Computational Cognitive Systems

Chapter 11

Emergence of linguistic and cognitive representations

Timo Honkela, Krista Lagus, Tiina Lindh-Knuutila, Matti Pöllä, Juha Raitio, Sami Virpioja, Jaakko J. Väyrynen and Paul Wagner

11.1 Introduction

Computational Cognitive Systems group conducts research on artificial systems that combine perception, action, reasoning, learning and communication. This area of research draws upon biological, cognitive and social system approaches to understanding cognition. Cognitive systems research is multidisciplinary and benefits from sharing and leveraging expertise and resources between disciplines. For example, statistical machine learning, pattern recognition and signal processing are central tools within computational cognitive systems research. Our research focuses on modeling and applying methods of unsupervised and reinforcement learning.

The general aim is to provide a methodological framework for theories of conceptual development, symbol grounding, communication among autonomous agents, agent modeling, and constructive learning. We have also worked in close collaboration with other groups in our laboratory, for instance, related to multimodal interfaces.

In the following chapters of this report, we describe the main results gained during 2006-07 in the four main thematic areas of the computational cognitive systems research group: *emergence of linguistic and cognitive representations*, *learning social interactions between agents*, *learning to translate* and *knowledge translation and innovation using adaptive informatics.* Each of these areas is described in a section of its own except for emergence of linguistic and cognitive representations that is a shared research area with Multimodal Interfaces research group and the results are mainly described in the chapter on Natural Language Processing. In the following, this area is discussed briefly. In general, the group has benefited strongly from the closeness and the collaboration with other research groups in the center. A notable example of such collaboration is the development of a *speech-to-speech machine translation system protytype*, developed in collaboration with the Multimodal Interfaces research group.

11.2 Research on emergence

The research on emergent linguistic and cognitive representations enables computers to deal with semantics: to process data having certain access to its meaning within multimodal contexts. Our focus is on the analysis and generation of conceptual structures and complex meanings.

The emergence of representations can be considered to consist of the following interrelated tasks: the discovery of elements of representation (e.g. words, morphemes, phonemes), their meaning relations (syntax and semantics), and structures or “rules” of their use in natural utterances (syntax and pragmatics).

An example of work on the first topic is the study of unsupervised machine learning techniques for finding the optimal segmentation of words into sub-word units called morphs, with the intent of finding realizations of morphemes (see Section 10.1 for details). Another example of the use of independent component analysis in discovering meaningful syntactic and semantic features for words (see Section 10.1 for details).

We have studied the emergence of linguistic representations through the analysis of words in contexts using the Independent Component Analysis (ICA). The ICA learns features automatically in an unsupervised manner. Several features for a word may exist, and the ICA gives the explicit values of each feature for each word. In our experiments, we have shown that the features coincide with known syntactic and semantic categories. More detailed description of this research is given in the section on Natural Language Processing in this report.

11.3 Events and projects

An important part of our activity has been the active role in organizing international scientific events. The main activity in 2006-2007 was the organization of the Scandinavian Conference on Artificial Intelligence [1].

In addition to the research areas discussed above, we continue to use the Self-Organizing Map (SOM) where applicable. The most important piece of research based on the SOM during this period was an analysis of conducted for the Academy of Finland. As a continuation to an earlier manually conducted qualitative analysis, the Academy of Finland, one of the country's largest funding agencies, commissioned a study to investigate whether text mining based on the Self-Organizing Map could be used to support assessment of the applications. A collection of 3224 applications was analyzed [2]. A collection of 1331 term candidates was extracted automatically. The 3224 application documents were encoded as term distribution patterns. The SOM algorithm organized the documents into a map in which similar applications are close to each other and in which thematic areas emerged.

Many parts of the research in this new group's agenda have been started during the reported period. Therefore, so far the results have mainly been reported in conferences and as technical reports. However, in early 2008 we were pleased to receive news about four accepted journal papers. Some of the related results are described already in this report, based on publications that have appeared during 2006 and 2007.

Some of the research activities by the group take place in projects funded by Tekes and EU Commission. The most notable examples are the projects Kulta (using adaptive informatics methods to model and simulate changing needs of consumers) and MeIEQ (developing quality labeling methods for medical web resources), discussed in some detail in the section on Knowledge translation and innovation using adaptive informatics.

References

[1] T. Honkela, T. Raiko, J. Kortela and H. Valpola (eds.) (2006). *Proceedings of SCAI'06, the Ninth Scandinavian Conference on Artificial Intelligence.* Finnish Artificial Intelligence Society, 239 p. Espoo.

[2] T. Honkela, M. Klami (2007). Text mining of applications submitted to the Academy of Finland [in Finnish], unpublished report. Helsinki University of Technology, Espoo.

Chapter 12

Learning social interactions between agents

Ville Könönen, Timo Honkela, Tiina Lindh-Knuutila, Mari-Sanna Paukkeri

12.1 Introduction

One important feature of an intelligent agent is its ability to make rational decisions based on its current knowledge of the environment. If the environment of the agent is not static, i.e., there are other active entities, e.g., other agents or humans in the environment, it is crucial to model these entities for making rational decisions.

Our earlier research has been concentrated on the theoretical aspects of the modeling other agents in the reinforcement learning framework. Reinforcement learning is a learning paradigm located between supervised and unsupervised learning. In an environment, the agent takes actions and receives reward signals corresponding to the success of these actions. Correct answers are not directly provided to the agent but it learns features of the environment by continuously interacting with it.

12.2 Applications of multiagent reinforcement learning

Reinforcement learning methods have attained lots of attention in recent years. Although these methods and procedures were earlier considered to be too ambitious and to lack a firm foundation, they have been established as practical methods for solving, e.g., Markov Decision Processes (MDPs). However, the requirement for reinforcement learning methods to work is that the problem domain in which these methods are applied obeys the Markov property. Basically this means that the next state of a process depends only on the current state, not on the history. In many real-world problems this property is not fully satisfied. However, many reinforcement learning methods can still handle these situations relatively well. Especially, in the case of two or more decision makers in the same system the Markov property does not hold and more advanced methods should be used instead. A powerful tool for handling these highly non-Markov domains is the concept of Markov game. In this section we introduce two applications of multi-agent reinforcement learning, namely a dynamic pricing problem and a communication game between agents.

Dynamic pricing

A dynamic pricing scenario is a problem domain that requires planning and therefore it is an ideal testbed for reinforcement learning methods. In the problem, there are two competing brokers that sell identical products to customers and compete on the basis of price. We have modeled the problem as a Markov game and solve it by using two different learning methods. The first method utilizes modified gradient descent in the parameter space of the value function approximator and the second method uses a direct gradient of the parameterized policy function. [1]

Communication game between agents

As another problem, we consider a multiagent system, where unsupervised learning is used in the formation of agents' conceptual models. An intelligent agent usually has a goal. For studying agent based systems formally, it is useful that the goal can be expressed mathematically. Traditional approach is to define a utility function for the agent, i.e. there is a scalar value connected to each possible action measuring the fitness of the action choice for satisfying the goal of the agent. The utility function is often initially unknown and must be learned by interacting with the environment, e.g. by communicating with other agents. [2]

The agents communicate and learn through communication leading into intersubjective sharing of concepts. Communication can be modeled as a mathematical game and the structure of the game can be learnt by using reinforcement learning methods.

References

[1] V. Könönen (2006). Dynamic Pricing Based on Asymmetric Multiagent Reinforcement Learning. *International Journal of Intelligent Systems*, vol. 21, pp. 73–98.

[2] T. Honkela, V. Könönen, T. Lindh-Knuutila and M. Paukkeri (2006). Simulating processes of language emergence, communication and agent modeling. In *Proceedings of STeP-2006, the 12th Finnish Artificial Intelligence Conference*, pp. 129–132. Espoo, Finland.

12.3 Meaning negotiation using simulated naming games

The acquisition of concepts can be viewed as the process of grounding them in language use. How do concepts form, to begin with, when no language exists in the environment? The relationship between cognitive and linguistic development can be pinpointed to the question: how do conceptual emergence and the formation of names for the concepts connect? An associated important question is how an agreement on the use of words is reached in a community of agents. This process of converging towards a shared use of words is called meaning negotiation.

In [4], we studied the emergence of associations between concepts and words with help of a computer simulation and consider the hypothesis that concepts are modeled as areas in some conceptual space (see [1]). Furthermore, we utilized the self-organizing map [3] as a model for an agent's conceptual memory [2]. The feasibility of this conceptual memory model was then studied using multi-agent language games, more specifically, naming games, in which agents learn to associate words to meanings in a communicative setting. Prior to learning word-meaning-associations, a conceptual representation is also learned individually by each agent based on sensory data.

Examples of conceptual maps from simulations with six agents are shown in Fig. 12.1. They show the conceptual space of one agent after the simulation run. The self-organizing map is labeled with the words that have been used during the simulation.

Figure 12.1: An example conceptual map from the simulation. The shades of gray denote distances in the original space: The larger the distance, the darker the color. One can see that there are eight clusters on each map which corresponds to the number of prototypical colors used in the input.

The results of the experiments show clearly that when using the observational game model and the SOM-based conceptual maps (1) the agents learned to communicate successfully on the topics of the games, and (2) a shared lexicon was developed during the simulations. According our definition of successful communication, the agents are also able to communicate successfully and develop a shared lexicon based on adaptation.

References

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

[2] T. Honkela. Neural nets that discuss: a general model of communication based on self-organizing maps. In S. Gielen and B. Kappen, editors, *Proceedings of ICANN'93, International Conference on Artificial Neural Networks*, pages 408–411, Amsterdam, the Netherlands, September 1993. Springer-Verlag, London.

[3] T. Kohonen (2001). *Self-Organizing Maps.* Third, extended edition. Springer.

[4] T. Lindh-Knuutila, T. Honkela, and K. Lagus (2006). 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*, pp. 168–179. Springer.

Chapter 13

Learning to translate

Timo Honkela, Mathias Creutz, Tiina Lindh-Knuutila, Sami Virpioja, Jaakko J. Väyrynen

13.1 Introduction

Learning to translate research focuses on developing methods and tools facilitating translations between different languages and even between different dialects or domains. Underpinning development of learning to translate methodology is the fact that contextual, experiential and/or disciplinary diversity impede interpersonal communication and understanding. Our research focuses on the use of unsupervised statistical machine learning. However, in comparison with the traditional approach in statistical machine translation (SMT), we want to take into account known linguistic levels and theories. This does not take place by encoding linguistic knowledge manually to the systems but through architectural choices. For instance, the basic statistical machine translation approach does not properly take into account the morphological or the semantic level. These issues are discussed in the following.

We have applied a method of unsupervised morphology learning to a state-of-the-art phrase-based SMT system [2]. In SMT, words are traditionally used as the smallest units of translation. Such a system generalizes poorly to word forms that do not occur in the training data. In particular, this is problematic for languages that are highly compounding, highly inflecting, or both. An alternative way is to use sub-word units, such as morphemes. We have used the Morfessor algorithm to find statistical morpheme-like units (called morphs) that can be used to reduce the size of the lexicon and improve the ability to generalize. This approach is described more in detail in Section 13.3.

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 that relate to the lack of semantic coverage and understanding of the pragmatic level of language. Statistical machine translation systems typically rely on applying Bayes' rule:

> We assign to every pair of strings, s (source) and t (target), in two languages a number $P(t|s)$, which is the probability that a translator, when presented with s, will produce t as the translation. Using Bayes' theorem, one can write $P(s|t) = P(t|s) * P(s)/P(t)$

Thus, in the basic SMT approach, the inputs and outputs are handled only as strings of symbols (consider, e.g., [1, 3]. The system does not receive or deal with information on the meaning of the expressions. To overcome this limitation, there are a number of systems with a hybrid approach, using, for instance, a parser that annotates the training samples with (syntactic and) semantic labels. However, as we wish to minimize the manual effort in development machine translation systems, we have chosen not to use traditional parsers or labeling schemes. Rather, we build on distributional information. Namely, the finding that word co-occurrence statistics, as extracted from text corpora, can provide a natural basis for semantic representations has been gaining growing attention. Words with similar distributional properties often have similar semantic properties. Therefore, it is possible to dynamically build semantic representations of the lexical space through the statistical analysis of the contexts in which words co-occur. In addition to distributional information on the word occurrences in text corpora, also other kinds of contextual information may be used.

The early work by Ritter and Kohonen with artificially generated short sentences as well as contextual information showed the feasibility of the approach outlined above [4]. This work was extended to natural data in [2]. In Section 13.4, we describe how the use of the Self-Organizing Map can be extended to multilingual processing in order to find semantic grounding for expressions in multiple languages. Some preliminary results on *visual grounding of meaning* are also discussed.

Before addressing the use of unsupervised learning in finding morphological and semantic models that are useful for machine translation, we consider in the following the structural complexity of a number of European languages in Section 13.2. The basic motivation for this analysis lies in the hypothesis that the translation between such two languages is relatively easier that encode information in a similar manner with respect to morphology and syntax. The results of the analysis should thus help in designing translation strategies for automated solutions for various pairs of languages.

References

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

[2] T. Honkela, V. Pulkki, and T. Kohonen (1995). Contextual relations of words in Grimm tales, analyzed by self-organizing map. In *Proceedings of ICANN'95, International Conference on Artificial Neural Networks*, vol. II, pp. 3–7. Nanterre, France: EC2.

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

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

13.2 Analyzing structural complexity of languages

The European Union has 21 official languages (including Irish from 1st of January 2007), which have approximately 407 million speakers. We have analyzed parallel corpora in these 21 languages using statistical, unsupervised learning methods to study the similarities and differences of the languages in different levels. We have compared these results with traditional linguistic categorizations like division into language groups, morphological complexity and syntactic complexity [3]. The aim of the study has been to evaluate the possibility of using statistical methods in different tasks related to statistical machine translation. For instance, for some language pairs the issues related to morphological analysis may be particularly relevant. For some other language pairs, one may have to pay particular attention to the word order. These kinds of questions can be taken into account when the statistical models to be used are chosen.

Use of compression as a measure for complexity is based on the concept of Kolmogorov complexity. Informally, for any sequence of symbols, the Kolmogorov complexity of the sequence is the length of the shortest algorithm that will exactly generate the sequence and then stop. In other words, the more predictable the sequence, the shorter the algorithm needed is and thus the Kolmogorov complexity of the sequence is also lower [4]. Kolmogorov complexity is uncomputable, but file compression programs can be used to estimate the Kolmogorov complexity of a given file. A decompression program and a compressed file can be used to (re)generate the original string. A more complex string (in the sense of Kolmogorov complexity) will be less compressible. Estimations of complexity using compression has been used for different purposes in many areas. Juola [2] introduces comparison of complexity between languages on morphological level for linguistic purposes.

To get a meaningful interpretation for the order of languages in the word order complexity counting, linguistic literature was consulted for independent figures. Bakker [1] has analyzed flexibility of language's word order, which is based on 10 factors, such as order of verb and object in the language, order of adjective and its head noun, order of genitive and its head noun, etc. The flexibility of the language in Bakker's counting can be given with a numeric value from 0 to 1: if the flexibility figure is close to zero, the language is more inflexible in its word order, if the figure is closer to one, the language is more flexible in its word order. In the information theoretic framework of the compression approach flexibility and inflexibility can be interpreted naturally as higher and lower degrees of complexity, i.e. predictability. In the table below, figures based on Bakker's counting of the flexibility values for the individual languages are given together with values given by compression analysis.

If one compares the figures given by Bakker in column 3 to figures given by compression based calculation in column 6, we can see, that the overall order of the languages based on these independent calculations converge well. The lower end of the scale is quite analogous in both analyses consisting of five same languages with differences in the order. There are also some differences in the orders given by the two analyses. The syntactic complexity of Lithuanian seems to be estimated higher by compression than by Bakker's flexibility value (rank 16 vs. 8). Slovene has also a higher flexibility value than its complexity value (rank 14 vs. 7). Greek is also higher in Bakker's counting than in complexity analysis (rank 17 vs. 11). In our compression calculations Finnish and Estonian are estimated almost equally complex, but in Bakker's analysis Estonian is less complex than Finnish (rank 18 vs. 13).[3]

Bakker's results			Compression results		
1.	fr	0.10	1.	fr	0.66
2.	ga	0.20	2.	es	0.68
3.	es	0.30	3.	pt	0.68
4.	pt	0.30	4.	ga	0.69
5.	it	0.30	5.	it	0.69
6.	da	0.30	6.	en	0.69
7.	mt	0.30	7.	sl	0.71
8.	lt	0.30	8.	nl	0.71
9.	en	0.40	9.	mt	0.72
10.	nl	0.40	10.	da	0.72
11.	de	0.40	11.	el	0.73
12.	sv	0.40	12.	sv	0.75
13.	et	0.40	13.	lv	0.75
14.	sl	0.50	14.	de	0.75
15.	lv	0.50	15.	pl	0.76
16.	sk	0.50	16.	lt	0.76
17.	el	0.60	17.	sk	0.77
18.	pl	0.60	18.	et	0.78
19.	fi	0.60	19.	fi	0.79

References

[1] D. Bakker (1998). Flexibility and Consistency in Word Order Patterns in the Languages of Europe. In Siewierska, A. (ed.): *Constituent Order in the Languages of Europe. Empirical Approaches to Language Typology.* pp. 381–419. Mouton de Gruyter, Berlin New York.

[2] P. Juola (1998) Measuring Linguistic Complexity: the Morphological Tier. *Journal of Quantitative Linguistics*, vol. 5, pp. 206–213.

[3] K. Kettunen, M. Sadeniemi, T. Lindh-Knuutila and T. Honkela (2006). Analysis of EU Languages Through Text Compression. In *Proceedings of FinTAL, the 5th International Conference on NLP*, pp. 99–109. Turku, Finland, August 23–25.

[4] M. Li, X. Chen, X. Li, B. Ma, P. M. B. Vitányi (2004). The Similarity Metric. *IEEE Transactions on Information Theory*, vol. 50, pp. 3250–3264.

13.3 Morphology-Aware Statistical Machine Translation

Statistical machine translation was applied to the direct translation between eleven European languages, all those present in the Europarl corpus, by [1]. An impressive number of 110 different translation systems were created, one for each language pair. Koehn discovered that the most difficult language to translate from or to is Finnish. Finnish is a non-Indo-European language and is well known for its extremely rich morphology. As verbs and nouns can, in theory, have hundreds and even thousands of word forms, data sparsity and out-of-vocabulary words present a huge problem even when large corpora are available.

It appears that especially translating into a morphologically rich language poses an even more substantial problem than translating from such a language. The study also showed that English, which has almost exclusively been used as the target language, was the easiest language to translate into. Thus it is natural to suspect that English as a target language has biased SMT research.

In the following, we describe how we have used morphological information found in an unsupervised manner in SMT [2]. We have tested the approach with the three Nordic languages, i.e., Finnish, Danish and Swedish. Danish and Swedish are closely related languages but differ considerably from Finnish. Danish and Swedish are grammatically very close and much of the vocabulary is shared except for some differences in pronunciation and orthography.

The parallel Europarl corpus [1] of European Parliament Proceedings was used to train our models. Word segmentation models for both source and target languages were trained using Morfessor. At this point, two data set were created for each alignment pair: one with the original word tokens and the other with morph tokens. This allowed us to create a comparable baseline system. We used standard state-of-the-art n-gram language models trained with the target language text. Phrase-based translation models were trained with Moses, an open-source statistical machine translation toolkit [3]. A phrase-based system translates short segments of consecutive words in contrast to word-based translation, which translates one word at a time. Phrases enable more natural language generation and flexibility in translating, for instance, idioms, collocations, inflected words forms and compound words in which the number of words may not stay the same across translation. The parameters of word-based and morph-based system were the same, except for the maximum number of tokens in a phrase, that was higher with morph-based systems to cover approximately the same number of words than word-based systems.

Figure 13.1 shows an example how in addition to the the different tokens, morph-based translation first segments words into morphs and finally after the translation constructs words from the translated morphs. Having morph tokens lowers the type counts greatly compared to words. The segmentation naturally increases tokens counts slightly. Reduced type counts help with sparse data, and this was especially prominent with Finnish. On the other hand, increased token counts seem to make the word alignment and translation process more complicated.

In the word-based translation model, only the words that were present in the training data can be translated. The other words are left untranslated, even though they may simply by an inflected form of a known word. Thus we expected to get less untranslated words with the morph-based system. This was true, as shown in Table. 13.1. An examination of the untranslated words reveals that a higher number of compound words and inflected word forms are left untranslated by the word-based systems.

As in most of the recent studies, we have used the BLEU scores [4] for quantitative evaluation. BLEU is based on the co-occurrence of n-grams between a produced transla-

<table>
<tr><td>a</td><td colspan="4">flera reglerande åtgärder behöver införas .</td></tr>
<tr><td>b</td><td>flera</td><td>reglerande åtgärder</td><td>behöver införas</td><td>.</td></tr>
<tr><td>c</td><td>eräitä</td><td>sääntelytoimia</td><td>on toteutettava</td><td>.</td></tr>
<tr><td>d</td><td colspan="4">eräitä sääntelytoimia on toteutettava .</td></tr>
</table>

<table>
<tr><td>e</td><td colspan="5">flera reglerande åtgärder behöver införas .</td></tr>
<tr><td>f</td><td colspan="5">flera$_0$ reglera$^*_0$ nde$_+$ åtgärd$^*_0$ er$_+$ behöv$^*_0$ er$_+$ in$^*_-$ föra$^*_0$ s$_+$.$_0$</td></tr>
<tr><td>g</td><td>flera$_0$</td><td>reglera$^*_0$ nde$_+$</td><td>åtgärd$^*_0$ er$_+$</td><td>behöv$^*_0$ er$_+$ in$^*_-$ föra$^*_0$ s$_+$</td><td>.$_0$</td></tr>
<tr><td>h</td><td>erä$^*_0$ itä$_+$</td><td>sääntely$^*_0$</td><td>toimi$^*_0$ a$_+$</td><td>on$_0$ toteute$^*_0$ tta$^*_+$ va$_+$</td><td>.$_0$</td></tr>
<tr><td>i</td><td colspan="5">erä$^*_0$ itä$_+$ sääntely$^*_0$ toimi$^*_0$ a$_+$ on$_0$ toteute$^*_0$ tta$^*_+$ va$_+$.$_0$</td></tr>
<tr><td>j</td><td colspan="5">eräitä sääntelytoimia on toteutettava .</td></tr>
</table>

Figure 13.1: Examples of word-based and morph-based Finnish translations for the Swedish sentence "Flera reglerande åtgärder behöver införas ." (*Several regulations need to be implemented .*) The top figure shows the word-based translation process with the source sentence (a), the phrases used (b) and their corresponding translations (c), as well as the final hypothesis (d). The bottom figure illustrates the morph-based translation process with the source sentence as words (e) and as morphs (f), the morph phrases used (g) and their corresponding translations (h), as well as the final hypothesis with morphs (i) and words (j). Each morph if either a prefix (−), a stem (0) or a suffix (+), marked by the lowerscript. A superscript (∗) marks the morphs that are not the last one in the word.

word / morph	→ Danish	→ Finnish	→ Swedish
Danish →		128 / 31	74 / 12
Finnish →	189 / 41		195 / 44
Swedish →	76 / 21	132 / 42	

Table 13.1: Number of sentences with untranslated words out of 1 000 with word-based and morph-based phrases.

tion and a reference translation. BLEU score has been critized, for instance, as in some cases human evaluation gives grossly different results. It is also clear that for morphologically rich languages, such as Finnish, it is harder to get good scores on a word-token based evaluation method. In Table 13.2, the differences between the scores for word-based and morph-based systems are shown, with statistically significant differences highlighted. According to these results, the translations based on morph phrases were slightly worse, but only in two cases the decrease was statistically significant.

References

[1] P. Koehn (2005). Europarl: A parallel corpus for statistical machine translation. In *Proceedings of MT Summit X, 10th Machine Translation Summit*, pp. 79–86. Phuket, Thailand, Sep 13–15.

[2] S. Virpioja, J. J. Väyrynen, M. Creutz and M. Sadeniemi (2007). Morphology-Aware Statistical Machine Translation Based on Morphs Induced in an Unsupervised Manner. In *Proceedings of MT Summit XI, 11th Machine Translation Summit*, pp. 491–498. Copenhagen, Denmark, Sep 10–14.

	→ Danish	→ Finnish	→ Swedish
Danish →		-0.60	-0.52
Finnish →	-1.23		**-2.14**
Swedish →	-0.46	**-1.14**	

Table 13.2: Absolute changes in BLEU scores from word-based translations to morph-based translations. The maximum phrase length was 7 for words and 10 for morphs. 4-gram language models were used for both. Statistically significant differences are marked with boldface fonts.

[3] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, & E. Herbst (2007). Moses: Open source toolkit for statistical machine translation. In *Annual Meeting of ACL, demonstration session*. Czech Republic, June.

[4] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu (2002). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of ACL 2002*, pp. 311–318. Morristown, NJ, USA.

13.4 Self-Organizing Semantic Representations for Machine Translation

Discussing the fundamental problems of translation, Quine has presented 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 [3]. 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. Even further, if one considers the symbol grounding problem [1], 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.

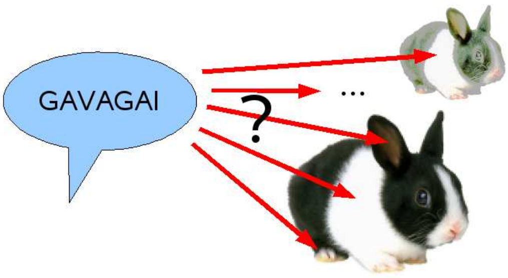

Figure 13.2: An illustration of the reference problem (see text for details).

Quine mentions that one can form manuals of translation [3]. 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, 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 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.

Finding a mapping between vocabularies of two different languages, the results of a new experiment are reported in the following. Maps of words are often constructed using distributional information of the 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. We have extended this basic idea to cover the notion of context in general. We have considered 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.

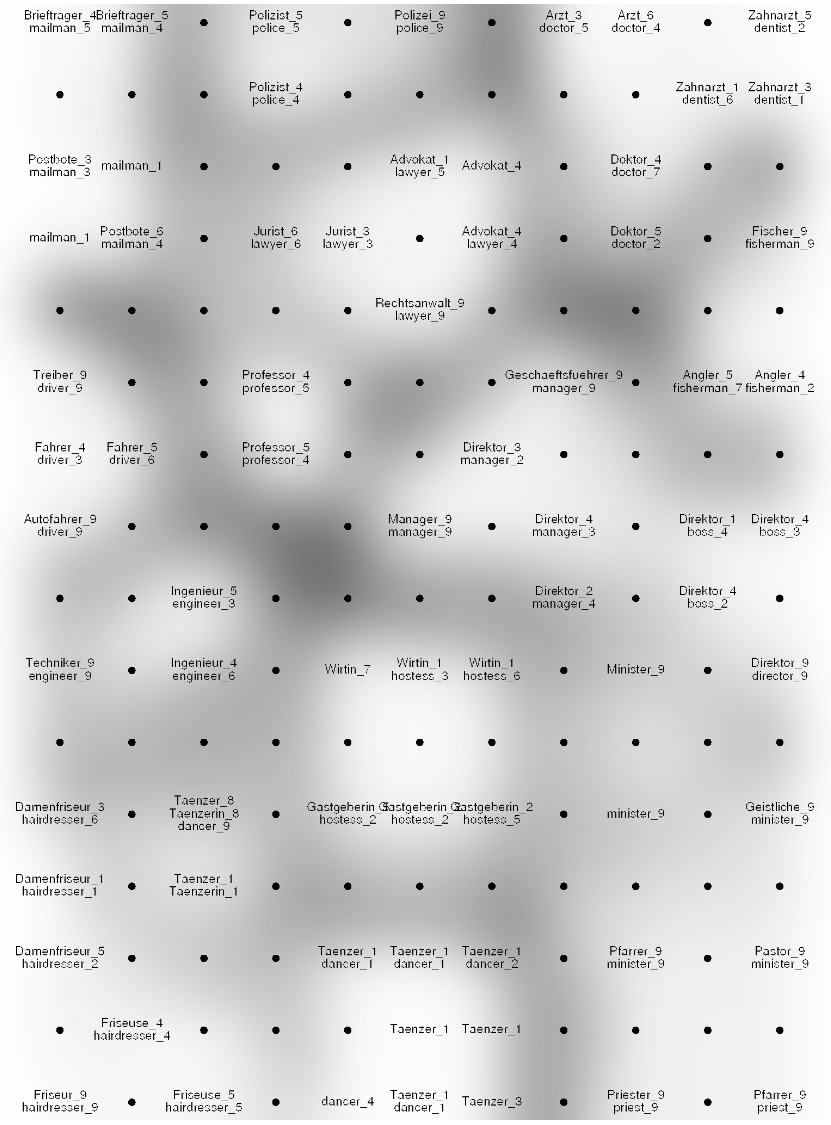

Figure 13.3: An illustration of the reference problem (see text for details).

Figure 13.3 presents the order of some words on a self-organizing map that serves simultaneously two purposes. First, it has organized different contexts to create a semantic landscape. 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 a 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 used was 19.

The map shows that those words in the two languages that have similar meaning are close to each 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 "hairdresser") and by two or more German words (for instance, "Arzt" and "Doktor" or "Friseur", "Friseuse" and "Damenfriseur").

The research on content-based information retrieval and analysis (see Chapter 7) provides a solid basis for future research in "translation through images". Some initial experiments show that names of concrete objects in different languages can be mapped with each other without any intermediate linguistic/symbolic representation (see [5] for details). In general, these results support the idea of symbol grounding [4].

References

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

[2] T. Honkela (2007). Philosophical Aspects of Neural, Probabilistic and Fuzzy Modeling of Language Use and Translation In *Proceedings of IJCNN 2007, International Joint Conference on Neural Networks.* Orlando, Florida, Aug 12–17.

[3] W. Quine, (1960). *Word and Object.* MIT Press.

[4] M. Sjöberg, J. Laaksonen, M. Pöllä, T. Honkela (2006). Retrieval of Multimedia Objects by Combining Semantic Information from Visual and Textual Descriptors. In *Proceedings of ICANN 2006, International Conference on Artificial Neural Networks*, pp. 75–83. Athens, Greece.

[5] M. Sjöberg, V. Viitaniemi, J. Laaksonen, T. Honkela (2006). Analysis of Semantic Information Available in an Image Collection Augmented with Auxiliary Data. In *Proceedings of IFIP, Conference on Artificial Intelligence Applications and Innovations*, pp. 600–608. Athens, Greece.

Chapter 14

Knowledge translation and innovation using adaptive informatics

Timo Honkela, Mikaela Klami, Matti Pöllä, Ilari Nieminen

14.1 Introduction

Knowledge translation can be defined as the process of supporting the uptake of research in a manner that improves the practices in the society and in industries through improved understanding, processes, services, products or systems. The term knowledge translation is used rather widely in health care: knowledge translation activities include: (1) research into the mechanisms of knowledge translation, and (2) evidence-based translation of knowledge (e.g., knowledge dissemination, technology transfer, knowledge management, knowledge utilization, synthesis of research results within a local or global context, development and application of consensus guidelines). Knowledge translation in any discipline requires a reciprocal relationship between research and practice. The goals of knowledge translation are to enhance competitiveness, innovation and quality of services and products. Adaptive informatics can provide tools for supporting knowledge translation and innovation.

14.2 Statistical machine learning systems as traveling computational models

In collaboration with the department of philosophy at University of Helsinki, we have considered statistical machine learning models from a more general level focusing on the question on what can be said about them as scientific methods and tools. This question has strategic interest when the use of these model is considered. Within science and technology studies, Dr. Tarja Knuuttila has for some time conducted research on studying scientific models as epistemic artifacts. Earlier her interest has focused on explicit models in computational linguistics such as parsers. Traditional linguistic models can be considered to be first-order models of language. Unsupervised learning methods, on the other hand, can be called second-order models: They do not model the phenomenon directly but through a process of emergence. In the following, the collaboration work is described more in detail based on [4] (see also [3]). We focus on neural network models and specifically on the Self-Organizing Map (SOM) [2]. The discussion, however, applies basically to any unsupervised statistical machine learning method.

Traditionally, it has been thought that models are primarily models of some target systems, since they represent partially or completely the target systems. Sometimes computational models not only have various roles within a scientific domain, but also travel across scientific disciplines. A travelling computational template is a computational method that has a variety of applications in different scientific domains [1]. Neural networks are good examples of such traveling templates. Initially, most of them were inspired by the idea of looking at brains as a model of a parallel computational device, but nowadays neural networks are applied in several different scientific domains, not all of which belong to the domain or neuroscience.

That a model can have so many and various applications, raises opinion some significant philosophical issues concerning the nature of models and how they give us knowledge. Both questions have been answered by philosophers of science by reverting to representation. On one hand models are considered to be representations, on the other hand they are thought to give us knowledge because they represent.

In the recent research in the philosophy of science it has been shown that neither isomorphism nor similarity can provide an adequate analysis of scientific representation. They lead to well-known problems. Firstly, the isomorphism view in fact assumes that there is no such thing as false representation, either the model and its target system are isomorphic, or then they are not, in which case there is no representation either. Secondly, both isomorphism and similarity are symmetric relations, which runs counter our intuitions about representation: we want a model to represent its target system but not vice versa. Both problems appear to be solved once the pragmatic aspects of representation are taken into account. The users' intentions create the directionality needed to establish a representative relationship; something is being used and/or interpreted as a model of something else. Taking into account human agency introduces also indeterminateness into the representative relationship: human beings as representers are fallible. Consequently, pragmatic approaches to representation solve many problems of the structuralist notion of representation-but this comes with a price. When representation is grounded primarily on the specific goals and representing activity of humans as opposed to the respective properties of the representative vehicle and its target system, nothing very substantive can be said about representation in general: There is nothing in the nature of the representation (the model) and its target system that guarantees the representational relationship between the two.

More importantly, even though the SOM were originally inspired by the neurophysiological structures of the cortex that does not explain their success in other domains. What is more, when SOMs are used in the fields of inquiry that lie quite afar from the neurophysiological research of the cortex, to conceive SOMs as representations becomes often vague. Instead, the SOM models reveal statistical structure in the data. To do this they rely on a "neurally inspired" algorithm, but this fact does not really make the SOM a representation of a neural representation of the domain of interest. If it represents anything, then it represents the data in a certain way. In this SOM models are alike simulation models on general, since often they are first and foremost appreciated for the output representations they produce. There was also a specific reason to consider the SOM as a sample of a neural network model rather than for instance, a backpropagation algorithm. Namely, as the SOM applies unsupervised learning paradigm, the end result of the analysis reflects relatively more the contents of the data than the supervised learning model sthat impose predetermined output categories on the analysis.

They should rather be conceptualized as multifunctional epistemic artifacts. More generally, the traditional philosophical view according to which models are first and foremost representations of some pre-defined target systems does not capture what seems to us the characteristic feature of modeling: the use of inherently cross-disciplinary computational templates.

References

[1] P. Humphreys. (2002). Computational Models. *Philosophy of Science*, vol. 69, pp. S1-S11.

[2] T. Kohonen. (2001). *Self-Organizing Maps.* 3rd edition, Springer.

[3] T. Knuuttila and T. Honkela. (2005). Questioning External and Internal Representation: The Case of Scientific Models. In L.Magnani (ed.) *Computing, Philosophy, and Cognition*, pp. 209–226. London: King's College Publishing.

[4] T. Knuuttila, A.-M. Rusanen and T. Honkela. (2007). Self-Organizing Maps as Travelling Computational Templates. In *Proceedings of IJCNN 2007, International Joint Conference on Neural Networks*, pp. 1231–1236. Orlando, Florida, Aug 12–17.

14.3 Modeling and simulating practices

In collaboration with Prof. Mika Pantzar (Helsinki School of Economics, on leave from National Consumer Research Centre), we have been developing a simulation model called *Pracsim*[3]. The systems demonstrates the basic concepts of practice theory. The theory is developed by Pantzar in collaboration with Prof. Elizabeth Shove (Lancaster University). In the theory, it is assumed that practices consist of three basic elements: material (materials, technologies and tangible, physical entities), image (domain of symbols and meanings), and skill (competence, know-how and techniques) [1]. Practices come into existence, persist and disappear when links between these foundational elements are made, sustained or broken: material, image and skill co-evolve. For instance, in the case of Nordic walking, Walking sticks are integrated to produce a proper Nordic Walking technique (linking material objects with skills). Furthermore, images of safety, fitness and nature can be integrated into the sticks themselves (linking image and material object) [2]. The basic motivation in considering practices as an application domain for adaptive informatics methods raises from its richness and complexity. The dynamics and conceptual content related to the phenomena of everyday practices sets a clear challenge for methodology development.

Theories on human action are often constructed either in such a way that the emphasis is on the social or on the individual level. Practice theory aims to build a bridge between these points of view. Based on the theory, Pracsim simulation system consists of two main parts: Simulation of practice dynamics and simulation of associated human population, the members of which adopt practices based on a variety of principles. Pracsim is an example of social simulation that refers to a general class of strategies for understanding social dynamics using computers to simulate social systems.

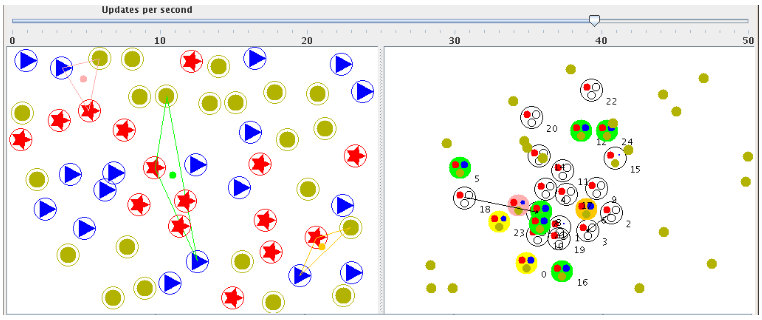

Figure 14.1: A screenshot of the Pracsim simulation. The symbols on the left hand side of the screen refers to the three different elements in the practice theory. The groups of three elements that are linked form a living practice. The right hand side of the screen includes a number of individuals, some of which have adopted a practice. Communication and diffusion within the community of agents is visualized with the links between the individuals. Small items in the “world” are instances of some material.

The pracsim simulation environment consists of a “world” in which a collection of items interact with each other. Following the practice theory, the items are either material, image and skill. Practices can be linked together into systems of practices. These systems are visualized by links between the participating practices. [3]

Pracsim system is applied in a Tekes project called Kulta that develops and applies methods that can be used in understanding, conceptualizing and anticipating the changing needs of consumers. The conceptual models of the practice theory are applied to analyze changes in the consumer society. These models are applied in the context of developing the business models of different kinds of companies. The data gained in this research and developing process are then analyzed and modeled using adaptive informatics methods. The research is focused on the strategic decision making within companies, and how new modeling techniques can be used in these processes. We have also been studied the usability of the modeling and simulation methods.

References

[1] M. Pantzar and E. Shove. (2008). *The Choreography of Everyday Life: Towards an Integrative Theory of Practice.* Forthcoming.

[2] E. Shove and M. Pantzar. (2005). Consumers, producers and practices: understanding the invention and reinvention of Nordic Walking. *Journal of Consumer Culture*, 1/2005, pp. 43-64.

[3] L. Lindqvist, T. Honkela and M. Pantzar. (2007). Visualizing Practice Theory through a Simulation Model. Helsinki University of Technology, Publications in Computer and Information Science, Report E9.

14.4 Analysis of interdisciplinary Text Corpora

We have presented means for analyzing text documents from various areas of expertise to discover groups of thematically similar texts with no prior information about the topics. These results show how a relatively simple keyword analysis combined with a SOM projection can be very descriptive in terms of analyzing the contextual relationships between documents and their authors.

Our analysis of text documents attempts to extract information about the area of expertise of the document using a set of keywords which are extracted from the documents automatically. To extract relevant keywords for each text document the frequency of each word is examined as an indicator of relevance. As the word frequency distribution for any source of natural language has a Zipf'ian form, special care has to be taken to filter out words, which occur frequently in all documents but are irrelevant in describing the topics. After the keyword extraction phase the documents are analyzed using a SOM projection of the keyword usage of the documents.

We selected two sets of documents from two distinctive fields of expertise: the first corpus A was collected from scientific articles published in the proceedings of the AKRR'05 conference with the topics of the papers mainly in the fields of computer science, cognitive science, language technology and bioinformatics. Corpus B consists of a collection of articles published by the Laboratory of Environmental Protection at Helsinki University of Technology with the topics ranging from social sciences to environmental managing.

Our experiments have shown that a combination of an automatic keyword extraction scheme combined with a clustering algorithm can be effective in describing the mutual similarities of text documents. Our statistical approach to the task has the additional benefit of being independent of the language that is being processed as no prior information about the processed language syntax is encoded into the algorithm.

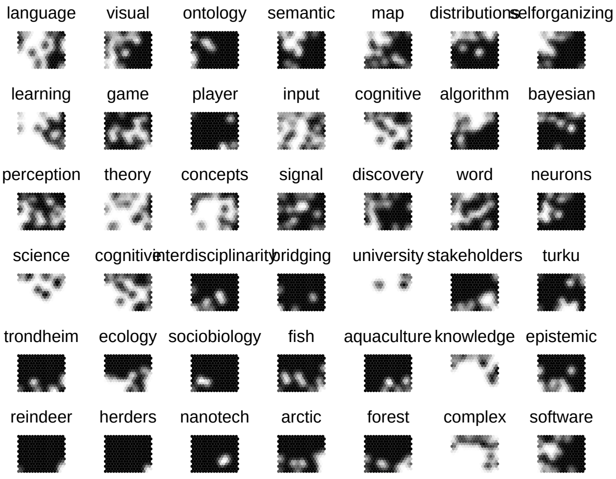

Figure 14.2: Component plane visualization of 42 keywords used in the analysis. Light shade corresponds to the occurrence of the keyword.

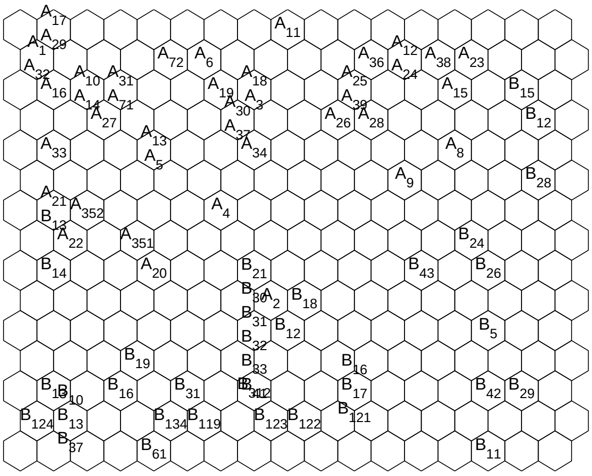

Figure 14.3: The documents of corpora A and B in a SOM projection. The two collections stem from the research of two different research groups which can clearly be seen in the clustering structure.

References

[1] M. Klami and T. Honkela. (2007). Self-Organized Ordering of Terms and Documents in NSF Awards Data. In *Proceedings of WSOM'07, Workshop on Self-Organizing Maps*. Bielefeld, Germany, Sep 3–6.

[2] M. Pöllä, T. Honkela, H. Bruun, and A. Russell. (2006). Analysis of interdisciplinary text corpora. *Proceedings of SCAI'06, the 9th Scandinavian Conference on Artificial Intelligence*, pp. 17–22. Espoo, Finland, Oct 25–27.

14.5 Quality analysis of medical web content

As the number of medical web sites in various languages increases, it is more than necessary to establish specific criteria and control measures that give the consumers some guarantee that the health web sites they are visiting, meet a minimum level of quality standards and that the professionals offering the information on the web site are responsible for its contents and activities.

We are a partner in a EU-funded project called MedIEQ that develops methods and tools for the quality labelling process in medical web sites. MedIEQ will deliver tools that crawl the Web to locate medical web sites in seven different European languages (Spanish, Catalan, German, English, Greek, Czech, and Finnish) in order to verify their content using a set of machine readable quality criteria. MedIEQ tools will monitor already labelled medical sites alerting labelling experts in case the sitesâ content is updated against the quality criteria, thus facilitating the work of medical quality labelling agencies. The overall objective of MedIEQ is to advance current medical quality labelling technology, drawing on past and original research in the area. Our work on automatic keyphrase extraction is used as a component of the MedIEQ AQUA system where relevant terminology about the content of medical web sites is used to facilitate the work of the human expert.

References

[1] T. Honkela, M. Pöllä (2006). Describing Rich Content: Future Directions for the Semantic Web. In *Proceedings of STeP 2006, Finnish Artificial Intelligence Conference*, pp. 143–148, Finnish Artificial Intelligence Society. Espoo, Finland, Oct 26–27.

[2] M. A. Mayer, V. Karkaletsis, K. Stamatakis, A. Leis, D. Villarroel, C. Dagmar, M. Labsky, F. López-Ostenero, T. Honkela (2006). MedIEQ -Quality Labelling of Medical Web Content Using Multilingual Information Extraction. *Studies in Health Technology and Informatics* vol. 121, pp. 183–190.