

Philosophical Aspects of Neural, Probabilistic and Fuzzy Modeling of Language Use and Translation

Timo Honkela

Abstract—Serious efforts to develop computerized systems for natural language understanding and machine translation have taken place for more than half a century. Some successful systems that translate texts in limited domains such as weather forecasts have been implemented. However, the more general the domain or complex the style of the text the more difficult it is to reach high quality translation. The same applies to natural language understanding. All systems need to deal with problems like ambiguity, lack of semantic coverage and pragmatic insight. In this article, some philosophical questions that underlie the difficulty of natural language understanding and good quality translation are first studied. These two areas of dealing with languages are actually closely related. Namely, for instance Quine’s notion of indeterminacy of translation have shown that the problem of translation does not only hold for translation between different languages but similar problems are encountered when communication between users of same language is considered. The term intralingual translation has been used e.g. by Roman Jakobson. Intralingual translation relates to translation between languages and to the problem of sameness of meaning. In this article, arguments and methods of considering translation and meaning within the framework of continuous-valued multidimensional representations, probability theory, fuzzy sets and neural adaptive systems are considered.

I. INTRODUCTION

In the following, a general introduction is provided to the philosophical issues related to the theme of the article.

A. Formal semantics

One underlying motivation of this article is the recognition of a potential need to increase the variety of methods that are used to deal with issues within philosophy of language. To express the basic situation in a slightly simplified manner, the formal methodological realm of philosophy of language is still largely dominated by predicate logic. Many philosophers, including Gottlob Frege, Alfred Tarski and Rudolf Carnap, have been more or less skeptical about formalizing natural languages, but many of them have relied on a certain level of formalization. Some of the prominent members of this tradition of formal semantics include Alfred Tarski, Rudolf Carnap, Richard Montague and Donald Davidson. Recent works on formal semantics have been conducted, for instance, by Jon Barwise, Robin Cooper and Barbara Partee [1]. The works on analytical philosophy of language tend to focus on some particular aspects of natural language such as truth conditions, and the role of quantifiers and connectives. Maybe the most striking example of formalization of natural

language is the work of Richard Montague. Montague’s thesis was that there is no essential difference between the semantics of natural languages (like English) and formal languages (like predicate logic), i.e., there is a rigorous way how to translate English sentences into an artificial logical language [2]. Montague grammar is an attempt to link directly the syntactic and semantic level of language. In order to do so, Montague defined the syntax of declarative sentences as tree structures and created an interpretation of those structures using an intensional logic. The end result was a focus on such aspects of language that nicely fit with the theoretical framework. Examples of language considered includes sentences like “Bill walks”, “every man walks”, “the man walks”, and “John finds an unicorn” [2]. It may be fair to say that most of the linguistic phenomena are set aside. Montague even assumes that the original sentences can be considered unambiguous even though ambiguity is a central phenomenon in language at many levels of abstraction. The idea of being rigorous may be considered a proper stand but it often leads to the negligence of the original complexity of the phenomenon being considered [3].

Many philosophers outside the analytical tradition have already for some time criticized the approach of logical formalization within philosophy of language. For instance, representatives of phenomenology (e.g. Edmund Husserl and Martin Heidegger), hermeneutics (e.g. Martin Heidegger and Hans-Georg Gadamer) and critical theory (e.g. Max Horkheimer and Jürgen Habermas) have presented alternative views. Richard Rorty [4] attacks the correspondence theory of truth (that truth is established by directly comparing what a sentence asserts to the “facts applying), and even denies that there are any ultimate foundations for knowledge at all. He calls for a socially-based theory of understanding. He also strongly criticizes the notion of truth: Truth is not a common property of true statements, and the good is what proves itself to be so in practice. Rorty combines pragmatism (cf. e.g. John Dewey and Charles S. Peirce) with the philosophy of language by later Wittgenstein which declares that meaning is a social-linguistic product.

B. From logical to statistical formalization

The distinction between translation within and translation between languages is made here to emphasize two matters. First, there is a strong connection between translation and understanding. Second, it is far from obvious that communication between speakers of one and same language would be based on commonly shared meanings as often suggested by the proponents of formal semantics, either

T. Honkela is with the Adaptive Informatics Research Centre (AIRC), Helsinki University of Technology, P.O. Box 5400, FI-02015 TKK, Espoo, Finland (email: timo.honkela@tkk.fi)

explicitly or implicitly. This leads to the rejection of the idea of an idealized language user and to the rejection of the possibility to consider central epistemological questions and natural language semantics without considering subjectivity and variability.

Rather than using first-order predicate logic, modal logic and other similar formal languages as a basis for theory formation within epistemology, it is strongly suggested that they might even be mostly replaced by probability theory, matrix algebra, dynamical systems theory and other statistical and mathematical methods that seem to be better suited for building epistemological theories in order to be able to deal with continuous, multidimensional and dynamical phenomena that are inherent in knowledge formation and natural language understanding.

Notions such as a symbol or a proposition may still be useful in some theoretical contexts but they may rather be seen as abstractions that are emergent outcomes of some highly complex processes. However, even in the context of philosophy of language, it might be less misleading to use terms 'word' and 'sentence' rather than 'symbol' and 'proposition' and discuss the emergence of symbol-like and proposition-like phenomena through theoretical tools that are suited to capture the nature of those emergent processes. Following the criticism by Richard Rorty towards the notion of truth, discussed earlier in this article, one may really question the usefulness of the notions of truth as a useful building block in realistic epistemological theories. Truthlikeness already seems to capture better the nature of 'sentential knowledge' but even that term might draw the attention away from the idea of language being primarily a tool for communication of various kinds. Moreover, human understanding of the world and of the relationship between language use and perception and action within the world is based on a long learning process for which the genotype gives a certain basis but which is mainly determined by the individual interaction with the world including other human beings and the social and cultural context. In general, the centrality of learning processes emphasizes the need to consider the statistical aspects related to learning and using language. Language learning seems to be essentially a statistical process. There are some researchers such as Jerry Fodor who suggest that linguistic skills and even conceptual contents are innate in the mind. However, it seems that the arguments supporting the centrality of learning proposed, e.g. by Patricia Churchland, Paul Churchland, Andy Clark and Paul Smolensky are more realistic. This line of thought leads directly to the use of neural network models in modeling processes of language learning, understanding and generation. In language use, many specific situations can be readily analyzed as probabilistic questions.

The statistical point of view also seems to be coherent with radical constructivism. Mickle [5] presents an interesting observation that radical constructivism in philosophy of science forces reconsider the status of statistical data. Radical constructivism argues that reality is not given and

explored but rather invented or mentally constructed. Mickle even considers theories as fictitious constructs. He also states, however, that there is no reason why useful fictitious constructs should not be analyzed by any means or methods, including theories of measurement, data, probability and statistics.

II. TRANSLATION THROUGH SELF-ORGANIZED CONCEPTUAL SPACES

In the following, translation or mapping between two language through conceptual spaces or using the self-organizing maps is considered. Construction of maps of words based on the self-organizing map algorithm is presented as an introductory theme.

A. Self-Organizing Map

The Self-Organizing Map (SOM) [6], [7] defines an ordered mapping, a kind of projection from a set of given data items onto a regular, usually two-dimensional grid. A model m_i is associated with each grid node. These models are computed by the SOM algorithm. A data item will be mapped into the node whose model is most similar to the data item, i.e., has the smallest distance from the data item in some metric. The model is then usually a certain weighted local average of the given data items in the data space. But in addition to that, when the models are computed by the SOM algorithm, they are more similar at the nearby nodes than between nodes located farther away from each other on the grid. In this way the set of the models can be regarded to constitute a similarity graph, and structured 'skeleton' of the distribution of the given data items.[8]

B. Maps of Words

Charniak [9] presents the following scheme for grouping or clustering words into classes that reflect the commonality of some property.

- Define the properties that are taken into account and can be given a numerical value.
- Create a vector of length n with n numerical values for each item to be classified.
- Cluster the points that are near each other in the n -dimensional space.

Handling computerized form of written language rests on processing of discrete symbols. How can a symbolic input such as a word be given to a numeric algorithm? One useful numerical representation can be obtained by taking into account the sentential context in which the words occur. Before utilization of the context information, however, the numerical value of the code should not imply any order to the words. Therefore, it will be necessary to use uncorrelated vectors for encoding. The simplest method to introduce uncorrelated codes is to assign a unit vector for each word. When all different word forms in the input material are listed, a code vector can be defined to have as many components as there are word forms in the list. This method, however, is only practical in small experiments. With a vocabulary picked from a even reasonably large corpus the dimensionality of

the vectors would become intolerably high. If the vocabulary is large, the word forms can be encoded by quasi-orthogonal random vectors of a much smaller dimensionality [10]. Such random vectors can still be considered to be sufficiently dissimilar mutually and not to convey any information about the meaning of the words. Mathematical analysis of the random encoding of the word vectors is presented in [11]. A typical result of applying the Self-Organizing Map in the analysis of words is presented in Fig. 1 (based on the analysis reported in [12]).

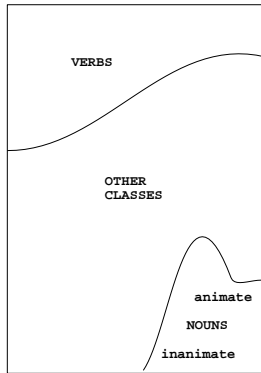


Fig. 1. A typical structure of a map of words.

C. Interlingual Mappings

To illustrate the idea of using the Self-Organizing Map in finding a mapping between vocabularies of two different languages, the results of a new experiment are reported in the following. Like discussed in earlier sections, the maps of words are often constructed using the sentential contexts of words as input data. The result is that the more similar the contexts in which two words appear in the text, the closer the words tend to be on the map. Here this basic idea is extended to cover the notion of context in general: We consider the use of a collection of words in two languages, English and German, in a number of contexts. In this experiment, the contexts were real-life situations rather than some textual contexts.

Figure 2 presents the order of a number of words on a self-organizing map that serves simultaneously two purposes. First, it has organized different contexts to create a conceptual landscape (see, e.g., [13]). Second, the map includes a mapping between the English and German words used in the analysis. The input for the map consists of words and their contexts. The German vocabulary includes 33 words (Advokat, Angler, Arzt, Autofahrer, ..., Zahnarzt) and the English vocabulary of 17 words (boss, dancer, dentist, director, ..., professor). For each word, there is an assessment by 10 to 27 subjects indicating the degree of suitability for the word to be used in a particular context. The number of contexts is 19. The resulting map is shown in Fig. 2.

The map shows that those words in the two languages that have similar meaning are close to each other on the map. In this particular experiment, the German subjects were usually

using a larger vocabulary. Therefore, in many areas of the map, a particular conceptual area is covered by one English word (for instance, “doctor” or “hairstresser”) and by two or more German words (for instance, “Arzt” and “Doktor” or “Friseur”, “Friseur” and “Damenfriseur”). It is important to notice that the model covers both translation between languages and within languages. Namely, rather than dealing with German and English, the same model can be built for the language used in medical contexts by experts and laypersons.

The results reported above are in an interesting contrast with another study in which words of two languages were presented in linguistic contexts [14]. Li and Farkas found out that the two languages were strictly separated on the map. It seems that the different ways of representing context explain these differences.

D. Semantic Holism and Self-Organizing Maps

Quine [15] presents a situation in which one is confronted with a situation in which one must attempt to make sense of the utterances and gestures that the members of a previously unknown tribe make. Quine claimed that it is impossible, in such a situation, to be absolutely certain of the meaning that a speaker of the tribe’s language attaches to an utterance. For example, if a speaker sees a rabbit and says “gavagai”, is she referring to the whole rabbit, to a specific part of the rabbit, or to a temporal aspect related to the rabbit. If one considers the point of view of radical constructivism [16], [3] and the symbol grounding problem [17], there can practically even be an infinite number of conceptualizations of the situation. Maybe the members of the tribe not only consider the whole rabbit or some parts or aspects of it as potentially relevant points of reference but, e.g., due to their cultural context they consider some other patterns of perception. Namely, considering the complex pattern recognition process, it is far from trivial to create a perception of a rabbit from the raw visual and auditory input. Quine [15] mentions that one can form manuals of translation. The observer examines the utterances as parts of the overall linguistic behavior of the individual, and then uses these observations to interpret the meaning of all other utterances. Quine continues that there will be many such manuals of translation since the reference relationship is indeterminate. He allows that simplicity considerations not only can be used to choose between competing manuals of translation but that there is even a remote possibility of getting rid of all but one manual. It seems that propositional logic as the underlying epistemological framework unnecessarily complicates the consideration. For Quine it was necessary to consider a number of logically distinct manual of translation hypotheses. However, if one considers the issue within the framework of statistics, probability theory and continuous multidimensional representations of knowledge (such as conceptual spaces [13]), one can consider the conditional probability of different hypotheses and partial solutions which do not need to be logically coherent. Moreover, the search for translation mappings can be seen as a process that may (or may not) converge over time. For Quine meaning is not something that is associated with a single word or

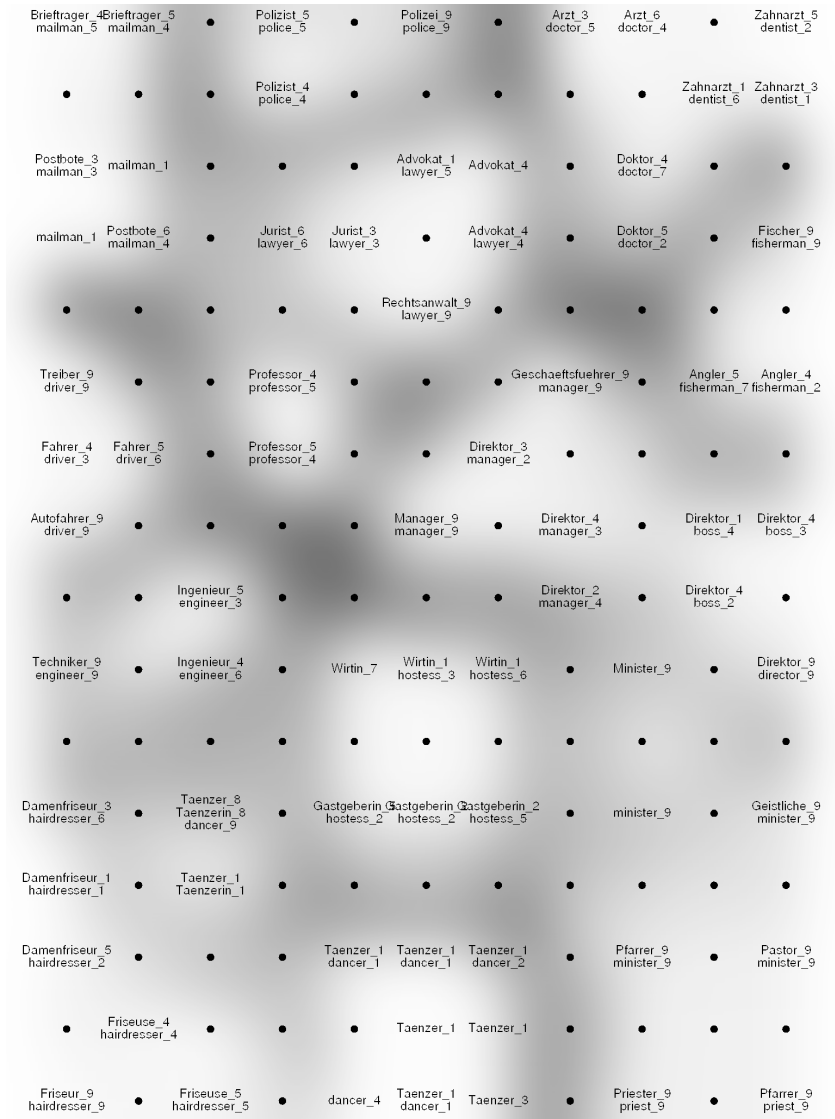


Fig. 2. A collection of German and English words positioned a conceptual landscape based on a Self-Organizing Map of contexts. The darker the shade of gray, the longer are the distances in the original input space. Thus, relatively light areas correspond to conceptual areas or clusters. The dots on the map denote empty prototypes, i.e., model vectors that are not the best match of any of the words under consideration.

sentence, but is rather something that can only be attributed to a whole language. The resulting view is called semantic holism. In a similar fashion, the self-organizing map specifies a holistic conceptual space. The meaning of a word is not based on some definition but is the emergent result of a number of encounters in which a word is perceived or used in some context. Moreover, the emergent prototypes on the map are not isolated instances but they influence each other in the adaptive formation process.

E. Comparison with Bayesian Models in Translation

Bayes' rule tells that

$$p(A|B) = p(B|A)p(A)/p(B), \quad (1)$$

where $P(A)$ is the prior probability or marginal probability of A , and $P(A|B)$ is the conditional probability of A , given

B . Let's consider a situation in which we wish to translate Finnish sentences, f , into English sentences, e . Bayes' rule gives $p(e|f)p(f) = p(e, f) = p(f|e)p(e)$ and reduces to the basic equation of statistical machine translation: maximize $p(e|f) = p(f|e)p(e)$ over the appropriate e . This splits the translation problem into a translation model ($p(f|e)$) and a language model ($p(e)$). The decoding algorithm, given these models and a new sentence f , finds translation e .

In their classical paper, Peter Brown and his colleagues described a series of five statistical models of the translation process and gave algorithms for estimating the parameters of these models applying the basic equation [18]. During recent years, this statistical approach has had considerable successes, based on the availability of large parallel corpora and some further methodological developments (consider, e.g., [19], [20]).

The main difference between the approach outlined in the previous sections and the Bayesian method, in its commonly used form, is that the semantic or conceptual space is explicitly modeled in the SOM-based approach. Thus, the mapping between any two languages is based on an intermediate level of representation. This approach resembles, to some degree, the idea of using a knowledge-based interlingua in machine translation. The underlying philosophical assumptions about knowledge are, however, quite different. In a knowledge-based interlingua, the semantics of natural language expressions are typically represented as propositions and relations in symbolic hierarchical structures. The SOM can be used to span a continuous and multidimensional conceptual space in a data-driven manner. Moreover, the approach provides a natural means to deal with multimodal data [21] and, thus, deal with the symbol grounding problem [17].

III. MODELING LANGUAGE USE

In this section, modeling language use based on fuzzy set theory and Bayesian probability theory is considered. As an introduction to the theme, subjectivity of understanding is discussed.

A. Subjectivity of Understanding

The language of a person is idiosyncratic and based on the subjective experiences of the individual. For instance, two persons may have a different conceptual or terminological density of the topic under consideration. A layperson, for instance, is likely to describe a phenomenon in general terms whereas an expert uses more specific terms. Moore and Carling [22] state that languages are in some respect like maps. If each of us sees the world from our particular perspective, then an individual's language is, in a sense, like a map of their world. Trying to understand another person is like trying to read the map of the other, a map of the world from another perspective [22].

If some persons speak the same language, many symbols in their vocabularies are the same. However, as discussed above, one cannot assume that the vocabularies of any two agents are exactly the same.

We take a broad view of learning, including any adaptation of internal representations that may occur as a result of interactions with other agents or the environment. As Kirsh [23] points out, in ecological systems each component of the system has a causal influence on the other. In the biological world organisms interact with their environment and with other organisms, who, of course, also tend to be part of each other's environment, the whole system of components being interdependent and interlocked. The result is a highly complex system displaying attractors, instabilities and cycles typical of dynamical systems [23].

B. Bayesian Models of Language Learning and Concept Formation

Joshua Tenenbaum [24] discusses the inductive concept formation problem presented by Quine [15]. Tenenbaum's

"Strong Bayes" model, presented as a solution to the concept induction problem, is based on the following ingredients:

- a constrained hypothesis space of possible extensions of a concept,
- a prior distribution over the hypothesis space reflecting the learner's relevant background knowledge,
- the size principle for scoring the likelihood of hypothesis, favoring smaller consistent hypotheses, and
- hypothesis averaging: integrating the predictions of multiple consistent hypotheses [24].

Tenenbaum aims to explain how it is possible to learn and generalize from just a few positive examples. Later, a related Bayesian approach has been applied to word learning [25].

Mike Dowman has considered the evolution of color terms [26]. In his simulation, ten agents could learn color term denotations by generalizing from examples using Bayesian inference. Conversations between the agents were simulated over several generations, and the languages emerging at the end of each simulation were investigated. The results corresponded closely with the observed typological patterns in basic color terms in human languages around the world [26].

C. Fuzzy Set Models of Language Use

Lotfi Zadeh has presented arguments for the idea that the brain has a crucial ability to manipulate perceptions [27]. Manipulation of perceptions plays a key role in human recognition, decision and execution processes. Zadeh also points out that measurements are crisp whereas perceptions are fuzzy. In general, the relationship between perceptions and their linguistic descriptions is not as straightforward as often suggested, for instance, by many logicians. It is this fundamental difference that makes it necessary to use fuzzy sets and fuzzy systems. Jonathan Evans points out that reasoning is highly contextualized by relevant prior knowledge and belief [28]. He also refers to the dual process theories of reasoning that make a division between a heuristic system and an analytical system. The heuristic system has evolved early, it is shared with animals, it is rapid and parallel, has high capacity and is pragmatic. According to Evans, the analytic system is conscious and it has evolved late in the evolution. It seems that many theories of language use consider only the level of the analytic system and fail to deal with the effects of the heuristic system.

Perhaps the most widely used mathematical method of modeling gradience in language use is fuzzy logic and fuzzy set theory [29]. Regarding a sentence like "John is tall", the 'truth' of this sentence needs to be considered a matter of degree. It is not possible to state a limit of tallness above which John would definitely be tall and under which not. In fuzzy logic, the truth of a proposition is expressed by a value between 0 and 1 rather than being just 'true' or 'false'. Accordingly, an item may belong to a set in a fuzzy manner. The degree of membership of a set is also specified by a graded value from the real interval between 0 and 1. Often, however, a single dimension such as height is not enough: to

state whether a person is tall one must also know her or his sex, age, genetic background, etc. The degree of membership becomes, thus, a surface in a multi-dimensional space rather than a representation based on one single variable.

A representation of linguistic vagueness or gradience as multidimensional fuzzy membership functions has been outlined above. Here one can come back to the issue of subjectivity of language use and understanding: How can subjectivity be represented within this framework? One can hypothesize that the interpretation of each expression by an individual in a specific context could be represented as a degree of membership, based on each person's unique system of membership functions. Intersubjectivity becomes a matter a degree in which the degree of shared meaning depends on the similarity or dissimilarity of the membership function systems between the individuals. The degree of communication success can be increased by iterated language learning within a community of individuals with shared contexts. In a related simulation study, a collection of simulated agents learned to use a shared emerging language [30]. The cognitive representation was based on the Self-Organizing Map. The distributions within the individual maps of the agents can be loosely considered as fuzzy membership distributions.

IV. CONCLUSIONS AND DISCUSSION

Various aspects of modeling translation and language use have been considered in this paper. Main methodological emphasis has been on the Self-Organizing Map that was used to illustrate the formation of a terminological mapping between two languages through an emergent conceptual space. Bayesian inference and fuzzy sets have also been discussed. Even though the proponents of different methodological approaches often consider critically other approaches, it seems that these three methods can be considered as complementary frameworks.

ACKNOWLEDGMENTS

This work was supported by the Academy of Finland through the Finnish Centre of Excellence Programme. Knowing that many important contributions are not mentioned, the author wishes to thank Tarja Knuutila for pointing out Richard Rorty's work and Anna-Mari Rusanen for many useful comments on an earlier version of the paper, many of which still remained unimplemented. Last but not least, the constructive comments by the reviewers are gratefully acknowledged, any remaining errors and problems are the author's own.

REFERENCES

[1] P. Portner and B. H. Partee, *Formal Semantics: The Essential Readings*. Oxford: Blackwell, 2002.

[2] R. Montague, "The proper treatment of quantification in ordinary english," in *Approaches to Natural Language: Proceedings of the 1970 Stanford Workshop on Grammar and Semantics*, J. Hintikka, J. Moravcsik, and P. Suppes, Eds. Dordrecht: D. Reidel, 1973, pp. 221–242.

[3] H. von Foerster, *Understanding Understanding*. New York: Springer-Verlag, 2003.

[4] R. Rorty, *Philosophy and the Mirror of Nature*. Princeton, NJ: Princeton University Press, 1979.

[5] H. C. Micko, "Statistical data: (empirical) facts or (theoretical) fictions," in *International Statistical Institute, ISI 54th session*, Berlin, Germany, 2003.

[6] T. Kohonen, "Self-organizing formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.

[7] —, *Self-Organizing Maps*, ser. Springer Series in Information Sciences. Berlin, Heidelberg: Springer, 2001, vol. 30.

[8] T. Kohonen and T. Honkela, "Kohonen network," *Scholarpedia*, p. 7421, 2007.

[9] E. Charniak, *Statistical Language Learning*. Cambridge, Massachusetts: MIT Press, 1993.

[10] H. Ritter and T. Kohonen, "Self-organizing semantic maps," *Biological Cybernetics*, vol. 61, no. 4, pp. 241–254, 1989.

[11] S. Kaski, "Dimensionality reduction by random mapping: Fast similarity computation for clustering," in *Proceedings of IJCNN'98, International Joint Conference on Neural Networks*. Piscataway, NJ: IEEE Service Center, 1998, vol. 1, pp. 413–418.

[12] T. Honkela, V. Pulkki, and T. Kohonen, "Contextual relations of words in Grimm tales, analyzed by self-organizing map," in *Proc. of ICANN'95, International Conference on Artificial Neural Networks*, F. Fogelman-Soulié and P. Gallinari, Eds., vol. II. Nanterre, France: EC2, 1995, pp. 3–7. [Online]. Available: citeseer.ist.psu.edu/honkela95contextual.html

[13] P. Gärdenfors, *Conceptual Spaces*. MIT Press, 2000.

[14] P. Li and I. Farkas, "Early lexical development in a self-organizing neural network," in *Bilingual sentence processing*. North-Holland, 2002, pp. 59–85.

[15] W. Quine, *Word and Object*. MIT Press, 1960.

[16] E. von Glasersfeld, *Radical Constructivism. A Way of Knowing and Learning*. London: Falmer Press, 1995.

[17] S. Harnad, "The symbol grounding problem," *Physica D*, vol. 42, pp. 335–346, 1990.

[18] P. F. Brown, S. A. D. Pietra, V. J. D. Pietra, and R. L. Mercer, "The mathematics of statistical machine translation: Parameter estimation," *Computational Linguistics*, vol. 19(2), pp. 263–311, 1993.

[19] P. Koehn, F. J. Och, and D. Marcu, "Statistical phrase-based translation," in *NAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*. Morristown, NJ, USA: Association for Computational Linguistics, 2003, pp. 48–54.

[20] R. Zhang, H. Yamamoto, M. Paul, H. Okuma, K. Yasuda, Y. Lepage, E. Denoual, D. Mochihashi, A. Finch, and E. Sumita, "The NiCT-ATR Statistical Machine Translation System for IWSLT 2006," in *Proceedings of the International Workshop on Spoken Language Translation*, Kyoto, Japan, 2006, pp. 83–90.

[21] J. Laaksonen and V. Viitaniemi, "Emergence of ontological relations from visual data with self-organizing maps," in *Proceedings of SCAI 2006, Scandinavian Conference on Artificial Intelligence*, Espoo, Finland, 2006, pp. 31–38.

[22] T. Moore and C. Carling, *The Limitations of Language*. Houndmills: Macmillan Press, 1988.

[23] D. Kirsh, "The context of work," *Human-Computer Interaction*, vol. 16, no. 2, 3 & 4, pp. 305–322, 2001.

[24] J. Tenenbaum, "A bayesian framework for concept learning," Ph.D. dissertation, Massachusetts Institute of Technology, 1999.

[25] J. Tenenbaum and F. Xu, "Word learning as bayesian inference," in *Proceedings of 22nd Annual Conference of Cognitive Science Society*, 2000.

[26] M. Dowman, "Explaining color term typology as the product of cultural evolution using a bayesian multi-agent model," in *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, R. Alterman and D. Kirsh, Eds. Mahwah, N.J.: Lawrence Erlbaum Associates, 2003.

[27] L. A. Zadeh, "From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions," *J. Appl. Math. Comput. Sci.*, vol. 12(3), pp. 307–324, 2002.

[28] J. S. Evans, "In two minds: dual process accounts of reasoning," *Trends in Cognitive Sciences*, vol. 7, pp. 454–459, 2003.

[29] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.

[30] T. Lindh-Knuutila, T. Honkela, and K. Lagus, "Simulating meaning negotiation using observational language games," in *Symbol Grounding and Beyond: Proceedings of the Third International Workshop on the Emergence and Evolution of Linguistic Communication*, P. V. et al., Ed. Rome, Italy: Springer, Berlin/Heidelberg, 2006, pp. 168–179.