Taking Fingerprints of Speech-and-Gesture Ensembles

Approaching Empirical Evidence of Intrapersonal Alignment in Multimodal Communication

Andy Lücking
CRC 673, B1

Alexander Mehler
CRC 673, A3, X1

Peter Menke*
CRC 673, X1

Bielefeld University, CRC 673 "Alignment in Communication"
{Andy.Luecking,Alexander.Mehler,Peter.Menke}@uni-bielefeld.de

Abstract

Co-occurring speech and gestures of natural language dialogues composes into meaning units, that is, they jointly describe discourse referents. We start from the idea that interlocutors tend to re-use this cross-modal information units if the discourse referent is referred to again: co-occurring speech and gesture are assumed to "align into" *bimodal ensembles* (BMEs). We further hypothesize that due to principles of dialogical economy interlocutors will exploit the impact of a BME's gesture to shorten its linguistic part of that BME. If this hypothesis is right, we expect that the words in multimodal communication exhibit a different frequency distribution from words in written texts, whose frequency distribution is known to obey Zipf's law. This hypothesis is tested for 24 direction-giving dialogues using two different frequency fits, rank frequency distribution and complementary cumulative distribution. According to the first fit, the hypothesis can be confirmed, according to the second one, it has to be rejected. In addition, we also propose a way to measure the strength of cross-modal informational association.

1 Introduction and Reasoning

This article presents some ideas about how to combine text-technological tools and linguistic research in the study of multi-modal dialogue, that is dialogue comprising speech and gesture. The term 'gesture' refers to gesticulations according to Kendon's continuum (Kendon, 1988), that is, 'gesture' is understood as a spontaneous co-verbal hand and arm movement which is linguistically significant and contributes to the narrative. McNeill (1992), alluding to a Peircean trichotomy, distinguishes different types of gesture, namely *deictic* gestures, *iconic* gestures, and *beats*. Beats are rhythmic stresses, deictic gestures are pointings. According to Peirce, icons are representations ("signs"), "whose relation to their objects is a mere community in some quality" (Peirce, 1867). That is, icons signify due to a certain resemblance between signifier and signified.

However, 'icon' is an "umbrella term" (cf. (Eco, 1976)) that covers a variety of different signifying methods. (Müller, 1998), drawing on the work of (Wundt, 1911), sets up a more fine-grained classification of gestures according to the distinction of four *modes of representation* on the ground of what the hands *do: Agieren* (Acting), *Modellieren* (Modelling), *Zeichnen* (Drawing), and *Repräsentieren* (Representing).

Ancient rhetoric already emphasizes the rhetoric connection between speech and gesture (Quintilian, 1st century; Maier-Eichhorn, 1989). In modern times, most notable Kendon (1980) claims that verbal utterances together with simultaneous accompanying deictic and iconic gestures coheres into single meaning units. However, it is not yet clear how the mechanism that binds together the two communication channels should be modelled – is it functional application (Rieser, 2004), rhetorical relation

*Authors' names are given in alphabetical order.

(Lücking et al., 2006; Lascarides and Stone, 2006) or something else? For the time being the pair of gesture and affiliated speech should be construed as an informational wholeness tied together by some kind of synchronicity principle (Jung, 1971). Take for instance an example from the study described in Section 2, where a subject is talking about one of two churches on a square which are, amongst others, distinguished by the type of their roofs:

(1) rechts die hat so'n [Giebel]
the one to the right it has such a [gable]
$\wedge$-*shaped gesture synchronous to bracketed speech*

The gesture from (1), which is displayed as Figure 1(a), is a *Posturing* gesture according to the modes of representation scheme introduced below. We assume that for the period of the dialogue the gesture gets associated with its accompanying speech,[1] or, as we will call it hereafter: The bracketed portion of the linguistic utterance together with the accompanying speech constitutes a *bimodal ensemble* (BME).

The linguistic part of a BME may comprise more than single words, as is illustrated in (2), where the subject talks about a chapel that is located within the "punch" of a surrounding "$\sqcup$"-shaped hedge, as indicated by a *Shaping* gesture (see Figure 1(b)).

(2) die hat ['ne grüne Hecke drumherum]
it has [a green hedgerow around it]
$\sqcup$-*shaped gesture synchronous to bracketed speech*

There is some discussion about the informational relation between speech and gesture: Is it redundancy or complementarity? (Cassell and Prevost, 1996; Bergmann and Kopp, 2006) We will, however, bypass this issue since our concern is purely quantitative: The linguistic part of BMEs is the input for the frequency distribution analysis given in Section 3.

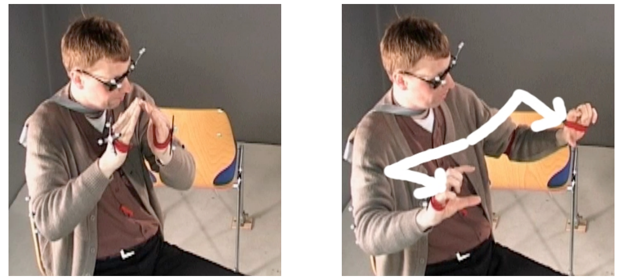
On a more abstract level, a BME is an assemblage comprising a set of parts of speech (classes of words) and a representation technique (class of gesture, e.g. *Shaping*). BMEs conceived this way enter into the determination of the Hartley information (Klir and Folger, 1988) (see Section 3).

The fusion of speech and gesture into a BME in dialogue is a precondition to the investigation pursued in this article. We investigate a **hypothesis concerning bimodal ensembles**: The use of gesture facilitates a merely partial recurrence or a paraphrase of the linguistic material of a BME. Since there is not yet data directed to and annotated for speech-and-gesture coupling over the time-course or a dialogue, we approach this issue by means of an indirect measuring. Think, for example, of an ensemble $e = (xy, g)$ manifested by some linguistic material xy in conjunction with a gesture g. The interlocutors of D may manifest e later on by the parts x, y of xy (or maybe even by some unit z which is sense-related to x, y or xy). The reason is that the simultaneously produced gesture g allows for correctly disambiguating the shortened or otherwise modified linguistic manifestation of the ensemble e.[2] For an illustration take the sample utterance (2). The BME (*'ne grüne Hecke drumherum*, $\sqcup$) might get shortened to:

(3) die hat ['ne Hecke]
it has [a hedgerow]
$\sqcup$-*shaped gesture*

To give an example for sense related substitution: The word *Giebel/gable* from (1) might be replaced by the hyponym *Pediment*:

(4) die hat so'n [Pediment]
it has such a [gable]



(a) Gable (b) Hedgerow

Figure 1: Two sample gestures.

[1]Most presumably, the association is established by some grounding mechanism (see for instance (Clark and Schaefer, 1989)), but we will not pursue this issue further here.

[2]It may also be the case that interlocutors reduce motor effort and produce simplified gestures. But this is a different story.

$\wedge$-*shaped gesture*

The described mechanisms leave an option to express the same concept in dialogical communication. Thus, any frequent usage of this method of reducing communication effort has an impact on the frequency distribution of lexical units within D: the same concept denoted by e is alternatively manifested by xy, x, y, z, z', z'', ... (Remember that z, z', z'' are sense related to x, y, or xy.) As this method of lexical choice is out of reach in written communication we expect an impact of using gestures on the frequency distributions of lexical units in dialogues.

Note that this argumentation presupposes that there is a usage-chain in dialogue D from the BME e to its shortening later on in D.

Note further that we do not expect this effect on the level of highly frequent words which, as expected, consist of function words and therefore rarely count as linguistic manifestations of bimodal ensembles.

The next section gives a brief overview of the study that underlies the data our investigation is based on. It also introduces the gestural representation techniques that enter into the determination of the Hartley information. The measuring procedures and its results are given in Section 3.

2 Experimental Study

Iconic gesturing is inherently spatial (Alibali, 2005). A kind of setting that has proved to elicit spatial discourse is the description of routes (Denis, 1997). Accordingly, the empirical data of our research consists of direction giving dialogues. The dialogues are about city tours one of the interlocutors has made in a town presented in a Virtual Reality environment (Kopp et al., 2008). Thus, our empirical study comprises two phases: At first, a participant undertakes a "bus ride" in a virtual town, see Figure 2(a) for an illustration. The sight-seeing tour passes five objects of interest, namely an abstract sculpture, a city hall, a church square with two churches, a chapel and a fountain. Subsequently, the first participant, called Router (R), has to explain to a second participant who does not know the virtual town which route he has driven and what landmarks he has seen. In order to elicit an elaborate spatial discourse the second participant, Follower (F), was made to believe that he will have to find the route through the virtual town and to identify all landmarks. Splitting up the virtual sight-seeing tour in a route and a landmark part, different types of spatial communication will come up, namely giving directions and describing shapes. Both are good candidates for iconic depiction.

In view of the frequency analysis to come, the employment of a virtual stimulus is a precondition for the inter-participant comparability of linguistic and non-verbal data, since it assures that all participants talk about the same thing.

2.1 Annotation

Annotation layers divide naturally into two different partitions, the one relating to speech the other relating to gestures. Speech transcription has been made using Praat[3] and has been done orthographically, i.e., on the level of words. Part of speech information is added automatically by means of POS-tagging (Gleim et al., 2007).

For gesture annotation, we delimit the gesture's semantic phase known as *stroke* (McNeill, 1992). Each stroke has been assigned a mode of representation. We have extended and terminologically modified Müller's set of representation modes in order to adjust it to the specific needs of route descriptions. Gesture has been annotated using the multimedia annotation software Elan[4]. The representation techniques we recognize are itemized and briefly commented upon in the following list.

Shaping The hands are sliding on the surface of a virtual object in gesture space, a shape emerges.

Sizing A configuration of hands or fingers that indicate a certain distance or size is called Sizing.

Posturing The hand (or both hands) represent an object involved in the described situation.

Drawing A single finger or the hand is used as a drawing tool to sketch an outline in the gesture space.

Pantomime The usage of an object or an action is displayed by imitation. Note that Pantomime,

[3]www.fon.hum.uva.nl/praat

[4]www.lat-mpi.eu/tools/elan

(a) Virtual bus ride (b) Route description (c) Ariadne system

Figure 2: Virtual environment stimulus and subsequent dialogue: the Data are managed in the Ariadne system.

in contrast to the other gesture practices, makes the gesturer himself a part of the depiction, not just his hands or arms.

Indexing A deictic gesture that singles out a point in the gesture space which thereby gets "semantically loaded", e.g., becomes a proxy for an object of the narrative.

Grasping If the hand touches or holds an object, but does not shape its body, then a Grasping-gesture is performed.

Counting If the fingers are used to enumerate things. Gestural counting can be seen as an iconic representation of a tally sheet.

Hedging Sometimes a wiggling or shrugging movement is used in order to depict uncertainty. We call this metaphoric gesture method 'Hedging'.

In sum, there are 25 direction-giving dyads with a total of 4961 gestures and 39.435 words.

Our multimodal dialogue data are stored, retrieved, transformed, and statistically explored within the Ariadne system (Gleim et al., 2007) which is used as an *Alignment Corpus Management System* (ACMS) – see the screen shot displayed as Figure 2(c).

2.2 Reliability

Since the classification of gestures in terms of representation modes is interpretive data, it is questionable whether it is reproducible (Krippendorff, 1980). Our evaluation of gesture classification data follows the discussion in (Stegmann and Lücking, 2005). A sample of gestures large enough to test for the reasonable agreement level of 70% with an α-error of 0.05 and a β-error of 0.85 (set in the run-up to the reliability study) has been classified by three expert annotators. The resulting first-order agreement coefficient AC_1 (Gwet, 2001) is 0.784. It's confidence interval is $(0.758, 0.81)$, so that the probability for agreement on gestures' representation modes – given that the agreement is not due to chance – is significantly greater than 75%.

3 Measuring Procedure and Results

Our starting point of indirectly measuring an impact of gestures on the choice of lexical units is Zipf's law (Zipf, 1972) which we denote as follows (Adamic, 2000):

$$n \sim r^{-\gamma} \tag{1}$$

n is the frequency of the rth most frequent word in the given text (or dialogue) for which Model (1) is fitted. Roughly, Zipf and related studies show that $\gamma \sim 1$ for written texts (Rapoport, 1982; Tuldava, 1998). Taking this as a reference value we expect – according to our hypothesis – a lower value of γ in the case of dialogical communication, that is, a flatter straight line which results from a log-log plot of the Rank Frequency Model (1). Note that $-\gamma$ is the slope of that line. Look, for example, at Figure 3(a), where we have fitted the power law $Cx^{-\gamma}$ to the *Rank Frequency Distribution* (RFD) of lexical units used by some interlocutor in a dialogue from our corpus. That is, the first rank is the one of the most frequent word, the second the one of the second most frequent word, and so on till we finally reach the ranks of *hapax legomena*. Fitting this empirical curve and plotting the result in a log-log plot we see that $\gamma = .678$ while the adjusted coefficient of determination $\bar{R}^2$ equals .9674. This indicates a good fit.[5] This result is in support of our hypothesis of a gesture-based impact on lexical choices – it does not falsify the hypothesis about the existence of this impact: as the exponent is smaller than one, the curve is flatter than suggested by the results derived from written texts.

However, according to (Newman, 2005) fittings change for the better by operating on the *Complementary Cumulative Distribution* (CCD), that is, on the probability function $P(X \geq x)$ of words which occur at least x times. In the case of our example the results of fitting to the CCD derived from the corresponding RFD are shown in Figure 3(b): Now, $\gamma = 1.145$ and $\bar{R}^2 = .9937$ what indicates a slightly better fit. *Can these two measurements be compared?* According to (Adamic, 2000) the exponent γ of a Zipfian RFD corresponding to a given CCD with exponent β is computed by $\gamma = 1/\beta$ – in the present case we achieve $\gamma \sim .873$.

That is, relying on the CCD which gives a better fit than the previously observed RFD and deriving the exponent of a RFD – which corresponds to the latter CCD on the same level of goodness of fitting – the absolute value of the exponent is raised ($.873 > .678$). This is what we actually observe in nearly all cases of our corpus of 24 dialogues.[6] The corresponding box plots of the 24 exponents γ and the corresponding determination coefficients $\bar{R}^2$ are shown in Figure 4: Not only are the absolute values of the exponents of the power laws fitted to the corresponding CCDs higher than the one of the primarily observed RFDs. More important is the observation of remarkably higher values of $\bar{R}^2$ – that is, as indicated by (Newman, 2005), CCDs are more reliable reference points of power law fitting. Thus, we additionally derive – according to the approach of (Adamic, 2000) – the exponents of those rank frequency distributions which correspond to the latter CCDs on the same level of goodness of fitting. As a result we see that we get on average higher values than in the case of the primarily observed RFDs (cf. Figure 4(c)). Moreover, the newly derived values disperse around 1 and are, therefore, in a good neighborhood of those values which were observed by Zipf. In this sense, our results do not indicate a difference between written and dialogical communication – at least under the regime of our experimental setting. Following this line of argumentation, there is *no* effect on the frequency distribution of lexical units. This hypothesis is only upheld by referring to the fittings of the left part of Figure 4 – however at the price of relying on worse fittings.

As our distribution analysis does not shed much light on the existence of bimodal ensembles we now compute a measure of interactivity between selections on the lexical and gestural layer. Such cross-modal selections are called interactive if, for example, the selection of lexical units constrains the selection of gestural units, that is, if there is a tendency of co-occurrence among lexical and gestural units. If we could measure such a tendency, this could be interpreted as a support of our hypothesis about the existence of bimodal ensembles. As we will see, this is not achieved.

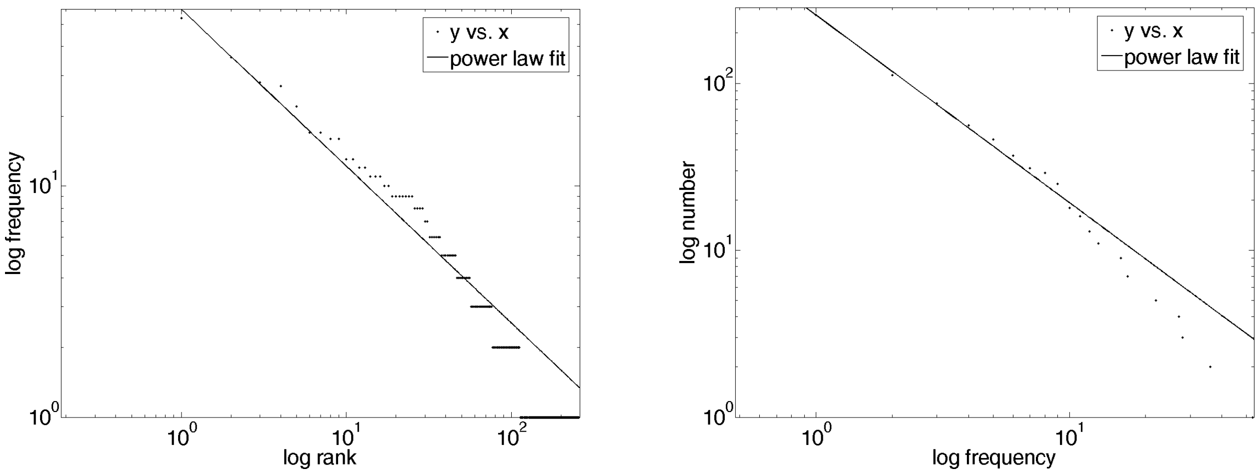
In order to get a first measure of the interaction of cross-modal selections we compute the information transmission between selecting from the set of parts of speech X and the set of representation techniques Y. Generally speaking, the information transmission between n sets $X_1, \ldots, X_n$ is defined as follows (Klir and Folger, 1988):

$$T(X_1, \ldots, X_n) = \sum_{i=1}^{n} I(X_i) - I(X_1, \ldots, X_n) \qquad (2)$$

where $I(X) = \log_2 |X|$ is the simple Hartley information of X (cf. (Klir and Folger, 1988) for the details of this and related definitions), $I(X,Y) = \log_2 |R|, R \subseteq X \times Y$, is the joint (Hartley) information. In our case R is the set of all co-articulated parts of speech and representation techniques: Generally speaking, the sets $X_1, \ldots, X_n$ are called *non-interactive* if $T(X_1, \ldots, X_n) = 0$, otherwise we observe that $T(X_1, \ldots, X_n) > 0$. Note that $T(X_1, \ldots, X_n) = 0$ if and only if $R = X_1 \times \ldots \times X_n$. In this case, any selection from set $X_i, i \in \{1, \ldots, n\}$,

[5] The adjusted coefficient of determination is a measure of goodness of fitting: the nearer its value to 1, the better the fit.

[6] Note that we deleted one dialogue from the corpus because of too many uncertain annotations.



(a) Results of fitting to the *Rank Frequency Distribution* (RFD – Zipfian scenario).

(b) Results of fitting to the *Complementary Cumulative Distribution* (CCD) derived from the latter RFD.

Figure 3: Two sample power law fittings of the frequency distribution of lexical units of a single interlocutor (in the role of the *router*). In both cases, the model $y = Cx^{-\gamma}$ is used.

may be combined with any selection from any other set $X_j, j \in \{1, \ldots, n\} \setminus \{i\}$. As the range of values of T is not limited, we standardize it as follows:

$$\hat{T}(X_1, \ldots, X_n) = \frac{T(X_1, \ldots, X_n)}{\sum_{i=1}^{n} I(X_i)} \in [0, 1] \tag{3}$$

Now, we see that for $\hat{T}(X_1, \ldots, X_n) \ll 1$ the sets $X_1, \ldots, X_n$ tend to be non-interactive, while they tend to be interactive if in contrast to this $\hat{T}(X_1, \ldots, X_n) \gg 0$. In other words: 0 indicates minimal and 1 maximal interactivity. In Figure 5 we report the results of measuring the interaction between the selection of parts of speech and of gestural practices by 24 interlocutors in 24 dialogues. Obviously, the sets are far from being interactive according to this measure of interactivity (which measures on an ordinal scale). However, as we do not yet know anything about expected values of such an interaction among elements of different modes in multimodal communication, we hesitate to value this as a falsification of our starting hypothesis. Anyhow, this hypothesis is not supported by both of our measurements, neither on the level of lexical distributions nor on the level of interactions of cross-modal choices. If we rely on the classical operation of rank frequency distribution, our hypothesis is not falsified. However, if we use the CCD we get a hint that there is no distributional difference.

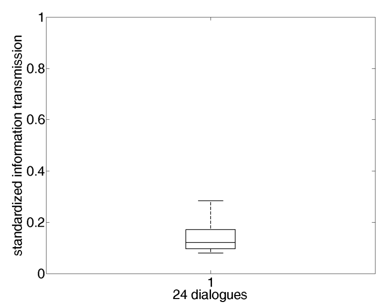
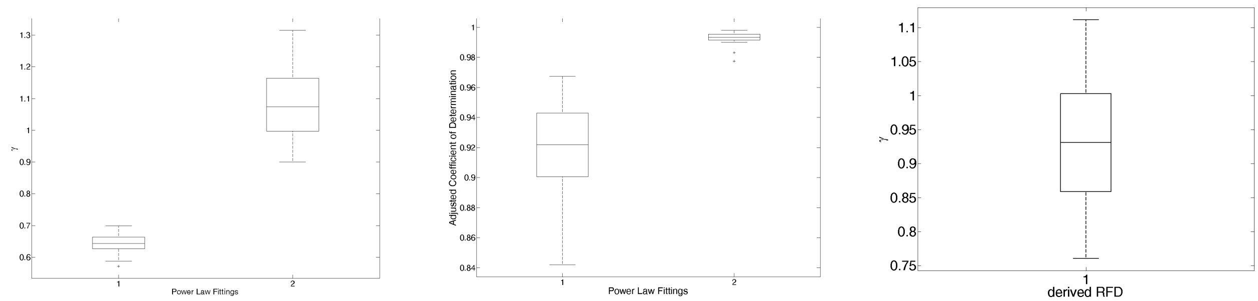


Figure 5: The distribution of information transmission between the selection of parts of speech and gestural practices by interlocutors in 24 dialogues.

4 Conclusion

One reason for the rather ambivalent result might be that its underlying presupposition does not hold. Ambivalent means that the rejection of the hypothesis depends on whether the fit is based on chosing the rank frequency distribution or the complementary cumulative distribution.

Recall from Section 1 that a BME *e* leaves a frequency distributional fingerprint only if there is a usage-chain connecting first occurrences of a fully specified *e* to subsequent shortened manifestations. The progressive rhematic structure of the direction-



(a) Exponent values

(b) Adjusted coefficients of determination

(c) Range of derived exponents of rank frequency distributions

Figure 4: Box plots of the exponent values (a), the corresponding adjusted coefficients of determination (b), and of the range of derived exponents of rank frequency distributions which correspond to the primarily observed complementary cumulative distributions (c) of all 24 dialogues of our corpus. The first column of both the (a) and the (b) sub-figures denotes the rank frequency model while the second column denotes the complementary cumulative model.

giving dialogues might block the establishment of a usage-chain for a certain BME *e*, leaving *e* an merely ephemeral phenomenon.[7]

As exposed in the preceding section, the rejection or affirmation of the hypothesis investigated in our analysis partly depends on "baseline values" for the different measuring procedures. Even if we cannot maintain our working hypothesis – and we have been very careful not to overstate our results, cf. Section 3 – analyses like the one carried out make up the pieces of the puzzle needed in order for a more comprehensive exploration of multimodal data. If BMEs indeed leave fingerprints that are measurable in the way explored in this article, this result clearly has an impact on cognitive theories, for instance theories of speech-and-gesture production. If there is an intra-personal alignment of words and gesture during a dialogue, the production of units on the respective modalities interacts. That is, empirical, quantitative research like the one presented here might help to collect evidence for or against different views of production processes as developed by (McNeill and Duncan, 2000; Kita and Özyürek, 2003; de Ruiter, 2000; Krauss et al., 2000).

Acknowledgements. Thanks to Hannes Rieser, Stefan Kopp, Kirsten Bergmann, and Florian Hahn for discussions and/or data annotation. We also want to thank the anonymous reviewers whose critique helped to clarify the ideas presented here. This research is partially supported by the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center 673 "Alignment in Communication".

[7]Note that the focus of our line of reasoning is a frequency *distribution*, *not* the *frequency* of (gesture accompanying) words. Thus our result does not contradict studies which prove the latter.

References

Lada A. Adamic. 2000. Zipf, power-law, Pareto – a ranking tutorial.

Martha W. Alibali. 2005. Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation*, 5:307–331.

Kirsten Bergmann and Stefan Kopp. 2006. Verbal or visual? how information is distributed across speech and gesture in spatial dialog. In David Schlangen and Raquel Fernández, editors, *brandial'06 – Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue*, pages 90–97, Potsdam.

Justin Cassell and Scott Prevost. 1996. Distribution of semantic features across speech and gesture by humans and computers. In *Proceedings of the Workshop on Integration of Gesture in Language and Speech*.

Herbert H Clark and Edward F. Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13:259–294.

Jan Peter de Ruiter. 2000. The production of gesture and speech. In David McNeill, editor, *Language and gesture*. Cambridge University Press.

Michel Denis. 1997. The description of routes: A cognitive appraoch to the production of spatial discourse. *Current Psychology of Cognition*, 16:409–458.

Umberto Eco. 1976. *A Theory of Semiotics*. Indiana University Press, Bloomington.

Rüdiger Gleim, Alexander Mehler, and Hans-Jürgen Eikmeyer. 2007. Representing and maintaining large corpora. In *Proceedings of the Corpus Linguistics 2007 Conference, Birmingham (UK)*.

Kilem Gwet. 2001. *Handbook of Inter-Rater Reliability*. STATAXIS Publishing Company, Gaithersburg, MD.

Carl Gustav Jung. 1971. Synchronizität als ein Prinzip akausaler Zusammenhänge. In *Die Dynamik des Unbewussten*, volume 8 of *Gesammelte Werke*, chapter XVIII, pages 475–577. Walter-Verlag, Olten and Freiburg im Breisgau. First published in *Naturerklärung und Psyche* (1952).

Adam Kendon. 1980. Gesticulation and speech: Two aspects of the process of utterance. In Mary Ritchie Key, editor, *The Relationship of Verbal and Nonverbal Communication*, pages 207–227. Mouton Publishers, The Hague.

Adam Kendon. 1988. How gestures can become like words. In F. Poyatos, editor, *Cross-cultural Perspectives in Non-Verbal Communication*, pages 131–141. Hogrefe, Toronto.

Sotaro Kita and Asli Özyürek. 2003. What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 48:16–32.

George J. Klir and Tina A. Folger. 1988. *Fuzzy Sets, Uncertainty, and Information*. Prentice Hall, Englewood.

Stefan Kopp, Kirsten Bergmann, and Ipke Wachsmuth. 2008. Multimodal communication from multimodal thinking—towards an integrated model of speech and gesture production. *International Journal of Semantic Computing*. To be published.

Robert M. Krauss, Yihsiu Chen, and Rebecca F. Gottesman. 2000. Lexical gestures and lexical access: A process model. In David McNeill, editor, *Language and gesture*. Cambridge University Press.

Klaus Krippendorff. 1980. *Content analysis*. SAGE Publications, Beverly Hills.

Alex Lascarides and Matthew Stone. 2006. Formal semantics of iconic gesture. In David Schlangen and Raquel Fernández, editors, *brandial'06 Proceedings*, pages 64–71, Potsdam.

Andy Lücking, Hannes Rieser, and Marc Staudacher. 2006. Multi-modal integration for gesture and speech. In David Schlangen and Raquel Fernández, editors, *brandial'06 Proceedings*, pages 106–113, Potsdam.

Ursula Maier-Eichhorn. 1989. *Die Gestikulation in Quintilians Rhetorik*. Peter Lang, Frankfurt am Main.

David McNeill and Sunsan Duncan. 2000. Growth points in thinking-for-speaking. In David McNeill, editor, *Language and gesture*. Cambridge University Press.

David McNeill. 1992. *Hand and Mind—What Gestures Reveal about Thought*. Chicago University Press, Chicago.

Cornelia Müller. 1998. *Redebegleitende Gesten. Kulturgeschichte – Theorie – Sprachvergleich*, volume 1 of *Körper – Kultur – Kommunikation*. Berlin Verlag, Berlin.

M. E. J. Newman. 2005. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46:323–351.

Charles Sanders Peirce. 1867. On a new list of categories. In *Proceedings of the American Academy of Arts and Sciences Series*, volume 7, pages 287–298.

Marcus Fabius Quintilian. 1st century. *Institutio Oratoria*.

Anatol Rapoport. 1982. Zipf's law re-visited. In H. Guiter and M. V. Arapov, editors, *Studies on Zipf's Law*, pages 1–28. Brockmeyer, Bochum.

Hannes Rieser. 2004. Pointing in dialogue. In Jonathan Ginzburg and Enric Vallduví, editors, *Catalog '04 Proceedings*, pages 93–100, Barcelona.

Jens Stegmann and Andy Lücking. 2005. Assessing reliability on annotations (1): Theoretical considerations. Technical Report 2, CRC 360, Universität Bielefeld.

Juhan Tuldava. 1998. *Probleme und Methoden der quantitativ-systemischen Lexikologie*. Wissenschaftlicher Verlag, Trier.

Wilhelm Wundt. 1911. *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythus und Sitte*, volume I: Die Sprache. Erster Teil. Wilhelm Engelmann, Leipzig.

George K. Zipf. 1972. *Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology*. Hafner Publishing Company, New York.