Multimodally Grounded Translation. Session 7 - Directions for Research

Timo Honkela

TRALOGY

Multimodally Grounded Translation

Timo Honkela
Aalto University School of Science
timo.honkela@aalto.fi

TRALOGY II - Session 7
Date d'intervention : 18/01/2013

In most cases computer systems processing symbols or language do not have access to the phenomena being referred to. In contrast, human beings can readily associate expressions with their non-linguistic experiences. As a direct consequence, the present day computational systems can only reason about the symbols themselves rather than about the inherent meaning or external references of those symbols. This problem can be addressed in various ways. A traditional solution is to formalize the domain, such that the symbols used are defined in relation to each other to allow algorithmic analysis. However, multimodal contexts are more naturally represented as patterns and signals which differ considerably from the discrete representation of symbols and expressions of symbolic languages. This finding has some important consequences on how theories of linguistic semantics and pragmatics should be formulated, and what is the methodology that is needed in the computational modeling of high quality language processing and translation. In this paper, contextuality, multimodality and subjectivity are discussed as foundations for high quality translation and future scenarios for language technology and machine translation are provided with suggestions on how to divide translation tasks between human translators and machine translation systems. Special emphasis is given on underlying philosophical assumptions that may often have decisive role in determining whether a technological solution is functional and effective or not.

lien video : http://webcast.in2p3.fr/videos-multimodality_grounded_translation_by_humans_and_machines

1. Introduction

Careful cross-linguistic research conducted, e.g., by Bowerman and her colleagues has shown that languages differ strikingly in their semantic structuring of even such basic conceptual domains as space (Choi & Bowerman 1991). They have shown that there is a complex interaction between the influence of the input language and learners' non-linguistic conceptual predispositions. In order to model computationally the meaning of spatial relations and the information in other similar domains including human body parts, the symbols (words, expressions) need to be grounded in multimodal information. In other words, it is not sufficient to model mappings between pairs of words or phrases in different languages when their way of dividing the conceptual space is fundamentally incompatible. The problems caused by this fact concern, of course, both human and computerized translators.

In cognitive linguistics it is argued that the meaning of linguistic symbols are representations in the mind of the language users that derive from the users' sensory perceptions, their actions with the world and with each other. For example, the meaning of the word 'walk' involves what walking looks like, what it feels like to walk and after having walked, how the world looks when walking. In certain domains it is possible to measure information that can be used to form a model of the phenomenon similar to the one that is hypothesized to exist in the human brain. In particular, it appears that the representation of space, of movement in space and actions among objects in space underlies much of linguistic cognition, and therefore affects the meanings of words and expressions, and the way we generate and understand language (Lakoff & Johnson 1991, Regier 1996).

2. Lessons learned in building a "language machine"

In the 1980s, a lot of optimism was around knowledge-based systems, whether rule-based expert systems (Buchanan & Shortliffe 1989) or natural language processing systems in which explicit linguistic rules are encoded manually. A large-scale project called CYC aimed at encoding an encyclopedic ontological structure that would encompass all main human knowledge (Lenat & Guha 1989). In the area of machine translation, Eurotra was a major effort that followed methodologically a similar path (Johnson et al. 1985). Even though the projects did not meet the expectations due to philosophical and practical reasons that are also discussed in this paper, there are even nowadays major investments into projects that have similar objectives and a closely related methodological basis. In order to promote learning from earlier experiences, a case study which has many similar characteristics is reviewed in the following (see also Honkela 2005).

In 1980s, a large project was developing a natural language database interface for Finnish. Methodologically the project (Kielikone, ``Language Machine'', funded by Sitra foundation) was following the traditional artificial intelligence approach (Honkela et al. 1988, Jäppinen & Ylilammi, 1986, Jäppinen et al., 1988, Lehtola et al. 1988). Systems for natural language processing as well as expert systems were commonly based on a collection of rules or some other symbolic representations. The architecture of this particular system was ambitiously designed to cover the levels of written language: lexicon, morphology, syntax, semantics, and even pragmatics. Modeling of the structural level of Finnish language, i.e, morphology and syntax, was successful (Jäppinen and Ylilammi, 1986, Valkonen et al., 1987). Finnish has complex morphology which can be exemplified by the fact that each verb may have about 12,000 inflectional forms (Karlsson, 1999). The rule-based system for morphological analysis reached a level comparable to the skill of native speaker of Finnish (Jäppinen and Ylilammi, 1986). As a parallel development, Koskenniemi (1983) deviced another high-quality system with a general two-level representation that could be easily applied to a large number of different languages.

In Kielikone the syntactic analyzer was based on the dependency grammar. The dependency grammar fits well with the quality of Finnish language in which the word order is relatively free. A natural language database interface is supposed to analyze the meaning in addition to the structure of the input sentences. When associated with a stock exchange data base, the system should be able to respond properly to expressions such as "Show me the largest companies in forestry!", or "How many companies have turnover larger than 1 billion euros?". Some users would also ask questions that require analysis of imprecise or fuzzy expressions like "What are the companies with a large turnover and a small number of employees", or even prediction: "Which companies will be profitable next year?". The user expectations grow higher if the system is coined with the term intelligent. In our project, the semantic analysis was conducted using a rule-based system that was transfering the results of the syntactic analysis, i.e., dependency trees into predication structures which were variants of expressions in predicate logic. All the basic assumptions of predicate logic as representation formalism were present.

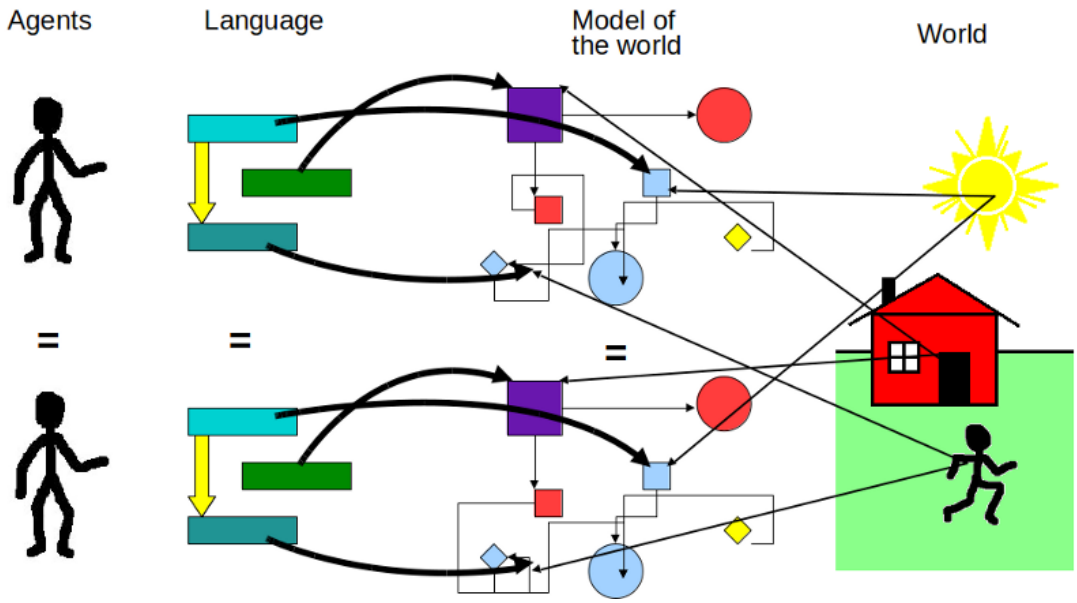

Figure 1. Logic-based view on the relationship between language, knowledge and the world.

One of the central ideas was that the world, or domain of interest, could be adequately modeled as a collection of objects and relationships between them (see Honkela 2005 for further discussion). This widely applied basic assumption is illustrated in Fig. 1. Well known proponents of this kind of view include early Wittgenstein (1922), Montague (1973) and Fodor (1975). This line of thinking was commonly held in the symbolic AI research in the 1980s and is still widely acknowledged, e.g., among those who develop and build upon semantic web technologies (Berners-Lee et al. 2001, Niles & Pease 2001, Noy 2004, Mellish & Sun 2006). This is the situation even though researchers such as Gärdenfors (2004) have given convincing arguments for the fact that the semantic web should not even be called semantic.

The efforts in developing the natural language interface in the Kielikone project as well as in many others before and after that one have shown that the manual coding of knowledge sufficient for the in-depth understanding of variety of questions and commands is very difficult. Considerable success has been gained only in systems that have limited domain of application. A classic example is SHRDLU system (Winograd 1972). SHRDLU was able to process commands related to a collection of items on the screen with varying sizes, colors and shapes.

The problems encountered in developing a natural-language database interface led into a re-evaluation of the status of the traditional artificial intelligence methodology and, specifically,

that of natural language processing. Many traditional philosophical underpinnings had also to be questioned (Honkela 2007). Both quantitative and qualitative issues can be considered. Firstly, the amount of knowledge needed in a successful large-scale natural language processing application is vast. Secondly, it seems that knowledge cannot be adequately modeled only by means of symbolic representation formalisms based on predicate logic or related formalisms (Honkela 1997).

3. From logical to statistical formalization

The distinction between translation within and translation between languages can made 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 explicitly or implicitly.

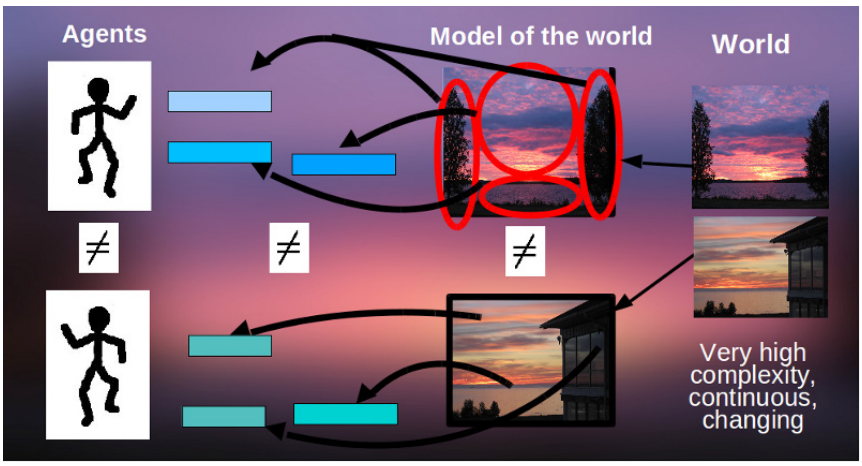

Figure 2. A dynamic view on the relationship between language, knowledge and the world.

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 (Honkela 2007). In essence, language and its use is a dynamic and statistical phenomenon. This point of view into language, knowledge and the world is illustrated in Fig. 2.

4. Linguistic models based on contextual information

Distributional hypothesis states that words that occur in same kinds of contexts tend to have similar meanings. Human intuition does not necessarily recognize this fact but it can be explained by considering corpora with millions or even billions of words. For instance, in the Europal corpus (Koehn 2005), words like "politician", "citizen", "agriculture", and "introduce" can be compared by collecting statistics on the words that precede these four words. For "politician" and "citizen", the most common preceding words are "a" and "any". On the other, more fine grained distinctions can be made when one notices that "politician" is often preceded by words such as "opposition" and "elected" that do not appear in the immediate context of the "citizen". Even though the

word "agriculture" is a noun, the most common preceding words differ from the two words discussed above. In this case, frequent contexts include "on", "of", "the" and "for", soon followed by specific words such as "sustainable" and "Mediterranian". The different syntactic category of the word "introduce"is reflected by the fact that the most common preceding words are "to", "and", "must" and "not". There are a number of different ways to collect context statistics and quite anumber of methods that can be used to analyze these statistics to detect linguistically relevants relationships and features. These options are discussed, e.g, in (Sahlgren 2006) and (Honkela et al. 2012).

Vector space models apply the distributional hypothesis and represent documents as the occurrences of words in it with the bag-of-words model (Salton et al. 1975). The transition from a sequence of symbols in a text corpus to a numerical matrix enables the use of linear algebra and general-purpose machine learning algorithms. Vector spaces can be built for any units for which co-occurrences can be computed. The validity of the distributional hypothesis has been successfully tested in various ways since late 1980s (Ritter & Kohonen 1989, Church & Hanks 1990, Deerwester et al. 1990).

It has recently been shown that data-driven semantic similarity judgments can be qualitatively comparable to human judgments (Lindh-Knuutila & Honkela 2013). Human beings can have different points of view into any matter, and similarly, unsupervised machine learning methods can give rise to different conceptual structures, each of which may be useful (Janasik et al. 2009). One unsupervised learning method, the self-organizing map (SOM) (Kohonen 1982, 2001) has been used widely in the analysis and visualization of complex data sets even though it was originally developed as an abstract model of the process of cortical organization.

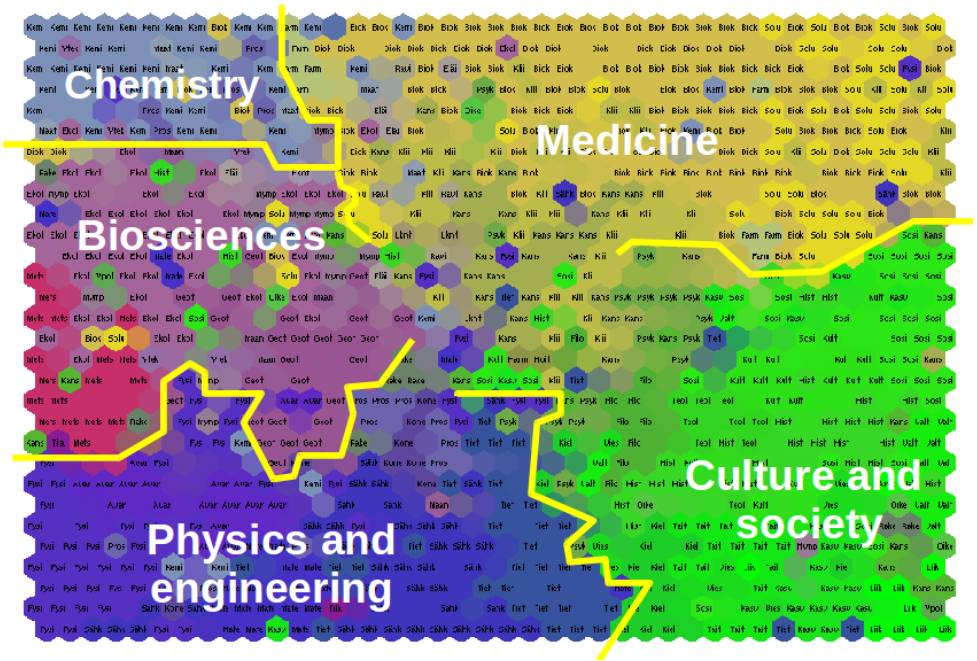

Figure 3. A self-organized map of Finnish science based on 3224 applications directed to Academy of Finland. The content analysis is fully automatic without any use of dictionaries, parsers, taxonomies, ontologies or any other similar manually constructed resources.

One of the applications of the SOM has been to create maps of documents in which two documents are close to each other in a two-dimensional space if their contents are similar (Kaski et al. 1998). An example of a document map is shown in Fig. 3. A collection of 3224 applications sent to Academy of Finland were analyzed fully automatically in a process where Language Independent Keyphrase Extraction (Likey) method (Paukkeri et al.. 2008) was used to extract relevant terminology which was then used to encode the documents to be analyzed using the

Self-Organizing Map algorithm (Kohonen 2001). The Likey method is able to find automatically terms (words or phrases) using an approach in which the frequency or rank of a word or phrase in the corpus at hand is compared with that in a large reference corpus (Paukkeri et al. 2008).

In addition to the self-organizing maps, linguistic vector spaces have been successfully analyzed using methods like singular value decomposition (Deerwester et al. 1990, Dumais & Landauer 1997), independent component analysis (Honkela et al. 2003, 2010) and latent Dirichlet allocation (Blei et al. 2003).

5. Multimodal contexts and pragmatics

Text corpora have proved to be useful resources when one wishes to create linguistic models cost effectively. However, if language technologies are to approach human level skills, the inherent lack of real world experience is a clear problem. This issue is often referred to as the symbol grounding problem (Harnad 1990). It is a relevant concern when, for instance, concrete objects, real world actions, and social relationships and processes are considered. Visual input is a primary source of real world experience for human beings. Computers can store images and videos but the link between visual patterns and symbolic descriptions is far from straightforward.

From the methodological point of view, multimodal contexts are more naturally represented as patterns and signals which differ considerably from the discrete representation of symbols and expressions of symbolic languages. This finding has some important consequences on the methodology that is needed in the computational modeling of these phenomena and processes. In particula, methods developed for pattern recognition are relevant here (cf., e.g., Oja 1983, Bishop 1995, Theodoridis & Koutroumbas 2003).

The use of text context as a source for learning automatically linguistic description was discussed above. The importance of context in interpretation has been recognized by many (see, e.g., Hörmann 1986). A context effect is demonstrated in Fig. 4 to illustrate that what would not be named as white without context, « becomes » white when the effect of shadows and other unoptimal viewing conditions are taken into account in a human mind. Similarly, « red » can refer to a rather wide range of colors, for instance, when the contexts « red shirt », « red skin » and « red wine » are considered.

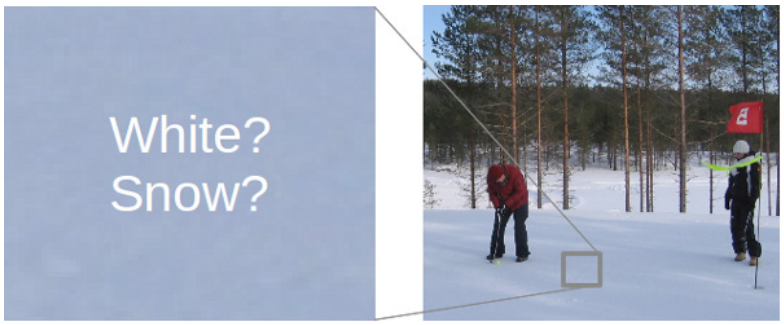

Figure 4. An illustration of a context effect related to naming, categorization and perception: the « input » color of a raw perception of color can be distant from prototypical white and even then it is perceived as white.

Contextuality is a central reason why it is extremely difficult to develop computerized systems that would have linguistic skills at human level. A person who has a lot of experience of some domain has a refined model of contextually relevant interpretations related to the domain and to the expressions used to describe phenomena within that domain. This is where human translators will have a « competitive advantage » over computerized systems at least for a long time. This requires, however, that the translator has a good understanding of the domain at hand. For

poetry and fiction, this often means human life in general with its high and low points. In specific domains, this underlines the importance of domain knowledge in addition to linguistic skills in the source and target language. In the future, it may be even more beneficial than nowadays as a human translator to specialize in some domains. An example that combines machine translation and human domain and linguistic knowledge can be given. A sentence is taken from a Russian web site that is known to have something interesting for the reader:

> "В последнее время для решения этой задачи активно применяются искусственные нейронные сети, реализованные в соответствии с парадигмой коннекционизма, благодаря которой была достигнута высокая точность классификации." (source: http://www.snipetz.com/neuro/1-4.htm)

However, the reader such as the presented author does not have any knowledge of Russian. Therefore, the Google Translate system is used to obtain the following result:

> "Recently, for solving this problem are actively used artificial neural network, implemented in accordance with the paradigm of connectionism, which was achieved thanks to the high accuracy of the classification."

For a human translator who does not have any domain knowledge related to the text, it would be difficult to determine how to translate or to post-edit this text in an appropriate manner. The present author, without any knowledge of Russian, is able to provide most likely a reasonable translation in the end. Here, one could even consider omitting the phrase "within the paradigm of connectionism" as redundant, because connectionistic models and artificial neural networks are more or less synonyms:

> **"Within the paradigm of connectionism, artificial neural networks have** recently been used for solving this problem, and thanks to them, a high classification accuracy has been achieved.**"**

Contextuality is an essential aspect in pragmatics which sometimes is defined to deal with meaning in context whereas semantics deals with prototypical meanings. Farwell and Helmreich (1999, 2006) have carefully considered pragmatics in translation and provide a large number of insightful examples to illustrate the importance of this issue.

6. Multimodally grounded language technology

At Aalto University School of Science, we have recently started a project, funded by Academy of Finland, on Multimodally Grounded Language Technology (MGLT). In the project, the basic scientific question is how to computationally model the interrelated processes of understanding natural language and perceiving and producing movement in multimodal real world contexts. The early results have shown that more robust manipulation of linguistic data becomes possible, e.g., when resolving ambiguities or when deeper inference is needed (Honkela & Förger 2013). In an earlier related work, Saenko and Darrell (2008) have studied how to visual information in relation to polysemous words like « mouse » that has multiple senses with visually very different results related to the animal, to computers and to the cartoon character.

In Fig. 5, the relationships between modalities that are considered in the MGLT project are illustrated. A Survey on Naming Human Movement has been started (see http://research.ics.aalto.fi/cog/mglt/ for details and to participate the survey). The purpose of the survey is to find out how people describe human motion including also attributes that are used. The data is being collected for many languages so that grounded mapping over language borders can also be conducted. Moreover, the data provides a chance to analyze variation in naming the movements in one language. Grounded Intersubjective Concept Analysis method has been developed for such purposes (Honkela et al. 2012). It applies Subject-Object-Context tensors that provide a rigorous framework for modeling subjective variation of understanding linguistic expressions in context (Honkela et al. 2012).

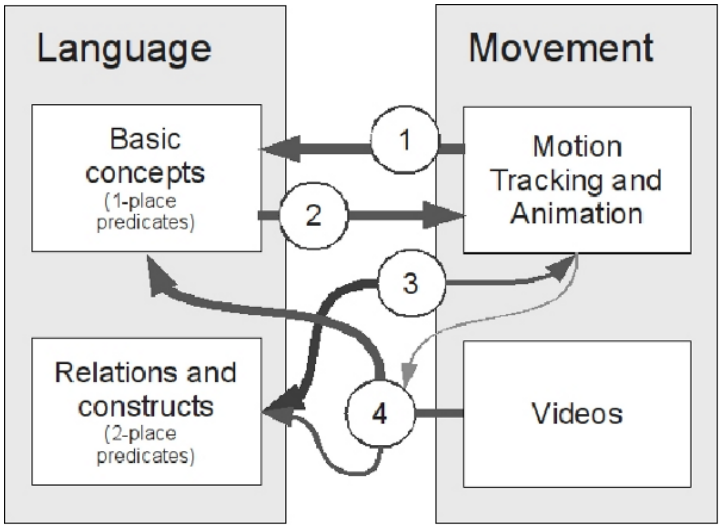

Figure 5. Relations between modalities under consideration in the Multimodally Grounded Language Technology project.

7. Translation strategies

Even though machine translation has been an active research topic for well over fifty years (Hutchins 2001), it can still be considered to be in its infancy. This statement may sound striking but one clear motivation is that machine translation could well be the most difficult task ever considered to be given to the computers. There are approximately six thousand languages in the world and many more dialects. Each language has from thousands to a million of lexemes and in some languages these lexemes give rise to even billions of different surface word forms. Modeling the syntax of any living language has proved to be a challenging problem and still much simpler than modeling the semantic mapping, i.e. defining the prototypical meaning. This task is again much simpler than modeling how language is understood in context, how it changes over time, how it varies over the population of speakers of one language, and how different languages divide the underlying conceptual space in different ways. This is not even a complete list of complexities that would need to be considered in building a "perfect" machine translator.

One interesting future challenge is related to variation. In his book "Exercices de style", Queneau tells a story in 99 ways, each time in a different style. In a similar vain, different kinds of translations are needed in different situations. Chesterman (1997) considers a number of different translation strategies. Syntactic strategies include literal translations, loans, changes of the part of speech, structural changes, etc. Potential semantic strategies are using a synonym, an antonym, a hyponym or a hyperonym. Condensing and expanding are also possible as well as modulation, e.g., from concrete to abstract. Pragmatic changes include addition, omission, explicitation, implicitation, domestication, foreignization, formality change, etc. Translators have thus a large number of alternatives to choose from.

Machine translation is still considered usually in the mind set of finding one « correct » target language expression that would correspond to the source language expression. The problem is, however, a high dimensional one with multiple, often contradictory criteria. Different contextual aspects are taken into account at best including the target audience. In the area of information retrieval, we have recently taken a step towards this direction by developing a method that can be used to assess the terminological readability of texts in a personalized manner (Paukkeri et al. 2013). Perhaps in the future we have machine translation systems that we can ask to produce texts in the target language so that we set the parameters according to some translation strategies. A related scenario was provided by Stanislaw Lem in his short stories on Trurl and Klapaucius (Lem 1975). In one story, Trurl builds a machine that could write

poetry with remarkable consequences. The translation of the story itself must be a challenge for human translators and goes far beyond the capacities of current machine translators.

8. Acknowledgements

The organizers of the Tralogy II conference are gratefully acknowledged for their kind invitation to give a talk in the interesting and important event. Research partners in the META-NET consortium are to be thanked for, especially Joseph Mariani and Pierre Zweigenbaum (LIMSI-CNRS, Paris), Stelios Piperidis (ILSP, Athens), as well as the Aalto colleague Jaakko Väyrynen who currently works at EC JCR in Ispra, Italy The author is also grateful to the collaborators within the Multimodally Grounded Language Technology project including Tapio Takala, Jorma Laaksonen, Markus Koskela, Harri Valpola, Klaus Förger, Xi Chen, Paul Wagner and Oskar Kohonen

Bibliography

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, *284*(5), 28-37.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford university press.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993-1022.

Buchanan, B. G., & Shortliffe, E. H. (1984). *Rule Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley Longman Publishing.

Chesterman, A. (1997). *Memes of translation: The spread of ideas in translation theory*. John Benjamins.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, *41*(1), 83-121.

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational linguistics*, *16*(1), 22-29.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, *41*(6), 391-407.

Dumais, S. T., & Landauer, T. K. (1997). A solution to Platos problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological review*, *104*(2), 211-240.

Farwell, D., & Helmreich, S. (1999). Pragmatics and translation. *Procesamiento de Lenguaje Natural*, *24*, 19-36.

Farwell, D., & Helmreich, S. (2006). Pragmatics-based MT and the Translation of Puns. Proceedings of EAMT'06.

Fodor, J. A. (1975). *The language of thought*. Harvard University Press.

Gärdenfors, P. (2004). How to make the semantic web more semantic. In *Formal Ontology in Information Systems* (pp. 19-36).

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1), 335-346.

Honkela (1997). *Self-Organizing Maps in Natural Language Processing*. PhD thesis, Helsinki University of Technology, Espoo, Finland.

Honkela, T., Hyvärinen, A., & Väyrynen, J. (2003). *Emergence of linguistic representations by independent component analysis*. Technical report, Helsinki University of Technology.

Honkela, T. (2005). Von Foerster meets Kohonen: Approaches to artificial intelligence, cognitive science and information systems development. *Kybernetes*, 34(1/2), 40-53.

Honkela, T. (2007). Philosophical aspects of neural, probabilistic and fuzzy modeling of language use and translation. *Proceedings of IJCNN*, pp. 2881-2886.

Honkela, T., Hyvärinen, A., & Väyrynen, J. J. (2010). WordICA—emergence of linguistic representations for words by independent component analysis. *Natural Language Engineering*, *16*(3), 277-308.

Honkela, T., Raitio, J., Lagus, K., Nieminen, I. T., Honkela, N., & Pantzar, M. (2012). Subjects on objects in contexts: Using GICA method to quantify epistemological subjectivity. In *Proceedings of IJCNN 2012, International Joint Conference on Neural Networks,* pp. 2875-2883.

Honkela, T. & Förger, K. (2013). Modeling Action Verb Semantics Using Motion Tracking. Unpublished manuscript. (When published, will be made available at http://research.ics.aalto.fi/cog/mglt/)

Hörmann, H. (1986). *Meaning and Context*. Plenum Press, New York.

Hutchins, W. J. (2001). Machine translation over fifty years. *Histoire epistémologie langage*, *23*(1), 7-31.

Janasik, N., Honkela, T., & Bruun, H. (2009). Text mining in qualitative research application of an unsupervised learning method. *Organizational Research Methods*, *12*(3), 436-460.

Jäppinen, H., & Ylilammi, M. (1986). Associative model of morphological analysis: an empirical inquiry. *Computational Linguistics*, *12*(4), 257-272.

Jäppinen, H., Honkela, T., Hyötyniemi, H., Lehtola, A. (1988). A multilevel natural language processing model. *Nordic Journal of Linguistics*, vol. 11 pp. 69-87.

Johnson, R., King, M., & des Tombe, L. (1985). Eurotra: A multilingual system under development. *Computational Linguistics*, 11(2-3), 155-169.

Kaski, S., Honkela, T., Lagus, K., & Kohonen, T. (1998). WEBSOM–self-organizing maps of document collections. *Neurocomputing*, *21*(1), 101-117.

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In *MT summit* (Vol. 5).

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological cybernetics*, *43*(1), 59-69.

Kohonen, T. (2001). *Self-Organizing Maps*. Springer.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. Basic Books.

Lehtola, A., Honkela, T., Hyötyniemi, H., & Jäppinen, H. (1988). Task Oriented Knowledge Representation Languages for NLP-Systems. In *Third International Symposium on Methodologies for Intelligent Systems (ISMIS'88),* Torino, Italy, pp. 250-259.

Lem, S. (1975). *The Cyberiad - fables for the cybernetic age*. translated by M. Kandel. Secker and Warburg, UK.

Lenat, D. B., & Guha, R. V. (1989). *Building large knowledge-based systems; representation and inference in the Cyc Project*. Addison-Wesley Longman Publishing.

Lindh-Knuutila, T., & Honkela, T. (2013). Exploratory Text Analysis: Data-Driven versus Human Semantic Similarity Judgments. In *Adaptive and Natural Computing Algorithms* (pp. 428-437). Springer, Berlin Heidelberg.

Mellish, C., & Sun, X. (2006). The semantic web as a < i> Linguistic</i> resource: Opportunities for natural language generation. *Knowledge-Based Systems*, *19*(5), 298-303.

Montague, R. (1973). The proper treatment of quantification in ordinary English. Approaches to Natural Language: Proceedings of the Stanford Workshop on Grammar and Semantics, J. Hintikka, J. Moravcsik, and P. Suppes, Eds. Dordrecht: D. Reidel, 1973, pp. 221-242.

Niles, I., & Pease, A. (2001). Towards a standard upper ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001* (pp. 2-9). ACM.

Noy, N. F. (2004). Semantic integration: a survey of ontology-based approaches. *ACM Sigmod Record*, *33*(4), 65-70.

Oja, E. (1983). *Subspace methods of pattern recognition* (Vol. 142). England: Research Studies Press.

Paukkeri, M. S., Nieminen, I. T., Pöllä, M., & Honkela, T. (2008). A language-independent approach to keyphrase extraction and evaluation. In *Proceedings of COLING,* pp. 83-86.

Paukkeri, M. S., Ollikainen, M., & Honkela, T. (2013). Assessing user-specific difficulty of documents. *Information Processing & Management,* 49(1):198–212.

Queneau, R. (1947). *Exercices de style*. Gallimard.

Regier, T. (1996). *The human semantic potential: Spatial language and constrained connectionism*. Bradford Book.

Ritter, H., & Kohonen, T. (1989). Self-organizing semantic maps. *Biological cybernetics*, *61*(4), 241-254.

Saenko, K., & Darrell, T. (2008). Unsupervised learning of visual sense models for polysemous words. *Advances in Neural Information Processing Systems*, *21*, 1393-1400.

Sahlgren, M. (2006). *The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces.* Doctoral dissertation, Stockholm University.

Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, *18*(11), 613-620.

Theodoridis, S., & Koutroumbas, K. (2003). Pattern recognition. *Academic Press, Burlington MA, USA*.

Valkonen, K., Jäppinen, H., Lehtola, A. (1987). Blackboard-based dependency parsing. *Proceedings of IJCAI'87, International Joint Conference on Artificial Intelligence*, pp. 700-702.

Winograd, T. (1972). *Understanding Natural Language*. Academic Press, New York.

Wittgenstein, L. (1922). Tractatus Logico-Philosophicus. Routledge & Kegan Paul, London.