories of 'relative position', namely 'object-instrument' (O-I) and 'object-space' (O-S), which shows that information about 'relative position: O-S' was sometimes represented linguistically, but never gesturally. This is also the case for information about the relative position of object and instrument, apart from one single exception, which we will refer to subsequently. And indeed, the statistical analysis shows that speech is significantly better than iconic gestures at conveying both of these categories of semantic information ([O-I] $\chi^2 = 15.6$, d.f. = 1, p < .001; [O-S] $G_{adj} = 187.71$, d.f. = 1, p < .001).

Examples from the present database are used again to demonstrate how different narrators represented information about the relative position of instrument and object as well as of an object in its surrounding space in gesture and speech:

Example 4 (O-I):

'[this spike is stuck in Tom's tail]'

[right hand rises in front of the narrator to about stomach height, clenches into a fist and moves down forcefully to about the height of the seat, then it is stopped abruptly]

In this example, the utterance provides information about the relative position of object and instrument relative to each other. A listener receiving only the verbal information would know where the instrument is placed relative to the object's body, i.e., on the tail. On the other hand, the iconic gesture here represents no information about this semantic aspect, since it shows that the agent is pushing the instrument towards an object, but not where it hits the object (Tom).

Example 5 (O-S):

'the dog's picked [him off and he's just holding him] so that his feet don't touch the ground'

[right hand rises in front of the narrator's body to about the height of the shoulders, the arm is fully extended, the hand is clenched into a fist]

The speech from example 5 provides the information that the object is off the ground, and thus it provides crucial information about the position of the object within its surrounding space. Linguistically, it is relatively easy to define an object's position in relation to the surrounding space, since we can do this by simply describing the object's position with reference to spatial landmarks, such as 'ground' or 'ceiling' or with reference to other

objects within the surrounding space, which are of a defined height, for example 'on top of the house' or 'in a tree'. However, iconic gestures are mostly not generated with reference to the 'real space', but happen within a particular gestural space, which McNeill (1992: 1) describes as a space that 'is not the speaker's space, but a fictional space, a narrative space that exists only in the imaginary world of the discourse'. For example, a narrator imitating punching someone in the face is not likely to rise from his seat just to be able to demonstrate the movement at the appropriate height at which it happened compared to the surrounding space. Similarly, the iconic gesture from example 5 provides no information about the relative position of the object within its surrounding space. The narrator uses here only his or her own body as a positional reference point, but does not rise from the seat to be able to demonstrate the exact height at which the cat is held by the dog.

Thus, in summary, with regard to our data on relative position, it is important to note that two of our more precise semantic categories of relative position, 'agent-object' and 'agent-instrument', seem to be in line with Beattie and Shovelton's observations that iconic gestures are particularly effective at communicating relative positional information, but two of the semantic categories, 'object-instrument' and 'object-space', do not conform to this pattern. This result highlights the importance of breaking the broad category of relative position down into its necessary components and suggests that the broad category of relative position as used by Beattie and Shovelton may have masked some important and indeed critical differences.

Information about relative size

Table 5 reveals some striking and extremely surprising results. Beattie and Shovelton found that iconic gestures were also particularly well suited to convey size information. But the present approach, scoring the information contained within the iconic gesture and within the speech, reveals quite the opposite pattern. Information about size of agent, object, and instrument seems almost universally to be done by the speech itself rather than by the iconic gesture. Indeed, we have only one case in which the iconic gesture seems to be adding information about the relative size of the respective instrument (this case is identical with the one case in which information about 'relative position: object-instrument' was represented gesturally, and which will be discussed subsequently).

The statistical analysis reveals that speech is significantly better than iconic gestures at conveying semantic information about the size of agent,

object, and instrument ([agent] $\chi^2 = 46.9$, p < .001; [object] $\chi^2 = 112.07$, p < .001; [instrument] $\chi^2 = 21.33$, p < .001 respectively). The following examples illustrate how size information has been represented by different narrators. Again, these examples are representative for the gestural and the linguistic data that most of the narrators generated referring to the respective 'semantic events', which represented the basis for the present analysis.

Example 6:

'Jerry's gonna [put this cue through his tail]'

[left hand rises to chest height, clenches into a fist, moves down with force, and stops abruptly near the height of the lap]

Although in this example the speech provides no information about explicit size (such as 'the *big* cat' or 'the *small* cat'), information about the size of all of the three entities involved is represented implicitly by speech. Information about the size of the agent is represented linguistically by mentioning the agent's cartoon name 'Jerry'. The name characterizes the agent's identity, and thus in the context of this cartoon also that the agent is a mouse, since the narrators used the cartoon names assuming that any interlocutor knows who the characters are that the names represent. Thus, by naming the agent, an interlocutor would know that in this case the narrator refers to the mouse, and of course, the cognitive concept (cf. Anderson and Robert 1995) of a mouse includes the information that the entity is of small size. Therefore, background assumptions, including an average size of any entity, have to be considered as information about size. Since in cartoon stories proportions are not necessarily true to life, the semantic category has been called 'relative size'. The mouse in the cartoon is compared to other features often bigger than in reality, but the relative differences are sustained throughout the cartoon (i.e., the mouse is always considerably smaller than the cat, and the cat is smaller than the dog). Apart from directly naming the cartoon characters by their cartoon names or their animal names, linguistic information about the size of entities has also been provided by referring to the entities with their pronouns, in this example the use of the word 'his'. A 'pronoun', of course, represents a grammatical substitute for the proper name that can be used when due to the context it is clear what the pronoun refers to. In other words, the pronoun 'his' provides size information about the object, since it is clear from the context of the story that it stands here for 'Tom', and thus, it is associated with the knowledge we have about the size of this cat. However, the iconic gesture accompanying this utterance demonstrates the stabbing movement that the agent performs, but does not

reveal any size information about any of the entities involved. Although the narrator transfers here into the role of the agent, he or she, of course, does not adopt the agent's size, which means that for a narrator imitating the action that an agent performs, it is hardly possible to represent gesturally information about the size of the agent at the same time. Concerning the object, the iconic gesture does not represent any information about the object's size either, since all the gesture in this example depicts, referring to the object, is that there is another entity, which the agent is acting on with the instrument, but it does not depict any specific information about the object itself.

Example 7:

> **'Jerry will prod him … with … with a pen or something that he's got, with a sharp instrument, either in his toe, or maybe in his tail, or [poke him in the bum or something]'**
>
> [right hand clenches into a fist, is swung up vertically in front of the narrator at the right side of the body and drops down again]

In example 7, information about the size of the agent and the object is, again, represented verbally by either directly mentioning the name or by mentioning the pronoun. And as in the previous example, the iconic gesture demonstrates the agent's action (the hand clenching into a fist to grab the instrument), but no information about the size of the agent or of the object is represented gesturally. Even though the gesture shows the agent gripping the instrument, it does not show what size the instrument is as a whole (in fact, the cartoon picture shows that the instrument is even taller than Jerry, which the iconic gesture does certainly not reveal). Only the segment of speech provides here information about the size of the instrument as a whole (the size of a pen). Thus, size information concerning the entities involved, again, has been provided exclusively by speech, and not by gesture.

Example 8:

> **'the warden is carrying him by … [by the scruff of his neck]'**
>
> [hand rises in front at the right side of the narrator's body, clenches into a fist, moves quickly up to the height of the shoulder and down again, then moves back into rest position]

The latter statement applies equally to example 8, as here information about the size of the agent and the object is represented linguistically, whereas the accompanying iconic gesture shows the agent's movement, or action, but it does not reveal any information about how big or

small the agent is performing the action, or about how big or small the object is, which the gesture demonstrates that the agent is picking up. Once again, the utterance provides this information by naming the agent (warden), which reveals that the agent must be the size of a relatively grown-up human, and the object is characterized concerning size by mentioning the pronoun 'him', which refers in this context to Tom the cat.

In conclusion, we would suggest that the statistical analyses of the semantic information within the gestural and linguistic channels reveal highly interesting results, since they, at times, seem to contradict the broad conclusions by Beattie and Shovelton (1999a, 1999b, 2002). In the following pages, we will offer further analyses that may provide some explanation as to why the findings of the present study suggest that information about position and size is less well encoded by iconic gestures than the studies of Beattie and Shovelton might suggest.

Gestural viewpoint and the representation of position and size information

As we have already discussed, Beattie and Shovelton's analyses provide evidence that differences in the gestural viewpoint might have an effect on the semantic information that iconic gestures convey. Here, we consider gestural viewpoint using a micro-analytic approach.

Since C-VPT gestures by definition involve the narrator him- or herself transferring into the role of the character so that the narrator's body represents the body of the character, this type of iconic gesture appears suitable to represent information about the positional relation of agent and second or third entities, such as objects and instruments. The transfer into the agent's role has the advantage that the relative position of entities other than the agent can be easily depicted by using the own body as a spatial reference point. O-VPT gestures, on the other hand, do not offer the possibility to represent information about entities in relative position to the agent as clearly and easily as C-VPT gestures. An O-VPT gesture means that the narrator's arm and hand represent the character as a whole. Thus, it is logical that it would be far more complicated to illustrate, for example, if an object is positioned behind or in front of the agent, and if it is being held at the height of the stomach or rather at the height of the face, since this would require that the narrator's arm and hand are exactly defined in terms of what represents the character's front and which part of the hand or arm represents the stomach, etc. Considering the categories that refer to positional information, but which do not involve the agent (i.e., the positional relation of object and instrument and of an object in its surrounding space), we can say that

these tend to be not represented by C-VPT gestures. Concerning the latter category (O-S), it has already been suggested that iconic gestures in general are not very good at representing information about the position of an object in 'real space', and that linguistic means seem to be more appropriate to provide this kind of information. And of course, the relative position of an instrument relative to an object can hardly be represented by C-VPT gestures, as here no agent is involved anyway (and as soon as the narrator would use his or her hands to illustrate positional information about other entities that does not refer to the agent, this would, logically, be defined as an O-VPT gesture again). However, O-VPT gestures that occurred in the present study did not represent any information about the relative position of object and instrument either. Instead, this information was provided linguistically. The reason for this can, again, be seen in the fact that it would be required to define, for example, which of the hands represented the object and which the instrument, or which part of one hand represents which body part of the object to infer, on which part of the body the object was hit by the instrument (e.g., tail). In short, it seems that positional information is represented best by C-VPT gestures, if we are talking about position that refers to an agent; otherwise, information about relative position seems to be represented best by speech, and O-VPT gestures seemed not to be involved in the representation of any kind of positional information in the present study (the single case where we found a gesture to represent information about 'relative position: O-I' was a special type of gesture; see example 11).

Indeed, O-VPT gestures only seem to represent one category of positional information, that is, the position of one object relative to another. Although this is a semantic category that does not appear in the present study, we decided to include this category in the discussion to complete our description of the effects that different gestural viewpoints can have on the representation of position and size information. To illustrate this point, we will therefore use an example taken from a different corpus of data.

Example 9:

'they're both jumping, [but Jerry is higher]'

[both hands have come to the front, the right hand is held at the height of the chest in front of the narrator, the left hand is held on the left side next to it, but at the height of the stomach. Both hands are flat, palms are pointing upward, the fingertips are pointing forward, and the hands move alternately up and down]

In this example, one hand represents Jerry, the other Tom, and the alternating up and down movements of the hands show the way they are jumping. Moreover, the fact that one hand is held higher than the other symbolizes here that one of the characters is positioned higher relative to the other, which demonstrates the representation of the relative position of two objects (as they are not actually acting upon each other, they are here seen as in the roles of objects, not as agents).

By investigating subtle differences between certain gestures, we recognized another feature that seems to characterize many O-VPT gestures, and which might also be relevant to the discussion of how different types of iconic gestures represent position and size information. Again, we use an example to illustrate our argument.

Example 10:

'he's just been stabbed with [that thing]'

[left hand rises in front of the narrator, the palm is pointing upward, the fingertips are pointing to the ceiling as well, the tip of the index finger and the thumb touch each other and let go again (4×) very quickly]

This gesture shows one of the few cases in which a narrator did not generate a C-VPT but an O-VPT gesture accompanying the verbal description of 'event 2' (see method section). Unlike a C-VPT gesture, the gesture here does not represent what Jerry has done with the poker, or Jerry directly holding the poker, but emphasizes the pointy feature about the object. This means that all kinds of positional information that has been represented by most of the C-VPT gestures referring to this event cannot be represented in this case, but therefore the gesture provides information about one particular feature of the instrument. Also, O-VPT gestures are frequently used to emphasize that objects are particularly big or small. Everyone knows probably the typical gesture, which involves that both hands are held at the right- and left-hand side in front of the body, often further apart from each other than the shoulders are broad, with the hands being flat and the fingers extended, the palms pointing inwards. This kind of gesture is generally used to show that something is particularly big, but it does not necessarily reflect an object's real size, since, for example, it may accompany a description of a remarkably big house as well as that of a big dog. All in all, it is therefore important to note that it seems as if some O-VPT gestures are generated to illustrate adjectival features that characterize actions or objects as being *particularly* big, quick, especially shaped (e.g., pointy, round), etc. This, in turn, can be brought in association with the fact pointed out before,

namely that some iconic gestures tend to represent different aspects of semantic information than other iconic gestures, and that when these differences in iconic gestures are not kept distinct or considered concerning the analysis, this can lead to distorted results.

Moreover, it shows that a gesture, which illustrates that something is especially big, and which is thus not necessarily referring to real size, differs in the information it represents from information about marked size or explicit size. And here we might consider why the results of the present study concerning size information contradict the findings by Beattie and Shovelton. In the current study, we captured information about marked size (man, cat, dog, etc.), whereas Beattie and Shovelton (1999a) asked decoders to state if any of the objects were *particularly* big or small. In some studies (1999b, 2002), Beattie and Shovelton used slightly different questions by asking instead, 'How big are any of the objects identified?' Apart from the fact that, compared to a question such as, 'What size are any of the objects identified?', the style of this question itself suggests that the objects are big rather than small, and apart from the fact that a typical feature of cartoons is that things are frequently particularly big, done particularly quickly, etc. (which are both aspects that might influence decoders' answers and thus distort the results), size information, indeed, turned out not to be significantly well conveyed by iconic gestures. This supports our argument that it is important to distinguish between size information referring to objects being particularly big or small and average size or marked size, since it hints at the fact that Beattie and Shovelton might have questioned their participants about different kinds of size information in their studies.

However, information about size has not been represented by C-VPT gestures (as defined by McNeill) either, but almost exclusively by speech. The reason for this, again, becomes clear if we consider how C-VPT gestures represent semantic information. Imagining C-VPT gestures providing information about the overall size of whole entities, such as that of an agent, an object, or an instrument, while the narrator depicts the agent and his or her action, would not only be far more complicated but rather impossible without performing a number of separate gestures, which would almost resemble a pantomimic presentation. For example, a narrator would have to provide the information that the object is a cat by depicting its ears, the tail, the height, and so on, and only after the information about the identity of the entity has been intelligibly depicted could the narrator go on to describe the recipient of the action in the same way, and in another step, the narrator would have to try to represent the relationship between these entities (i.e., who the agent is, what the agent is doing, and towards whom).

As pointed out before, in one exceptional case a gesture was seen to represent information about size as well as about the relative position of object and instrument — both semantic aspects that were represented purely linguistically in all other cases. Of course, we were curious to find out in which way this gesture differed from the rest of the gestures. Since we found, in the course of further analysis, that the way iconic gestures represented certain semantic features explained a lot of differences concerning the representation of semantic information, we traced back which of the gestures that were analyzed constituted the one exception (example 11).

Example 11:

'but [that spike is still in his tail]'

a) [the right hand rises quickly from the height of the lap in front of the narrator's body to about shoulder height, while the index finger is extended. When the hand reaches shoulder height, it immediately moves down again the same way]

b) [when the hand reaches the starting point again (height of the lap in front of the body), without stopping, the hand moves directly from this position horizontally backwards, still with the index finger extended, so that at the end of the movement the hand points behind the narrator's back]

A closer look at this gesture-speech compound reveals that we are here confronted with a very special type of iconic gesture: one that starts as an O-VPT gesture and finishes as a C-VPT gesture. McNeill (1992) defined this kind of iconic gesture as incorporating a 'dual viewpoint', which he exemplified himself by the following gesture:

Example 12:

'and then he [throws] a five-hundred-pound weight'

[hands move forward and down, showing throwing motion, and end in points, showing the direction of the weight]

(extracted from McNeill 1992: 98).

McNeill described this gesture as a 'dual viewpoint gesture', because it consists of two components in the sense that 'the gesture begins as a character viewpoint of throwing and ends in pointing, which has

the viewpoint of an observer' (McNeill 1992: 98). Equivalently, the gesture in example 11 begins as an O-VPT gesture a), since here the index finger demonstrates the dimensions of the stick (index finger moving up and down along the dimensions of an imagined stick), then it continues as a C-VPT gesture b), because the hand with the extended index finger moves slowly backwards, away from the narrator, showing the dimensions of an imagined tail, and at the end, the hand points towards the end of the imagined tail to demonstrate where the stick has been put through.

Thus, we can say that this gesture, incorporating a special kind of viewpoint, is partly composed of an O-VPT gesture that represents size information. Why, in this case, the size of the instrument was represented gesturally, unlike in all of the other cases, we cannot explain, but one assumption would be that it has something to do with the special kind of composition of the gesture and the dual viewpoint. Concerning the fact that in this example information about the relative position of object and instrument was also provided gesturally, even though this was not the case in any of the other cases, we can here identify the reason more precisely by saying that the gesture differs subtly from a C-VPT gesture as defined by McNeill. Although the narrator transfers here into the role of the character, which is indicated by the fact that he or she demonstrates the tail as starting at the own backside and the spike as being stuck behind the narrator at a place where the imagined tail ends, the narrator's arms and hands do not represent the character's arms and hands (they obviously perform movement, which the character's hands do not perform). Since, in association with the analysis of other data, we found a number of other examples in which the narrator transferred into the role of a character and at the same time used his or her hands to specify a region on the body of the character, or a body part (see example 13), and since these gestures thus bear crucial differences to the C-VPT gestures as defined by McNeill, we decided to distinguish them as a special type of C-VPT gestures by terming them 'BARP-gestures' ('Body as Reference Point').

Example 13:

'he's [got soot or whatever it is on him]'

[right hand moves close towards the narrator's chest, the hand is flat, the palm is facing the chest, the hand moves up and down quickly in small movements to indicate the area on the dog's body where the soot is]

Once again, we thus have to note that subtle differences indicate that we are here dealing with a special type of gesture, which is also characterized

by the fact that it differs from other iconic gestures in the way it refers to a category's position and size. Although the separation between both channels of communication concerning size information might seem clear, example 11 proves that it is possible that iconic gestures *can* represent size information.

In conclusion, we can say that it seems that crucial differences exist between C-VPT gestures and O-VPT gestures in the way that they represent semantic information about position and size. Moreover, within both of these categories, we recognized subtle but important differences in how they provide position and size information. Different types of iconic gestures seem to be more or less effective at encoding information about position and size, even if in some cases these differences seem to be rather subtle. Generally speaking, our findings support those of Beattie and Shovelton (2002), since our analyses show that C-VPT gestures are better at representing positional information than O-VPT gestures; but our analyses go beyond Beattie and Shovelton's findings, since we show more clearly how C-VPT gestures represent information about relative position, and thus we offer a plausible argument as to why positional information is predominantly represented by C-VPT gestures. Furthermore, our analyses reveal why O-VPT gestures tend not to represent positional information. However, there is one particular aspect of relative position, at which O-VPT are particularly effective, and even better than C-VPT gestures. Therefore, we conclude that differences in the gestural viewpoint are a very significant aspect of how iconic represent and communicate semantic information.

Discussion

The findings of the present study revealed a clear pattern showing how iconic gestures and speech represent information about position and size. These findings suggest that certain types of positional information are represented gesturally and others linguistically, and furthermore, that size information in the present study seems to have been exclusively encoded linguistically, except in one particular case. Some of these findings thus stand in contrast to those of Beattie and Shovelton (1999a, 1999b, 2002), who claim that information about both size and position are especially well represented by iconic gestures.

There are a number of possible explanations for this divergence. One possible reason might be that the present study is based on the information that is actually encoded in gesture and speech (as seen from an analyst's perspective), rather than what information is actually

transmitted to decoders. This may lead us subsequently to the question of which types of information within a signal are picked up by decoders and which are not.

Furthermore, in the present study, we analyzed gestural and verbal data that has been generated during dialogues instead of during monologues. As we argue in the introduction, there is solid evidence that the frequency of gestures increases dramatically in association with conversational interaction. In the present study, we found that only 13.2 percent of the participants (5 out of 38) did not produce any gestures at all, which stands in considerable contrast to studies based on monologue situations, and in which on average about 50 percent of the participants do not produce any gestures (see Beattie and Shovelton 1999a). Thus, we would claim that the present study based upon conversational data, in which a high proportion of participants use gestures, is more representative of everyday talk and clearly entails a higher proportion of our sample of participants being involved in the actual analysis.

Furthermore, in terms of research on gestures and speech, McNeill had employed a traditionally linguistic method, with the analysis based exclusively on clausal units, which Beattie and Shovelton also attempted to apply in their studies, with the notable exception of their first study (1999a), in which the size of the speech extracts varied (example 15 below represents just such a case, in which Beattie and Shovelton used a speech extract considerably larger than an individual clausal unit). In the introduction, we hinted that one consequence of this focus on the clause might be that certain aspects of semantic information were ignored because they were represented linguistically in those parts of speech outside the clausal units under consideration. Thus, in the present analysis we used ideational units, instead of clausal units, as exemplified below.

Example 14:

> **'Jerry will prod him … with … with a pen or something that he's got, with a sharp instrument, either in his toe, or maybe in his tail, [or poke him in the bum or something]'**
>
> [right hand clenches into a fist, is swung up vertically in front of the narrator at the right side of the body, and drops down again]

Considering this example, it becomes obvious that the clausal unit with which the iconic gesture co-occurs ['or poke him in the bum or something'] contains only a fragment of the semantic information that is uttered verbally to describe the semantic idea completely. Here, it would not be possible to argue that the gesture refers only to the verbal

description of the movement of the action, i.e., the poking movement, since the narrator himself transfers into the agent and shows the grip around the instrument, which the narrator refers to verbally, but previous to the respective clausal unit. This illustrates that here an analysis on a clause-by-clause basis would have revealed a slightly distorted picture, since the result would show that there is no linguistic information about either the agent or the instrument.

As seen in the results section, the name of the cartoon character also conveys, at least implicitly, information about the character's size. In this case, size information about the agent is provided verbally by mentioning the word 'Jerry'; however, this word is not part of the clausal unit that the iconic gesture accompanies. Thus, in association with the present study, this would have been scored as speech providing information about the size of the agent, whereas an analysis based on a clause-by-clause analysis would have revealed that linguistically there is no information about the size of the agent represented. This is only one example, but it should make clear that such differences accumulated over the whole analyses of a large corpus of data might be considerable, and that therefore this might be another reason for the discrepant results of the present study compared with the Beattie and Shovelton studies.

Furthermore, differences concerning the scoring criteria that are applied to the data can lead to quite different results (even when identical semantic categories are being considered). This applies in particular to information about size, since Beattie and Shovelton seemed to have used slightly different scoring criteria than we used here. In their first study (1999a), referring to the iconic gesture described in example 15, they asked their participants, 'Is the bench seat extremely large?' (1999a: 16), and they considered 'no' as being the correct answer, because 'the gesture consisted of circular movements of the index finger, and the circular movements were small' and therefore, 'the gesture was believed to convey the idea that the bench seat was not particularly large' (1999a: 16).

Example 15:

'Um they take the table home and it's actually a round table with a bench sort of [seat around it (pause)], which looks alright, because it doesn't have wizzy chairs or anything'

[left hand's index finger makes small circular movements]

But was this really the most appropriate way to score the responses? One might here argue that the iconic gesture in this example was in the first

place used to depict the round shape of the bench seat, but not its size, since the bench seat was neither particularly small nor particularly large. If the gesture here were considered as representing size information, the only logical interpretation would be that it shows that something is particularly small, which is represented by the extremely small circle movements of the index finger. However, this is not what the cartoon suggested, and therefore it does not seem plausible to assume that the iconic gesture represented size information at all. Beattie and Shovelton, however, considered the gesture to convey size information due to the fact that the gesture did *not* show that the respective entity was particularly big, since they argued that the gesture thus shows that the bench seat was not extremely large. Thus, Beattie and Shovelton's particular method for scoring size information seems to rule out the possibility that an iconic gesture may represent no size information at all, and that they may represent information about size only when something is remarkably big or small. Logically, using a methodology like this might have prevented Beattie and Shovelton from finding that iconic gestures tend not to represent information about average size.

If a different narrator represented the same semantic aspect with the same gesture, apart from the fact that the circles generated were of double the diameter, would decoders answering the same question then have stated that the bench seat was particularly large? In other words, the point is here what is being considered as *critical* size information, and also what is possible to be depicted gesturally. It is hardly possible to represent the actual size of the bench seat gesturally, unless we stand up and walk around in a circle imitating the size of the circumference of the bench with our hands. Furthermore, referring to the kind of size information that is considered as being relevant, it has to be noted that it makes a difference, if any kind of size information that is represented gesturally is scored, or if only size information is considered as being relevant that meets certain criteria. This becomes clearer, if we consider example 6 from the results section:

'Jerry's gonna [put this cue through his tail]'

[left hand rises to chest height, clenches into a fist, moves down with force, and stops abruptly near the height of the lap]

Of course, the iconic gesture in this example, if seen by a decoder in the absence of speech, could be judged as conveying some kind of size information. For example, the decoder might state that the narrator imitates holding a *short* knife in the hand by inferring this information

from the size of the fist, which appears to embrace an imagined instrument. However, the stimulus material shows the mouse holding an instrument, which is of a far larger size than the mouse itself, and of which the mouse's hand covers only about a tenth of the actual length. This makes clear that, although the iconic gesture might provide here *some* size information, it does not necessarily mean that this is also *critical* size information, since the instrument is here not small at all. In the present study, we determined that one of the criteria for scoring size information is 'size of entities as a whole', which in this case has been solely provided linguistically, as the word 'cue' describes an instrument of a relatively defined size. If the information of gesture and speech is not matched according to strict criteria such as these, i.e., with close reference to the scene, confusion of different aspects of size information can occur, which is similar, for example, to the confusion of different types of positional information. Since it seems that Beattie and Shovelton did not determine such strict criteria and thus combined different kinds of size information (information that is critical with reference to the scene versus some kind of size information), as well as size information referring to different entities, this might be another explanation for the contradictory findings.

Furthermore, we have argued that there are different types of iconic gestures particularly with regard to the viewpoint that they incorporate, and we demonstrated that these differences in gestural viewpoint can have a considerable effect on the representation of position and size information. Thus, it is logical that an analysis of the semantic features that are represented by iconic gestures, which is not based on a careful differentiation of 'subtypes' of iconic gestures, makes any accurate interpretation of the results considerably harder. In the present study, we tried to control for this factor by analyzing in the first place C-VPT gestures, and only in cases in which such data was rare, we included other types of iconic gesture (10 gestures out of 58). However, we took care to ensure that it was possible to trace these cases back at any point during the analysis and also concerning the interpretation of the results. We were able to show that C-VPT gestures do seem well suited to represent positional information, and we suggested why this is. It turned out to be beneficial that we employed a micro-analytic approach, since the detailed semantic categories that we used, as well as the careful differentiation made between subtle differences in gestural viewpoint, revealed that Beattie and Shovelton's analyses may have masked the fact that O-VPT gestures seem able to represent one particular aspect of positional information even better than C-VPT gestures.

To sum up, the fact that Beattie and Shovelton's findings were based on an analysis of a combination of iconic gestures which differed in viewpoint, whereas the present findings were derived from an analysis mainly based on C-VPT gestures, may account for some of the differences that emerged from these two studies.

Summary

This study, using a micro-analytic approach, revealed some of the detailed ways in which the linguistic system and the gestural system interact in the communication of semantic information. We used an approach similar to that of McNeill (1985), rather than an approach using decoders like Beattie and Shovelton (1999a, 1999b, 2002), and we attempted to determine what information was present in the linguistic system and what was present in the gestural system. The research suggested that relative position for at least two of the possible categories of relative position information was conveyed gesturally and that for two other categories it was conveyed verbally. In terms of size information, the information tended to be conveyed verbally. But what is interesting is that the two systems rarely overlapped in terms of the communication of semantic information. In other words, they are two systems, which are extremely complementary in this domain.

Of course, the fact that in this study we did not find conclusive evidence that information about position and size was exclusively done through iconic gestures raises some interesting issues about how we can reconcile these findings with those of the decoder studies of Beattie and Shovelton. We suggested a number of ways in which the two types of studies may differ. Of course, the information may be present in the speech signal or in the gestural signal, but may not be employed by decoders, and this could be a rather simple explanation for the discrepancy. But, we also argued that, because we took extracts from more natural types of conversational data in the present study, which involved a higher proportion of participants, this must operate in favor of the ecological validity of the present data. We also raised specific issues about how Beattie and Shovelton have scored positional and size information. What we are not attempting to do here is to substitute one type of methodology for another, rather we are trying to obtain a fuller picture of how iconic gestures and speech work together in the communication of semantic information.

Concerning McNeill's communication theory (1985), the present study enables us to gain more insight into the kind of role that iconic gestures

play in the communication of semantic information. By applying a range of very detailed semantic categories, it was possible to carry out a fuller analysis of the types of position and size information that was contained in gesture and speech. Interestingly, the results reveal that information about position and size is clearly partitioned between iconic gestures and speech in the way that certain positional aspects are exclusively represented by iconic gestures, whilst other aspects of relative position, as well as information about size, are almost exclusively represented by speech. Thus, the present findings support McNeill's hypothesis that the gestural and the speech systems must be closely interlinked, since they seem to be well tuned in the way that they provide semantic information about position and size. In other words, apart from one case, none of the information provided by gesture or speech about 'relative position' and about 'relative size' was redundant. Thus, we can be more precise concerning statements about the interaction of iconic gestures and speech than McNeill; the findings of this study suggest that it seems that both channels of communication, gesture and speech, complement each other in a near-perfect fashion, at least concerning the representation of positional and size information.

   Since this research emphasizes the communicational role of iconic gestures, one goal of future research must consist of testing the interaction of iconic gestures and speech using a broader range of semantic categories. Only by investigating an even fuller range of semantic aspects can we discover whether the pattern of complementarity between both systems of communication holds more generally. Only then will we truly understand how the linguistic and the gestural systems cooperate to communicate meaning in everyday talk.

Appendix

Transcription conventions:
Segments of speech analyzed are marked using 'single quotes'. Dots indicate pauses in speech. That part of the verbal utterance that was directly accompanied by the iconic gesture is marked using [square brackets], thus the brackets indicate the start and end points of the iconic gesture with respect to the speech. The iconic gesture that accompanied the verbal utterance is described underneath the extract of speech in each case, and this description is also contained within square brackets. In the present study, we considered the iconic gesture as a whole in our transcriptions, whereas in the gesture transcriptions extracted from

McNeill (1992) only the start and end points of the stroke phase of the iconic gesture are marked (see McNeill 1992: 12).

References

Anderson, John R. and Robert, John (1995). *Cognitive Psychology and its Implications*, 4th ed. New York: W. H. Freeman.

Beattie, Geoffrey and Aboudan, Rima (1994). Gestures, pauses and speech: An experimental investigation of the effects of changing social context on their precise temporal relationships. *Semiotica* 99 (3/4), 239–272.

Beattie, Geoffrey and Shovelton, Heather (1999a). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica* 123 (1/2), 1–30.

—(1999b). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology* 18, 438–462.

—(2002). An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *British Journal of Psychology*.

British Psychological Society (2000). Code of conduct, ethical principles and guidelines [brochure]. Leicester: The British Psychological Society.

Butterworth, Brian (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research* 4, 75–87.

Efron, David (1941). *Gesture and Environment*. New York: King's Crown Press.

Ekman, Paul and Friesen, Wallace V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica* 1 (1), 49–98.

Kelleher, Pat (2000). *Tom and Jerry* 30, 22–35. Manchester: Just Publishing Ltd.

McNeill, David (1985). So you think gestures are nonverbal? *Psychological Review* 92 (3), 350–371.

—(1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

Schegloff, Emanuel A. (1984). On some gesture's relation to talk. In *Structures of Social Action: Studies in Conversation Analysis*, J. Maxwell Atkinson and John Heritage (eds.), 266–296. Cambridge: Cambridge University Press.

Judith Holler is a Ph.D. student at the University of Manchester <judith.holler@stud. man.ac.uk>. Her research interests include the interaction of iconic gesture and speech in conversational interaction.

Geoffrey Beattie is Head of the Psychology Department at the University of Manchester <beattie@psy.man.ac.uk>. His research interests include discourse analysis, nonverbal communication and its relationship to speech, and ethnography. His major publications include 'An experimental investigation of the role of iconic gestures in lexical access using the tip-of-the-tongue phenomenon' (with J. Coughlan, 1999), 'Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech' (with H. Shovelton, 1999), 'Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation' (with H. Shovelton, 1999), and 'Iconic hand gestures and the predictability of words in context in spontaneous speech' (with H. Shovelton, 2000).