

The Act of Speaking: Spoken  
Language and Gesture in the  
Determination of Definiteness  
of Intention

Richard Hirsch

Chapter in the book *Indexicality: Papers from the Symposium "Indexikalia Tecken"*,  
*University of Göteborg, November 1995*  
Editor: Christiane Pankow  
Series: SSK KII 9604  
ISBN: 91-630-4786-1

# THE ACT OF SPEAKING: SPOKEN LANGUAGE AND GESTURE IN THE DETERMINATION OF DEFINITENESS OF INTENTION

Richard Hirsch  
Linköping University  
richi@lin.se

## Introduction

The roman rhetorical tradition acknowledged the importance of gesture and made the appropriate use of gesture an important part of the 'actio' of a speech. Quintillian devoted a large portion of one of the four books of his *Institutio Oratoria* to a discussion of the proper use of gesture by an orator. Mainstream modern linguistic theorizing has had a condescending or downright antagonistic attitude toward gesture. Due to a Cartesian dualistic bias where body and mind are strictly separated and to a concentration on the enterprise of accounting for linguistic competence rather than linguistic performance, gesture occurring in connection with spoken language has generally been ignored as irrelevant.

This situation is however changing. Linguists are coming together with communication scientists, anthropologists, psychologists, and others to study the actual use of spoken language in a variety of everyday situational contexts. In this regard I would refer the reader to the excellent work being done by Charles and Marjorie Goodwin on the analysis of video recordings of language use in natural settings (cf. Goodwin & Goodwin 1992). Recent psycholinguistic research shows that speech and gesture are probably neurophysiologically related (cf. McNeill 1992 and Feyereisen & de Lannoy 1991).

Most studies of the integration of gesture and speech have been 'syntactically' oriented, i.e. determining the temporal order of occurrence of the gesture and the corresponding speech segments. Usually as a effort to investigate the process of speech production in relation to thought (cf.

Feyereisen & de Lannoy 1991). McNeill (1992) has however started to move in a more semantic direction and has studied the use of illustrative and metaphoric 'imagistic' gestures in connection with speech.

In contrast to Decartes, C.S. Peirce realized that knowledge or cognition has three basic semiotic dimensions; iconic, indexical, and symbolic. Peirce claimed that these three dimensions of cognition were grounded in intuitions of similarity, causality, contiguity in space-time and part-whole, and arbitrary conventional connections between objects (abstract or concrete) of attention. In a Peircian semiotics the iconic and indexical dimensions of signs are primarily non-verbal, the symbolic dimension is primarily verbal.

The question to be addressed in this paper is how the non-vocal, non-verbal aspects of gestures are related to the vocal, verbal aspects of spoken language (speech). The relationship that will be explored and discussed is a semantic one. To this end I have turned to Arne Naess's (1953) Theory of Interpretation and Preciseness for inspiration. Naess's theory is meant to be a tool for semantic analysis of communicative language use either spoken or written. I am going to generalize Naess's semantic insights in a semiotic direction to cover both gesture and speech.

### **Terminology and Definitions**

Naess (1953) uses the term *expression* to refer to a linguistic formulation, usually a statement or a sentence. I use *expression* in a wider sense to refer to a linguistic vocal, verbal sign and/or a non-vocal, non-verbal sign. Based on a reformulation of Grice's (1957) distinctions between natural and non-natural meaning to be found in Allwood (1976) and Hirsch (1989) I define *expression* as follows.

#### **Expression**

**Definition:** A notable movement of the body or part's of the body that,

- i) **indicates**, i.e. functions as a source of information to an observer
  
- ii) **displays**, i.e. is intended to make a receiver at least apprehend or attend to certain information, through some manner of apprehension like direct perception or inference.

iii) **signals**, i.e. is intended to make a receiver at least apprehend a display of certain information, through direct perception or inference. (cf. Grice's notion of Meaning NN )

iv) **symbolizes**, i.e. functions by convention as a representative display of information.

Different types of expression are related to the distinctions verbal, non-verbal, and vocal, non-vocal as illustrated in table 1.

**Table 1: Expression Types**

	<b>Vocal</b>	<b>Non-vocal</b>
<b>Verbal</b>	Spoken Language	Written Language
<b>Non-verbal</b>	Vocalizations	Gestures

As can be seen in table 1, gestures are non-verbal, non-vocal expressions. I define gesture for the purposes of this paper, as follows.

### **Gesture**

Definition: semiotic phenomena characterizable as non-conventional, non-vocal, nonverbal, non-alter contact communicative behavior produced by movements and/or configurations of the upper extremities of the body, i.e. the hands, fingers, arms, shoulders and head.

The aim of analysis in a Naessian semantics is to gain insight into how people go about dealing with communication problems involving the achievement of cognitive understanding or the avoidance of cognitive misunderstanding, i.e. how people arrive at mutually confirmed agreement about what is meant by what is expressed even though parties may disagree as to whether what is meant is true or false, practical or impractical, etc. Naess introduces the notion of *definiteness of intention* (which he and others following him later called depth of intention) to capture the insight that speakers can vary as to the number of cognitive distinctions that they are

aware of and prepared to take into consideration in a particular situation in connection with the use of a particular expression. This holds also for the same speaker in different communicative situations.

Situational and personal relativity in the cognitive understanding connected with expressions in communication is part of a more general and fundamental insight that I sometimes refer to as the Naess semantic uncertainty principle which can be formulated in a weak or a strong version.

**Weak version:**

It is not necessarily the case that speakers have a particularly well determined definiteness of intention in mind when expressing themselves.

**Strong version:**

In many/most cases in everyday life, speakers do not have a particularly well determined definiteness of intention in mind when expressing themselves.

The semantic uncertainty principle leads to the following consequence.

**Consequence:**

What *definiteness of intention* speakers have/had in mind when expressing themselves can, in certain (weak)/most (strong) cases, only be *determined by a process of interpretation*.

Naess's semantic theory was primarily developed to deal with the interpretation of verbal - ~~verbal~~ <sup>vocal</sup> expressions or formulations. For Naess, interpretation consists basically of the determination of definiteness of intention by means of various types of reformulations. My extension of Naess's theory attempts to incorporate the interpretation of verbal and non-verbal expressions within the same basic framework where the determination of definiteness of intention is effected by a combination of vocal, verbal and non-vocal, non-verbal means.

The framework I am going to use for analysis of the determination of definiteness of intention in face-to-face spoken interaction consists of the following basic assumptions and definitions. In these definitions it should be born in mind that the term *expression* covers both vocal, verbal and non-vocal, non-verbal expressions, i.e. speech and gesture.

1. The *definiteness of intention* of expressions varies according to situation and personal interpretations of the expressions.

Comment: This is a direct consequence of the Naess semantic uncertainty principle discussed above.

2. An expression T is an *interpretation* of an expression U if and only if there is at least one situation in which for at least one person T and U are understood cognitively in the same way or have the same cognitive effects, i.e. are equivalent in cognitive adequacy.

Comment: For instance, let T be the vocal, verbal expression "OK" and U be the gesture 'thumb-index ring'. In certain situations T:"OK" is used as an interpretation of U: 'thumb-index ring'. Especially in situations where U: 'thumb-index ring' occurs first in face-to-face spoken interaction.

3. Two expressions (T and U) are cognitively non-equivalent for at least one person in at least one situation if and only if there is one *imaginable* set of circumstances for which the person(s) would claim that the one expression (T) was cognitively adequate and the other expression (U) cognitively inadequate.

Comment: This can be illustrated in the case of two vocal, verbal expressions, say T:"block" and U:"cube". These two expressions are cognitively non-equivalent in a situation in which speakers are talking about a collection of blocks where some blocks are cubes and others are not. T: "block" would, for instance, be cognitively adequate as an expression referring to a pyramid, whereas U: "cube" would not. In a situation where a speaker held the gesture 'thumb-ring index' to be an expression referring to money, as for instance in Japan, T:"OK"(or the Japanese translation equivalent) and U:'thumb-index ring' would not be cognitively equivalent.

4. A necessary but not sufficient condition for equivalence of cognitive understanding is that two expressions are not cognitively non-equivalent.

Comment: In the case of vocal, verbal expressions this condition rules out T:"Good food is not cheap" being equivalent to U:"Cheap food is not good"

although both can be reduced logically to the same expression, namely 'No food is both cheap and good', i.e. both can be shown to have the same truth-conditions. The expressions are, however, obviously not cognitively equivalent.

5. Two expressions (T and U) are equivalent in *definiteness of intention* if and only if for at least one person in at least one situation all interpretations of T are interpretations of U and all interpretations of U are interpretations of T. All ways of understanding T are going to be ways of understanding U and vice versa.

Comment: In our example T:"OK" and U:'thumb-index ring', it is possible that for at least one person in at least one situation all ways of understanding T:"OK" will be ways of understanding U:'thumb-index ring', namely 'good', 'right', 'agreed', etc.

6. Two persons have the same definiteness of intention in relation to an expression if and only if the expression is equivalent in cognitive adequacy for both persons, i.e. all ways of understanding the expression by one person are going to be ways of understanding the expression by the other person, and vice versa, in at least one situation.

Comment: This definition covers the case when we have mutual understanding and confirmed agreement on the meaning of an expression by two persons in a communicative situation.

7. Two persons have sufficiently equivalent definiteness of intention in relation to an expression if and only if they can use a set of interpretations of the expression to arrive at a mutually confirmed agreement as to the meaning of the expression in order to solve at least one communication problem in a least one situation.

Comment: This definition covers the normal state of affairs in what is usually referred to as 'successful communication' in face-to-face spoken interaction. For instance, in a certain interaction involving reference to blocks, two persons may be able to use a set of interpretations of an expression such as "the square one" to solve a communication problem in the ongoing interaction, although the concept of a square block may be conceived of as a contradiction in terms outside of the ongoing interaction.

8. An expression T is more precise or effects a greater definiteness of intention than an expression U if and only if all interpretations of T are interpretations of U and there is at least one interpretation of T which is more clearly decideable in its application and non-application to any given entity or phenomenon than other interpretations of U.

Comment: The set of interpretations of T is a proper subset of the set of interpretations of U, and the certainty of application and non-application associated with T is greater than that associated with U. All ways of understanding T: "a three dimensional closed figure with only square faces" are going to be ways of understanding U: "block" and at least one interpretation of T, e.g. "cube", is more clearly decideable than any other interpretation of U. All ways of understanding the expression T: 'thumb-index ring' are going to be ways of understanding the expression U: 'configuration of thumb and index finger' and at least one interpretation of T: 'thumb-index ring', e.g. "agreed" is more clearly decideable than other interpretations of U: 'configuration of thumb and index finger'.

9. An expression T is less precise, i.e. more vague (expresses less definiteness of intention) than an expression U if and only if all interpretations of T are interpretations of U and there is at least one interpretation of T which is less clearly decideable in its application and non-application to any given entity or phenomenon than other interpretations of U.

Comment: The set of interpretations of T is a superset of the set of interpretations of U. And the certainty of application and non-application associated with T is less than that associated with U. This is the exact inverse of the precision relation between expressions.

10. An expression T is an elaboration of an expression U if and only if all interpretations of T are interpretations of U and all interpretations of U are interpretations of T, and T is neither more nor less precise than U.

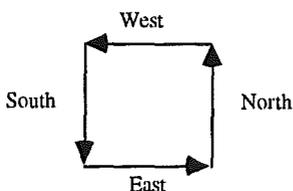
Comment: Elaborations of an expression are either some sort of paraphrases or have the effect of making the expression more or less specific rather than more or less precise (cf. Naess 1953, Pinkal 1986, Hirsch 1989).

In the next section of this paper, I am going to demonstrate with empirical evidence how speakers' vocal, verbal and non-vocal, non-verbal expressions collaborate in the incremental determination of definiteness of intention in ongoing spoken interaction. Speakers' expressions will be seen to be part of an historical process of determination of definiteness of intention within an interactional context where there is an incremental development of information involving an integration of indicative, display, signal, and symbolic expressions over the course of one speaker's turn at talk or across turns and speakers in the course of the spoken interaction.

### **Empirical investigation**

The experimental situation consisted of a studio video recording of three students participating in a problem-solving task involving the construction of a marble track according to a specification given by me, the experiment leader. The communicative behavior exhibited by the participants was not elicited or directed by the experiment leader. The construction task involves many complex geometrical and spatial relations and can therefore be expected to generate gestures in connection with speech in the interaction (cf. Feyereisen & de Lannoy 1991). The construction specification consisted of asking the students to build a track with a set of blocks that would allow a small steel marble to roll west, south, east, and finally north so that the marble ended up directly below the position at which it started as illustrated in figure 1.

**Figure 1.**



In this task situation most gesturing occurs in connection with a direct manipulation of the construction blocks. I define a direct manipulation gesture as follows.

### **Direct Manipulation**

Definition: Grasping and/or moving a block or a part of the blocks construction with one or both hands.

Table 2 contains a nonexhaustive collection of examples of cases of vocal, verbal expressions used in conjunction with non-vocal, non-verbal direct manipulation gestures which occurred in the video recorded problem-solving session. These gestures can be seen as illustrations of what Searle has called the limiting case of saying, namely saying that involves showing (Searle 1969: 88).

The direct manipulation gesture may be characterized as an expression which has less indefiniteness of intention than the preceding vocal, verbal expression. Here the speaker is actually manifesting an intention in practical action in full view of the co-participants. The direct manipulation gestures in table 2 can be classified as signal expressions according to our typology of expressions in section 1. The vocal, verbal symbolic expressions are complemented by non-vocal, non-verbal signal expressions to accomplish a demonstration of the speakers' intentions.

**Table 2.**

1. så <gesture> <i>like this</i>	<DM>	2. så där <gesture> <i>like that</i>	<DM>
3. vi ska ha en sån <gesture> <i>we need one like this</i>	<DM>	4. ta en sån där <gesture> <i>take one like that</i>	<DM>
5. det måste var så <gesture> <i>it should be like this</i>	<DM>	6. den här måste också ha ett hål så <i>this one has to have a hole in it so</i> kulan kan trilla ner <gesture> <i>the marble can roll down</i>	<DM>
7. vi ska ha en sån <gesture> <i>we should have one of these</i>	<DM>	8. ta en sån här med hål i <i>take one with a hole in it</i>	<gesture><DM>
9. sätt på en sån här istället <gesture> <i>put one like this instead</i>	<DM>	10. kan man inte byta plats <i>couldn't you change places</i>	<gesture><DM>
11. kan man inte sätta den under <i>can't you put that under</i> <gesture>	<DM>	12. om vi vänder den och sätter den så <i>if we turn it and put it like this</i> istället <gesture> <i>instead</i>	<DM>
13. vi kan prova med det här <i>we can try with this</i> <gesture>	<DM>	14. men om man lägger den <i>but if you lay that</i> <gesture>	<DM>
15. den här lutar kan man säga <i>this one slants you might say</i> <gesture>	<DM>	16. den där så <gesture> <i>that one like this</i>	<DM>
17. vi skulle ha den <gesture> <i>we should have that one</i>	<DM>	18. den här då kanske kan åka in i den <i>this one could maybe roll into that one</i> <gesture>	<DM>

(See note 1 at the end of the paper for transcription conventions.)



Frank: <gesture> <both hands out palms up>

The vocal, verbal question "what are we doing?" that accompanies the expression <torso lean back hands out palms up> incorporates the non-vocal, non-verbal display into the ongoing interaction as a stretch of behavior relevant to the purpose of the problem-solving exercise.

The course of development continues with a series of direct manipulation gestures introduced and accompanied by vocal, verbal expressions.

de här ä mycket smidigare att  
*this is much more elegant to*

göra så här  
*do like this*

<ta den> (...) <DM>  
*take this*

<så sätter vi upp den> <DM>  
*and we put this up*

<då gör vi så> <DM>  
*then we do like this*

What I want to propose is that direct manipulation gestures like those exemplified above can be characterized as non-vocal, nonverbal precisification operations on the definiteness of intention of the vocal, verbal expressions. These gestures are simply more restricted in their interpretative potential than the preceding speech expressions. There are fewer ways of understanding or interpreting these gestures than there are ways of understanding or interpreting the speech expressions. The sets of interpretations of the demonstrative direct manipulation gestures are proper subsets of the sets of interpretations of the speech expressions and there is more certainty concerning their application or non-application to the situation at hand in the ongoing interaction.

This does not mean, of course, that all such direct manipulation gestures can be viewed as precisification operations. Neither should it be taken to mean that all vocal, verbal expressions have non-vocal, non-verbal counterparts of

equivalent cognitive content and vice versa. However, in this particular type of situation many vocal, verbal expressions are made more determinate as to their definiteness of intention by being combined with a direct manipulation gesture.

In example 2, we find the participants cooperating to produce what I call an intersubjective course of development of definiteness of intention. The speakers work on each others' expressions trying to better determine an initially indeterminate definiteness of intention.

### Example 2

Jane: det måste vara fyra våningar högt  
*it has to be four stories high*

<gesture> eller någonting      <right index rise>  
*or something*

Frank: för man<tappar ju en för varje><right hand illustrates  
*because you lose one for each      stepwise drop>*

Linda:                      <går ner till den nivå sen går ner  
*go down to that level then go down*

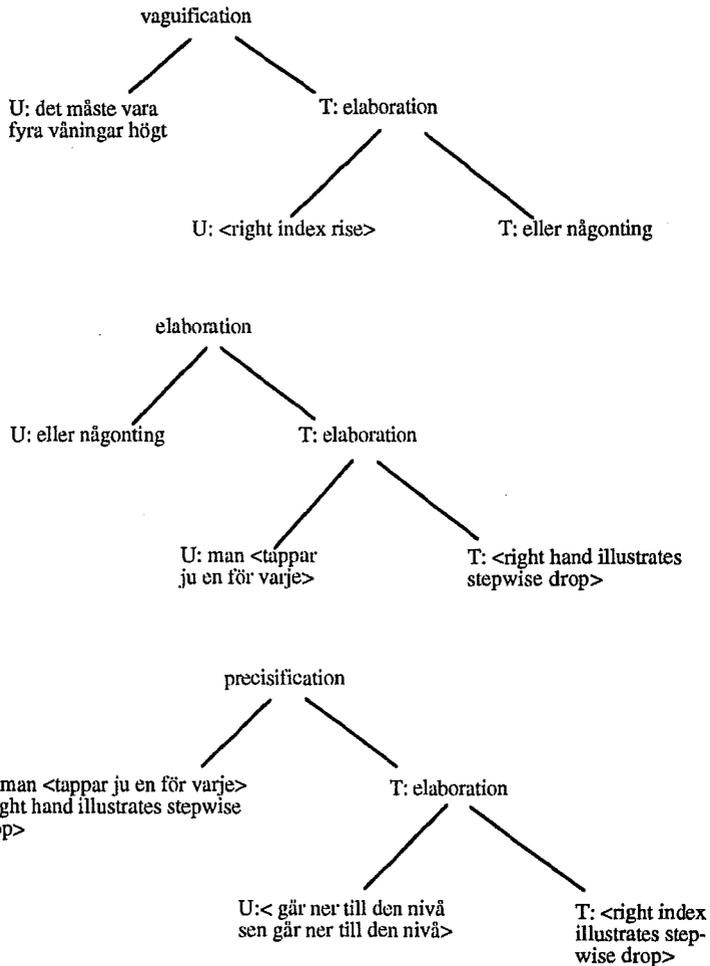
till den nivå>                      <right index illustrates  
*to that level                      stepwise drop>*

The development takes place in a basically recursive manner where an expression by one speaker is made more precise, more vague, or elaborated on in some manner, either by the use of expressions from the same speaker or by expressions from other speakers. Figure 2 contains an analysis of the recursive structure of the development. The expressions labeled U are expressions that are determined in some manner or other by expressions labeled T (see the definitions above). An expression labeled T may itself consist of expressions which consist of a determined part (U) and a determiner part (T).

As figure 2 illustrates, vocal, verbal and/or non-vocal, non-verbal expressions by a speaker are subjected to determination by means of other vocal, verbal and/or non-vocal, non-verbal expressions, either by the same

or another speaker. The determination is seen to be recursive in nature where, for instance, a precisification may be elaborated on or an elaboration may in turn be elaborated on.

Figure 2.



### Conclusion

The discussion of these examples of the integration of speech and gesture in face-to-face spoken interaction demonstrate a point made by Grice in the

closing line of his 1957 article 'Meaning' where he claims that "to show that the criteria for judging linguistic intentions are very like the criteria for judging nonlinguistic intentions is to show that linguistic intentions are very like nonlinguistic intentions". Both linguistic and nonlinguistic intentions are matters of interpretation and vary in definiteness: compare the direct manipulation gestures that accompany the short vocal, verbal expressions in table 2 and the vocal, verbal accompaniment of Frank's <torso lean back hands out palms out> expression in example 1. This variation in definiteness is something that must be dealt with and worked out to the satisfaction of the participants within the ongoing face-to-face interaction.

I hope to have shown that we may profitably employ a semiotically generalized version of Arne Naess's theory of Interpretation and Preciseness to account for the interpretative work that is carried out by participants in face-to-face spoken interaction. Even if I am proven wrong in the details of the analyses offered above or even in the broad theoretical outlines, I think I am still right in saying that a more empirically oriented linguistic methodology is going to run head on into the problem of the integration of speech and gesture and that there must be a place provided in linguistic theorizing for gestures in combination with speech.

## Note

### 1. Transcription conventions:

Vocal, verbal expressions in the left-hand column of the transcription within < > brackets are accompanied by non-vocal, non-verbal expression to be found directly to the right within matching < > brackets.

<words in transcription>	<description of non-vocal, non-verbal expression>
--------------------------	--

The cases in which non-vocal, non-verbal expressions occur without overlapping with vocal, verbal expressions is indicated as follows.

< <i>gesture</i> >	<description of non-vocal, non-verbal expression>
--------------------	--

Inaudible speech is indicated by dots enclosed in parentheses as follows.

## References

Goodwin, C. & Goodwin, M.H. (1992) 'Context, activity, and participation'. in P. Auer and A. di Luzio (eds.) *The Contextualization of Language*. Amsterdam: John Benjamins.

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

Feyereisen, P. & de Lannoy, J-D. (1991) *Gestures and speech: Psychological investigations*. Cambridge: Cambridge University Press.

Searle, J. (1969) *Speech Acts*. Cambridge: Cambridge University Press.

Allwood, J. (1976) *Linguistic Communication as Action and Cooperation*. Gothenburg Monographs in Linguistics 2. Dept. of Linguistics. University of Göteborg.

Hirsch, R. (1989) *Argumentation, Information, and Interaction*. Gothenburg Monographs in Linguistics 7. Dept. of Linguistics. University of Göteborg.

Naess, A. (1953) *Interpretation and Preciseness*. Oslo: Universitetsforlaget.

Peirce, C.S. (1955) 'Logic as Semiotic: The Theory of Signs' in Buchler, J. (ed.) *Philosophical Writings of Peirce*. New York: Dover.

Grice, H.P. (1957) 'Meaning'. *Philosophical Review*, 66, 377-388.

Pinkal, M. (1986) *Logik und Lexikon*. Berlin: de Gruyter.