JYVÄSKYLÄ STUDIES OF COMPUTING 87

Steve Legrand

Use of Background Real-World Knowledge in Ontologies for Word Sense Disambiguation in the Semantic Web

Esitetään Jyväskylän Yliopiston informaatioteknologian tiedekunnan suostumuksella julkisesti tarkastettavaksi Agoran Auditorio 1:ssä joulukuun 17. päivänä 2007 kello 12.

Academic dissertation to be publicly discussed, by permission of the Faculty of Information technology of the University of Jyväskylä, in the Agora's Auditorium 1, on December 17, 2007 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2007

Use of Background Real-World Knowledge in Ontologies for Word Sense Disambiguation in the Semantic Web

JYVÄSKYLÄ STUDIES OF COMPUTING 87

Steve Legrand

Use of Background Real-World Knowledge in Ontologies for Word Sense Disambiguation in the Semantic Web

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2007

Editors
Seppo Puuronen
Department of Computer Science and Information Systems
Irene Ylönen, Marja-Leena Tynkkynen
Publishing Unit, University Library of Jyväskylä

URN:ISBN: 9789513930639
ISBN 978-951-39-3063-9 (PDF)

ISBN
ISSN 1456-5390

Jyväskylä University Printing House, Jyväskylä 2008

ABSTRACT

Legrand, Steve
Use of Background Real-World Knowledge in Ontologies for Word Sense Disambiguation in the Semantic Web
Jyväskylä: University of Jyväskylä, 2008, 73 p. (+ included articles)
Jyväskylä Studies of Computing,
ISSN 1456-5390; 87
ISBN 978-951-39-3063-9
Finnish Summary
Diss.

The purpose of this thesis is to show how word sense disambiguation (WSD) can be improved with background real-world knowledge encoded in ontologies and, especially, in ontologies based on psychological considerations. Ontologies are used, because conceptualized background knowledge is not available directly, from texts, to WSD systems. Although it is possible to disambiguate text to some extent without using ontologies, employing this kind of knowledge for WSD is of great help, especially in an environment like the Semantic Web, which has been the principal motivating factor behind this thesis. Some of the real-world knowledge, which is indispensable for human understanding, cannot be readily encoded in conventional ontologies either. One of the fundamental types of this kind of embodied knowledge is basic-level categories. After showing that conventional ontologies can be used to automatically group and label concepts in a text for disambiguation purposes with the help of self-organizing maps, the idea is extended to ontological structures based on basic-level categories. The thesis shows that the use of basic-level categories in WSD significantly improves accuracy. It also shows that linguistic phenomena, such as metaphoric expressions, can be manipulated structurally to reduce them to basic-level components with the potential to use them in WSD.

The approach used here proves fruitful and can be used as a starting point for designing an application that not only disambiguates using hybrid systems (including ontological real-world component) but also selects the best applicable disambiguation system for a particular word.

Keywords: word sense disambiguation, WSD, basic-level categories, real-world knowledge, background knowledge, Semantic Web, ontology.

ACM COMPUTING REVIEW CATEGORIES

I.2.7 Artificial Intelligence: Natural Language Processing:
Language parsing and understanding, Text analysis

I.5.3 Pattern Recognition: Clustering:
Algorithms, Similarity measures

Author

Steve Legrand
Department of Computer Science and
Information Systems
University of Jyväskylä
P. O. Box 35
40014 Jyväskylä, Finland

Supervisors

Prof. Pasi Tyrväinen
Department of Computer Science and
Information Systems
University of Jyväskylä, Finland

Prof. Alexander Gelbukh
Centre for Computing Research (CIC)
Instituto Politécnico Nacional, Mexico

Reviewers

Ass. Prof. Eneko Agirre
Computer Science Department
University of the Basque Country, Spain

Prof. Fred Karlsson
Dept. of Linguistics,
University of Helsinki, Finland

Adjunct Prof. Atro Voutilainen
Dept. of Linguistics,
University of Helsinki, Finland

Opponent

Prof. Barbara Lewandowska-Tomaszczyk
Dept. of English Language and Applied
Linguistics,
University of Lodz, Poland

ACKNOWLEDGEMENTS

I would like to thank my supervisors Prof. Pasi Tyrväinen and Prof. Alexander Gelbukh for their valuable and appreciated guidance and support during the course of this thesis. I also feel greatly indebted to Prof. Pekka Neittaanmäki, the head of Agora, for his valuable support, and Prof. Seppo Puuronen, the head of the COMAS[1] Information Systems Programme, for his excellent introduction to scientific work in general. COMAS (Jyväskylä Graduate School in Computing and Mathematical Sciences) provided the financial means for most of my investigation, and its courses helped in the orientation of my scientific work. I would also like to thank all my colleagues both in Finland, mainly at the Faculty of Information Technology in the University of Jyväskylä, and in Mexico at Instituto Politecnico Nacional and the University of Colima, where I undertook some of the crucial experiments related to this thesis. Many of the articles published during the course of the investigation were authored either with Harri Saarikoski from the University of Helsinki or with Rafael Pulido from the University of Colima.

My two stays in Mexico, which all in all spanned almost two years, were very eventful both in regards of my investigation and my private life. Prof. Pasi Tyrväinen allowed me free hands in selecting and pursuing my topic, which finally lead me to Mexico and to a meeting with Prof. Alexander Gelbukh, an expert in the area of my investigation, who became my second advisor. In Colima I met my present wife, Concepcion Ramirez, with whom I worked in the interface of Spanish and English languages related to Computational Linguistics. Without her understanding and unflinching support my stay in Mexico would certainly have been much poorer and less eventful. I also created strong professional ties with both Prof. Alexander Gelbukh and Dr. Rafael Pulido without whose cooperation and support I would not have been able to achieve what I did.

Overall, the process of writing of this thesis has enriched both my private and professional life more than I could have expected when I started.

Jyväskylä, December 7, 2007.

Steve Legrand

LIST OF FIGURES

LIST OF TABLES

CONTENTS

LIST OF ORIGINAL ARTICLES

I. Legrand, S., Tyrväinen P., Saarikoski, H. Bridging the Word disambiguation Gap with the help of OWL and Semantic Web Ontologies. In V. Pallotta, A. Todirascu (Eds.): Ontologies and Information Extraction International Workshop held as part of the EUROLAN 2003, Bucarest 28th July – 8th August 2003, 29-35.

II. Legrand, S., Tyrväinen, P. Connecting Lexical Knowledge to Distributed Real-World Knowledge, In T. Cameron, C. Shanks, K. Holley (Eds.): Proceedings of the Fifth Annual HDLS Conference, 1-2 November 2002, High Desert Linguistic Society, Albuquerque, New Mexico, 109-120.

III. Legrand, S., Pulido J.R.G. A Hybrid Approach to Word Sense Disambiguation: Neural Clustering with Class Labeling. In P. Buitelaar, J. Franke, M. Grobelnik, G. Paaß, V. Svatek (Eds.): Knowledge Discovery and Ontologies (KDO-2004) workshop in 15th European Conference on Machine Learning (ECML) Pisa, Italy, September 24, 2004, 127-132.

IV. Legrand, S. Word Sense Disambiguation with Basic-Level Categories. In Advances in Natural Language Processing. In A. Gelbukh (Ed.): Research in Computing Science. Vol.18, Mexico: IPN, 2006, 71-82.

V. Legrand, S., Structuring metaphors with basic-level concepts for word sense disambiguation. In J. Šķilters (Ed.): Baltic International Yearbook of Cognition, Logic, and Communication. Vol.2: Complex Cognition and Qualitative Science, University of Latvia, Riga, 2007, 171-188.

VI. Pulido, J.R.G, Herrera, R, Aréchiga, M., Block, A, Acosta, R, Legrand, S. Identifying Ontology Components from Digital Archives for the Semantic Web. In S. Sahni (Ed.): The IASTED Conference on Advances in Computer Science and Technology. Puerto Vallarta, Mexico, January 23-25, ACTA Press 2006, 7-12.

VII. Saarikoski, H., Legrand, S., Building an Optimal WSD Ensemble Using Per-Word Selection of Best System. In J. F. Martínez-Trinidad, J. A. C. Ochoa, J. Kittler (Eds.): Progress in Pattern Recognition, Image Analysis and Applications : 11th Iberoamerican Congress in Pattern Recognition, CIARP 2006 Cancun, Mexico, November 14-17, 2006. Lecture Notes in Computer Science. No. 4225. Berlin / Heidelberg: Springer, 2006, 864-872.

1 INTRODUCTION

When communicating, in any natural language, both the speaker (writer) and the listener (reader) must disambiguate words uttered (written), to correctly get their meaning. For example, a word such as *bat*, which has many senses, can homonymously represent an animal or a sport implement. To be able to understand and to make themselves understood people can use context and employ background real-world knowledge. This real-world knowledge is not coded in the language itself, but is available, for the communicating parties, in a conceptualized form.

The human ability to disambiguate is a very desirable skill for computer applications, as it can be used to help in information retrieval, translation, classification and categorization of documents and concepts and in many other tasks. The Semantic Web, following on the footsteps of the World Wide Web, is all about meaning. There we need an ability to extract the meaning of natural language to make machine-human interfaces more transparent to its users. This ability might be analogous to that of human beings who can disambiguate written text or spoken word in order to understand it. There are many computational applications, both within and outside the Semantic Web, which would benefit of automated word sense disambiguation. However, a lot needs to be done before a machine can rival a human being in this very important task.

The subject of the thesis deals with enhancement of automatic (unsupervised) word sense disambiguation (WSD) with the help of ontologies and basic-level categories which both encapsulate background real-world knowledge. This chapter introduces the motivation for the research, the main research objectives and methods, and presents the organization of the thesis.

1.1 Motivation for the research

Consider the sentence: "Yesterday, Peter was found dead in the street." If a person has been dead for 1 day only during the last 100 days and alive in 99 of them, what is the statistical chance of him being alive now? A human being can correctly surmise it being 0%, but a computer without the same background knowledge about the concept of death, might offer the figure of 99%. Even worse for the machine if this background knowledge were packaged within a metaphoric expression: "Peter kicked the bucket yesterday."

One may come across many similar situations, when trying to disambiguate a word by statistical methods based on training data alone. One inconformant and unpredictable real world fact can defeat statistics collected. For this reason, purely statistical methods have their limitations in WSD, and in fact they might have reached these limits already, witnessed by the diminishing returns in WSD accuracies obtained by such methods. This is why we need real world knowledge in addition to lexical knowledge for natural language understanding. Ontologies contain real-world knowledge; even a lexical database, such as WordNet, incorporates a great deal of real world knowledge in its assumptions, glosses and relations. Therefore, it makes sense trying to capture this knowledge through the use of ontologies and use it for natural language tasks such as word sense disambiguation. Like statistical methods alone, knowledge based methods alone have their limitations. However, the encouraging news is that by combining these two, we might still be able to improve unsupervised WSD disambiguation accuracies to reach a level in which they can usefully be employed by machines.

With the development of the Semantic Web (SW) we are going to delegate more and more tasks to SW agents that will act on our behalf in the information space and look after our interests there. Various data-formats and representation languages have been standardized and gradually developed to fit that new environment. However, based on the experiences with SGML and XML among other formats, it is unlikely that a majority of individuals or companies would take the trouble to learn fairly complicated formats such as XML and OWL (Web Ontology Language) in order to sense-tag all their web pages and build ontologies. Many of the web-pages would remain untagged in these circumstances. Currently, the only conceivable way to make sense of them, the way that a computer could use them without human intervention, would be to employ some kinds of natural language understanding agents as suggested by Java et al. (2006) that would use real-world knowledge instead of or in addition to OWL and other formats, whenever available, to correctly disambiguate any ambiguous terms.

The limitations of statistical methods in WSD and the emergence of the Semantic Web together with the need for real-world knowledge to be used for disambiguation by the natural language understanding agents are the motivation behind this research, the modest aim of which is to show that ontologies and basic-level categories can be gainfully employed to that end.

1.2 Objectives

The principal objective of this thesis is to show that word sense disambiguation can be improved by using hybrid WSD methods combined with background real-world knowledge encoded in ontologies (Article III) and basic-level categories (Article IV).

Ontologies are used, because conceptualized background knowledge is not available directly, from texts, to word sense disambiguation systems. Although it is possible to disambiguate text to some extent without it, employing this kind of knowledge for WSD is of great help.

Real-world background knowledge that can be used in WSD is often encoded and available in ontological relations. Nevertheless, some of the real-world knowledge, which is indispensable for human understanding, cannot be readily encoded in conventional ontologies either. One of the fundamental types of embodied knowledge is basic-level categories.

The second objective is to show that the idea has a universal dimension and is not dependent on any particular grammar or ontology, and for this reason different grammars and ontologies are used in the publications (I, II, III, IV). Also the methodology used in each case can vary.

The third objective is to create a base for a WSD system that combines both the best WSD systems for each word to be disambiguated and allows the use of real-world background knowledge in disambiguation while doing so (Article VII). This is to be the starting point for future research.

1.3 Methods and results

In this thesis we consider unsupervised WSD as an artificial intelligence tool, and view the Semantic Web as the context for the use of this tool. For this reason, the Semantic Web, although not elaborated to a great detail in the included publications, forms the general framework within which the research is situated. To be able to use a WSD application for Semantic Web purposes in an unsupervised manner, its present state of development needs some enhancements. We aim to do that with the help of real-world background knowledge present in ontologies and basic-level categories. The approaches used in this study are a combination of theoretical, constructive and experimental. Each constructive and experimental stage, which provide most of the results, are based on theoretical background study.

The results show the following:

1. The harnessing of ontologies for WSD does not seem dependent on the type of ontology used. In our research, the ontologies used include WordNet, and SUMO.
2. These ontologies can be used in connection with various grammar systems such as FDG, HPSG.
3. The use of ontology improves word sense disambiguation accuracy as it helps in labelling the concepts when used in conjunction with self-organized maps.
4. The use of basic-level categories improves word sense disambiguation accuracy.
5. The basic-level categories can also be used to structure metaphoric expressions making them better suitable for WSD.
6. It is possible to use self-organized maps also in the task of selecting the best WSD system for a particular word. This would make it easier to combine the WSD system selection and WSD in the same application.

2 RESEARCH BACKGROUND

This chapter describes the research background on word sense disambiguation, dealing with word senses, central terminology to WSD, historical background, State-of-Art, theoretical bases, and WSD evaluation.

2.1 Word Senses

What we mean by word senses cannot be clearly defined without first agreeing on what we mean by ambiguity, vagueness and other terms central to WSD.

2.1.1. Ambiguity and vagueness

If a meaning of a word, phrase or sentence has more than one interpretation, it is said to be ambiguous. There are many words such as *bank*, *bass*, *tank* etc. and phrases like *Finnish geography teacher* which out of context are difficult to interpret correctly. In word sense disambiguation, there are many different levels of ambiguity that need to be addressed. The most familiar of these are structural ambiguity and lexical ambiguity, which includes semantic ambiguity. Although the aim for WSD is to resolve the semantic ambiguity by allocating a word to its correct semantic class, or sense, structural disambiguation and wider lexical disambiguation is also needed for that task. Also, the senses of the word must be agreed upon before disambiguation can take place.

Vagueness does not give rise to multiple representations; rather, a vague expression is characterized by some inexactness, leaving details to be filled out. In vague statements, the falsity or truth cannot be established in such a manner that would satisfy all the interpretations. For example, if someone tells us that a *person is fat* or a *train is late*, we might have a different mental image depending on our cultural background, our notion of time etc. Also, it is possible, that the speaker might intentionally leave room for different interpretations, being unsure about the exact state of affairs. WSD does not address vagueness, therefore, although ambiguous words may involve vagueness.

2.1.2 Structural ambiguity

When a sentence or a phrase has two or more alternative syntactic representations it is structurally ambiguous. Jurafsky and Martin (2000) divide structural ambiguity into:
a) attachment ambiguity: *We saw the Eiffel tower flying to Paris*, (we flying / Eiffel tower flying);
b) coordination ambiguity: *old men and women* (old men and woman / old men and old women); and
c) noun-phrase bracketing ambiguity: *Can you book TWA flights?* (on behalf of TWA / flights run by TWA).

2.1.3 Lexical ambiguity

According to Jurafsky and Martin (2000) there are two main kinds of lexical ambiguity:

a) category ambiguity (part-of-speech): *I made her duck* (verb / noun)
b) semantic ambiguity (word sense): *I made her duck* (create / cook)

It is this word sense lexical ambiguity that WSD principally tries to resolve, and in particular, the semantic ambiguity. However, one needs to remember that before we can resolve senses on the semantic level we often also need to resolve the ambiguities on morphological, syntactic and other levels of NLP. Word sense disambiguation deals usually with homonymy and polysemy.

2.1.3.1 Homonymy and polysemy

Homonymous words, such as *bank* (river bank v financial institution) are often defined as words that have the same lexical form but different meaning, whereas polysemous words are taken to mean words such as *walk* (a physical activity, a place for walk) that mirror different aspects, including part-of-speech categories, of the same meaning. Nevertheless, the distinction between these two terms is often quite arbitrary due to the multiple ways that words with the same lexical form can be related, and in the case of word sense disambiguation does not greatly matter, except that homonyms are easier to disambiguate due to the greater sense difference in words in context. Sometimes homonymy is seen just as a special case of polysemy.

For polysemous words it is often hard to pinpoint the number of senses, and the more senses are identified the harder the disambiguation task becomes. Humans tend to use polysemous words more often than words that do not exhibit polysemy (Miller et al., 1993b). Often there is no real necessity to make very fine-graded sense distinctions for automated WSD, nor it is recommended, unless the task at hand specifically requests it. eXtended WordNet (Mihalcea and Moldovan, 2001) is a project to modify the original WordNet (Miller et al., 1993a) to make the sense distinctions more coarse grained so that more useful NLP tasks can be performed. In our experiments, we do not feel it necessary to make a clear distinction between homonyms and otherwise polysemous words; we treat them both as targets of disambiguation for both humans and machines. With WSD, it makes more sense to use the term 'ambiguity' to cover both of the terms. Humans can deal with that ambiguity quite easily, but for machines it has proved a major stumbling block in natural language understanding tasks.

2.1.3.2 Sense distinctions

It is often not easy to say how many senses a word has, and this has been admitted by many lexicographers (Malakhovski, 1987; Robins, 1987; Ayto, 1983). Some argue that a word has a slightly different sense in each context (Bloomfield, 1933) or is a mere function of distribution (Harris, 1954) forcing us to redefine what we mean by sense distinctions and the related assumptions in WSD. Most of the human experts agree on at least the coarser types of sense distinctions, but disagreements arise when finer distinctions are sought (Slator and Wilks, 1987). Dictionary definitions have been used for the purpose, but many of the dictionaries have been put together in an unsystematic way and have been found wanting. Cruse (1986, p. 55) listed three principles or rules to help in deciding whether a word is ambiguous:

> 1. If there exists a synonym or one occurrence of a word form which is not a synonym of a second syntactically identical occurrence of the same word in a different context, then that word form is ambiguous, and the two occurrences exemplify different senses. (*match* --> *lucifer* / *contest*).
>
> 2. If there exists a word or expression standing in a relation of oppositeness to one occurrence of a word form, which does not stand in the same relation to a second, syntactically identical occurrence of the same word form in a different context, then that word form is ambiguous, and the two occurrences exemplify different senses. (*light* --> *dark* / *heavy*).
>
> 3. If there exists a word which stands in a paronymic relation to one occurrence of a word form, but does not stand in the same relation to a second, syntactically identical occurrence of the same word form in a different context, then that word form is ambiguous, and the two occurrences exemplify different senses. (*race* --> *to race* / *racist*).

Before assigning words to senses through WSD, one needs, therefore, define the number of senses through which this ambiguity is expressed.

Schütze (1998) devised an algorithm in which the target words are grouped automatically into clusters representing a word sense. His method of word sense discovery which he calls word sense discrimination, does not name the clusters, however. Since then many other automatic word sense induction (WSI) methods have been proposed (Yarowski, 1995; Dorow and Widdows, 2003; Pantel and Lin, 2002). Our own method, which includes unsupervised word sense labelling (Article III) addresses this issue.

2.2 Real-World Knowledge in Ontologies

Ontology has two basic meanings. In philosophy it covers a very general topic: a study of being or existence and its ultimate relationships. In computer science the term has a different interpretation. According to Gruber (1993), ontology is an explicit specification of conceptualization. This definition was modified by Studer et al. (1998), who proposed that ontology is an explicit specification of shared conceptualization. The consensus element, *shared*, in the definition might be disputed on the grounds that a person can create an ontology only for his own purpose without sharing it. Shared conceptualization was later made explicit by Fensel (2001) to refer to understanding between people and heterogenous and widely spread application systems. From this it can be seen that the definition of *ontology* itself, as used throughout this thesis, is nothing absolute but also a shared understanding among people working in the computer science. Thus, we do not refer to its philosophical meaning when we use the term.

Simply put, ontology is a system of concepts connected with relations. Concepts and the relations between them can be represented symbolically in many different ways giving rise to different representation languages, ontological models, and tools to work with. The reason for this great variety of ontologies is that the majority of them are created for various purposes, and there is no specific ontology that would be the best suited for every imaginable task. Those ontologies which are not task-specific are called general ontologies or upper-level ontologies and subsume more specific domain ontologies. Like SUMO (Suggested Upper Merged Ontology)(Niles and Pease, 2001), they themselves are hybrids of many other more general ontologies. One of the best-known linguistically motivated ontologies is WordNet (Miller et al., 1993a) used also in WSD research. (See Section 2.4.2.1 for knowledge based disambiguation methods.)

The reason why ontologies might prove valuable knowledge sources for WSD is that in their relations between concepts there is a lot of real-world knowledge embedded, which can be tweaked out with certain methods and inference tools. For example the *part-of* relation can ideally construct an entity such as *car* from top to bottom, create a blueprint of it so to speak. Functional relations can be combined with structural relations. Inference engines can utilize this relational knowledge concluding, for example, that when a mechanic is working under a car and turns a wheel when repairing it, it is most likely one of the car's wheels underneath the car and not the steering wheel, helping to disambiguate the concept wheel. This might not make much difference for people reading an English text – the readers would understand it by the context – but a machine translator, translating it to Finnish, could just as well translate it to *ratti* or *ohjauspyörä* (driving wheel) instead of *pyörä* (car wheel). If the translation was to the opposite direction, from Finnish to English, Finnish car wheel, *pyörä*, might be translated to English *bicycle* without the help of such background real-world knowledge. For complicated machinery, such as cars, paper machines, and airplanes, these kinds of ambiguity induced errors in translated operation manuals can make them downright life-threatening to people who use the machines and consult their manuals. These examples abound.

The idea of this thesis, however, is not to concentrate on what is the background real-world knowledge that ontologies contain or can contain, but how we can employ that knowledge through the use of ontologies for WSD in SW. Nor will we dwell here on the structure of different ontologies, as there is a host of good textbooks for that (see Sharman et al., 2006; Staab and Studer, 2004; Simons, 1987). What we have found important, if we want to optimize any of the existing linguistically motivated or other ontologies for WSD, is to include concepts and relations that are also psychologically motivated, because cognition and language have a very close connection. The so-called basic-level categories (Section 2.4.1) and their relations are well studied but hardly applied to WSD.

Ontologies with conventional relations are useful in disambiguating sentences containing real-world knowledge such as:

If you do not give food to that animal it will just keep on barking.

where they can at least be used in inferring that the animal meant is *dog*. This can be based on a simple inference rule such as '*if dog is an animal, and dog is barking, then animal is barking*'. A conventional taxonomical ontology usually can offer only this subsumption relation (dog is-an animal), but the other relations are rarely available. But we also know that the dog is probably hungry and wants to eat something except, among other things, oranges, watermelons and fruits in general, we are irritated by the noise it makes etc. An automated WSD application usually needs to be told all this background knowledge explicitly. Functions like *eat* and *bark* are easier to find in basic-level relations which are not usually taken into account when building ontologies. Real-world knowledge, of which basic relations form part of, may be thought of as non-linguistic background knowledge that is not explicitly expressed in the communication, because it is assumed to form part of the common background knowledge for the participants in the communication. It includes common sense reasoning, general knowledge and facts about certain specialized domains among other things (Arnold et al., 1994). Although rarely done, it is possible to code these and any other types of relations to an ontology and use them subsequently for tasks such as WSD.

2.3 WSD: Some historical background

According to Wilks and Stevenson (1966) WSD is an "intermediate task", i.e., one of the necessary tasks in natural language processing applications such as machine translation, information retrieval, content and thematic analysis, grammatical analysis, speech processing, and text processing. Ide and Veronis

(1998) divide WSD into a) early WSD work in MT b) AI-based c) knowledge-based, and d) corpus-based methods.

The earliest WSD methods were used by machine translation in the 1950's, the most notable experimenter being Masterman (1957) who used Roget's Thesaurus in Latin-English translation for which most frequently employed thesaurus categories in a Latin sentence were exploited. AI-based methods, in vogue mainly in the 1960s and 1970s, suffered from the so-called "knowledge acquisition bottleneck" (Cullen and Bryman, 1988), but many of the present methods are based on these. There is a variety of AI-based methods, which can be roughly divided into symbolic and connectionist methods. The symbolic methods include a) semantic networks (Quillian, 1968), b) frames and networks (Hayes, 1977; Hirst, 1987), and c) case-based approaches (Wilks,1975, Boguraev, 1979), whereas connectionist methods consists mainly of spreading activation networks (Meyer and Schvaneveldt, 1971; Collins and Loftus, 1975; Anderson, 1983; McClelland and Rumelhart, 1981) and were influenced somewhat by Quillian's work.

There are many other ways to make these divisions and subdivisions. From the viewpoint of multiple WSD methods, therefore, our treatment of WSD here is, by no means, exhaustive. Apart from Ide and Veronis' (1998) and Escudero et al's (2002) articles on WSD in general, there are good textbooks, with sections on WSD, to complete the picture (Manning and Schütze, 1999, Ch. 17; Jurafsky and Martin (2000). A more extensive treatment of the subject can be found in a book by Agirre and Edmonds (2006) published more recently, and the first book to cover the entire topic of WSD. Most of these have sufficient historical background about the reasons leading to the present state of art in WSD for anyone interested.

2.4 WSD: State-of-Art

Nowadays, WSD techniques are usually divided in two main types: statistical supervised systems using tagged training data and unsupervised knowledge-based systems. Strictly speaking, however, completely unsupervised systems are not possible: without sense labeling these systems would merely discriminate the senses by clustering. Most of the practical applications are hybrids of some sort: unsupervised systems may make use of training data, and supervised systems may use lexical resources. Also, Manning and Schütze (1999) in their treatment of the subject offer a word of caution against simplistic classifications: according to them, the most important question to ask is: *What knowledge sources are needed for use of this method?* Nevertheless, one can make this rough distinction for comparative purposes, while keeping in mind that the methods form an uneven continuum from supervised to unsupervised methods. Here we discuss supervised and unsupervised WSD methods, on the one hand, and knowledge-based and corpus-based WSD methods, on the other. Another way to classify WSD methods would be to divide them into knowledge-based methods and corpus-based methods.

2.4.1 Supervised methods v unsupervised methods

Escudero et al. (2002), in their working paper, further divide the supervised methods into methods based on probabilistic models, similarity of examples, discursive properties, discriminating rules, methods based on rule authentication, linear classifiers and Support Vector Machines. Unsupervised methods can be similarly further subdivided. Therefore, our division here is on a very general level and made principally to exemplify the concepts.

2.4.1.1 Supervised methods

In supervised disambiguation labeled, feature-coded input is used to build a statistical classifier which can assign labels to new inputs based on feature similarity. Supervised disambiguation methods include Bayesian classifiers, decision lists and trees, neural networks, logic learning systems etc.

Naive Bayes classifier A Bayesian classifier takes a large context window and examines words around the target word. Each content word contributes some information for the identification of the target word sense. A naïve-Bayesian classifier is called naïve, because it makes the naïve assumption that the individual features used for description are all conditionally independent. It has been shown that for classification problems dealing with predicted categorical values the independence assumption need not to be strictly adhered to (Domingos and Pazzani, 1996; Domingos and Pazzani, 1997). Although the estimates of probability may be inaccurate, the classifier usually assigns maximum probability to the correct class. As regards WSD, the bag-of-words model used does not take into account any linear ordering or structural arrangement of the words in a sentence. The model also ignores any dependencies between the words in the bag. Despite these simplifications naïve-Bayes classifier has outperformed most of the other supervised disambiguation methods, as shown by Mooney (1996).

Given a sense, the probability of the entire vector consisting of the words in the bags is the product of the probabilities of its individual features, and the naïve-Bayes classifier selects the most likely classification V_{nb} given the attribute values a1, a2...an:

$$V_{nb} = \operatorname{argmax}_{v_j \in V} P(v_j) \prod P(a_i|v_j)$$

P(ai | vj) is estimated as

$$P(a_i|v_j) = \frac{n_c + mp}{n + m}$$

n = the number of training examples for which $v = vj$
nc = number of examples for which $v = vj$ and $a = ai$
p = a priori estimate for $P(ai \mid vj)$.

2.4.1.2 Semi-supervised methods

Semi-supervised methods are sometimes called weakly supervised, or minimally supervised, depending on the viewpoint.

Bootstrapping Yarowsky (1995) refers to his method as "unsupervised WSD", but many other authors prefer the term bootstrapping for this and similar methods, and regard them as minimally supervised methods, at best. Despite the fact that the WSD process in bootstrapping is, by and large, unsupervised, a very small number of so-called "seed" lexemes are needed to train the initial classifier, which can use any supervised method (Bayesian, neural nets etc.) for the classification task. Yarowsky used a decision-list classifier (Yarowsky, 1994) for his pioneering work.

In Yarowsky`s algorithm (Yarowsky, 1994), suitable collocations representing each sense are identified in the corpus. Only a small portion of the example corpus is tagged (perhaps 1% for each identified sense of the word to be disambiguated), leaving the bulk of the corpus as untagged residual. When a supervised classification algorithm is trained using these tagged seed sets, other collocations or features are identified which can be used to distinguish senses. The resulting classifier is applied on the entire sample set, the new sense-tagged examples from the residual set are added to seed-data, and the classifier is trained again. This iterative bootstrapping continues until most of the data is sense-tagged and only a small, stable residual remains.

There are some corrective procedures to Yarowsky`s algorithm (Yarowsky, 1994): one sense per discourse property which either augments collocations to specific senses or filters them, and an escaping mechanism from initial misclassifications whenever any later probabilistic evidence provided by the classifier demands adjustments.

Yarowsky (Yarowsky, 1994, 1995)reports results equaling the results achieved on the same data by supervised methods, questioning the need for large tagged corpuses. His results are better than those achieved by Schütze (1998) (see the Section below).

2.4.1.3 Unsupervised methods

When we talk about unsupervised disambiguation methods, we often really mean clustering, or word sense discrimination as in Schütze (1998). One of the first and the best known algorithm used in WSD that can be called unsupervised, EM (Dempster et al., 1977), tended to extract wrong patterns and had other problems as well. Recently, however, bootstrapping algorithms have improved to the point that it might be possible to talk about a "real" unsupervised WSD quite soon.

"Strapping" Building directly on Yarowsky`s bootstrapping algorithm (1995, ref above), Eisner and Karakos (2005) show that it is sometimes possible to eliminate the last bit of supervision in a bootstrapping method when building word sense classifiers for ambiguous words. They call their method strapping, relating it to terms such as bagging and boosting. It differs from bootsrapping in that instead of selecting the initial seed words, a number of plausible seeds are guessed, a classifier is built for each of them, and then a determination is made as to which classifier has grown the seeds given to it best. Unlike in Yarowsky's method, where the seeds are manually selected, in strapping it is possible to automate seed selection by using seed words that do not co-occur or do so very rarely but have high pointwise mutual information with their respective target senses. By using pseudowords (Gale et al., 1992c; Nakov and Hearst, 2003), the artificiality in seed selection is further emphasized, and manual tagging of senses becomes unnecessary. In practice, the method might not work for all ambiguous words, and mixed strategies involving the use of supervision might need to be used. Nevertheless the results are promising and are one more step towards wholly unsupervised disambiguation.

Context-group discrimination It was Schütze (1992) who coined the term 'word sense discrimination' to describe sense induction for his context-group discrimination method. As the disambiguated senses are not labelled, this is not regarded by many as fully-fledged WSD but rather as a very effective clustering method for word sense discovery. For some tasks, such as document queries, sense discrimination is an internal process, no sense labelling is needed, and the method is wholly adequate. Before strapping (see above) it was thought that this would be as far as unsupervised disambiguation methods would get.

In context-group discrimination senses are represented as clusters of similar contexts of the ambiguous words. A multidimensional vector space, Word Space, in which proximity equals semantic distance, is created for words, context and senses. The contexts of ambiguous words are assigned to the same cluster if the words with which they co-occur, occur with similar words in the training corpus, i.e., second-order co-occurrence vectors are created. The clusters are represented by their centroids, i.e., their average values. The second-order representation for the context of the word to be disambiguated is computed and then assigned to the cluster whose centroid is closest to that representation. As this process can create very high dimensional spaces, Schütze (1998) experimented with dimensionality reduction techniques, including singular value decomposition (SVD) with encouraging results.

Self-organizing maps Perhaps better known by its acronym, SOM, self-organizing maps is one of the most popular artificial neural network algorithm used in many different fields of science (judging by the number of SOM-related publications). Its use specifically on WSD has been scant; in rare occasions, though, its application has been directed to fields related to WSD. As its use is central to this thesis, it is described in more depth in Section 2.6.

2.4.2 Knowledge-based methods v corpus-based methods

Knowledge-based methods use external knowledge sources to help in defining senses, and corpus-based methods use statistical techniques and machine-learning to extract word usage models from large text collections. Knowledge-based WSD methods are normally considered supervised, although the knowledge sources can in theory be built either using supervised or unsupervised methods, and corpus-based either supervised or unsupervised. Yarowsky's (1995) bootstrapping method might be considered as a hybrid between these two approaches.

2.4.2.1 Knowledge-based methods

The most commonly used knowledge sources in knowledge-based WSD are machine-readable dictionaries, thesauri and knowledge bases of various kinds. The knowledge bases used include lexical databases (WordNet) (Miller et al., 1993a), domain taxonomies, and ontologies (Cyc) (Lenat, 1995). Sometimes it is not possible to make a very clear distinction between them - here they all are dealt with under the name ontologies. Historically, knowledge-based methods have their origins in AI. Gruber (1993) defined ontology as a formal explicit specification of a shared conceptualization.

WordNet. WordNet (Miller et al., 1993a) has been referred to as a lexicon, a thesaurus, a hierarchical digital dictionary and with many other terms. The term lexical ontology can be applied here as well, as we are dealing with a structure common to any ontology where concepts are joined with relations. The WordNet (2.1) defines itself as a *machine-readable lexical database organized by meanings* and, as a generic term (to cover other wordnets such as EuroWordNet (Vossen, 1999), Extended WordNet (http://xwn.hlt.utdallas.edu/) etc.) *any of the machine-readable lexical databases modeled after the Princeton WordNet*. One should emphasize here that this database is linguistically motivated.

After the appearance of WordNet 1.7.1 Harabagiu et al. (1999), among others, critizised its shortcomings: its lack of connections between noun and verb hierarchies, limited connections between topically related words, lack of morphological relations, absence of thematic relations and selectional restrictions, missing concepts and relations, and lack of uniformity in manual gloss definitions. WordNet 2 addressed some of these concerns, at least partially, but left some still unaddressed.

WordNet (http://wordnet.princeton.edu/) currently contains just under 150,000 words which are organized into 115,000 synsets forming a total of 203,000 word-sense pairs. The ontology is divided into four independent parts by part-of-speech: nouns, verbs, adjectives and adverbs. Concepts that are related are organized into sets of *synonyms* or *synsets* which act as the principal links within the ontology. Apart from being a member of a synset, nouns, verbs, adjectives and adverbs participate in other lexically motivated relations. In WordNet 2 there are nine separate noun hierarchies connected with hypernym relations (is-a-kind-of) as in "*a car is a kind of vehicle*", but there are other relations as well: *hyponym, meronym, holonym*, domain, familiarity, coordinate terms, and derivationally related forms.

For verbs, there are over 600 separate hierarchies, which means that their relatedness appears smaller than for nouns and that their hierarchies are much shallower. For this reason it is mainly its noun hierarchies that have been used in WSD. The main relation connecting verbs is *troponymy* (way-of-doing) as in "*saunter is a way of walking*".

Even though the principal authors behind the WordNet claim, with some justification, that it is based on psycholinguistic considerations (Miller et al., 1993a) and even though Miller treats basic-level relations at length in that publication, these relations have not been explicitly included. One might argue that part-of relations and glosses contain at least some of these basic-level relations. Nevertheless, the unsystematic nature of this inclusion leaves huge gaps and seems, thus, unsatisfactory. Another criticism directed against WordNet, by Lenat (Lenat et al., 1995) regards the number of relations used in WN insufficient for NLP tasks (less than 10), when compared with Cyc's (see below) approximately 1000 relations. Miller readily agreed with this in the same article. Miller's point is that while WordNet is a lexical database for English only, Cyc is a much more ambitions undertaking: a commonsense knowledge database adequate for any AI system.

The WordNet bibliography (http://mira.csci.unt.edu/~wordnet/) contains over 300 publications related to the use of this application in WSD and in other natural language processing task. The list is by no means exhaustive, and the number of WordNet related publications is growing constantly.

Mikrokosmos. According to Nirenburg et al. (2004), the growing popularity of WordNet and other well-known lexical knowledge resources, as witnessed by the related publications, is not altogether a good thing. They argue that the resources such as WordNet that are used by a majority of the scientists involved in WSD research are either unsuitable for machines, inaccurate, or low in vocabulary coverage, and that too much time and energy is being spent on trying to make use of less than ideal software applications for NLP-related task. Some of the problems, they claim, are due to the restrictive formalism of these applications rather than anything inherent in the language itself.

Their critical attitude did not, however, prevent the team from making use of WordNet when building their own resource, Mikrokosmos (Nirenburg and Raskin, 2004), to address the problem, which indicates that they do acknowledge the contributions other knowledge sources can offer. Mikrokosmos - recently dubbed to OntoSem - is specifically tailored for machine translation. As MT involves WSD, Mikrokosmos/OntoSem can also be seen as a potential contributor to WSD. Ambiguity resolution is, in fact, seen as its major challenge and major contribution. As a lexical database aimed at machine translation between English and Spanish and with a great amount of real world knowledge included, Mikrokosmos/Ontosem can be situated somewhere between WordNet and Cyc in its ambitions.

The input for OntoSem is unrestricted raw text. After tokenization, morphological, syntactic, and semantic analysis are carried out. The end result is in the form of text meaning representations (TMR's) (Nirenburg and Raskin, 2004). Most of the lexical and real world knowledge that TMR's rely on must be manually acquired, a clearly labor intensive task which the authors justify by comparing their efforts with those designed to adapt existing a priori inadequate lexical knowledge sources to NLP tasks. The estimated time-frame, 8 years, to create a usable system, is seen as reasonable (Wordnet took about 20 years to reach its present shape, Cyc has projected an even longer time-frame).

In the version developed so far, TMR's play the central role. The difference in concept representation between OntoSem TMR's and WordNet hierarchy is highlighted with the verbs of change such *as increase, bend,* and *buckle*. While each of these have different superordinate nodes in WordNet (*change magnitude, change shape, and change surface*, respectively) in order to create a semantic differentiation between them, their common ancestor node remains *change*. In OntoSem the semantic differentiation is accomplished with the help of THEME: in the case of the verb *increase*, for example, if the theme is *cost*, then the precondition, effect, and time (components of the semantic differentiation) can be set accordingly.

The cost, thus, is more at the end of the specified time than in the beginning of it, i.e., it has *increased*. The latest version of WordNet has added the category of derivationally related forms for verbs which can be used much to the same effect: if the verb is derived from sense x of a noun, then the theme can in many cases be deduced. The simple example above does not, however, fully cover the rich semantic structure of OntoSem.

SUMO. What makes SUMO (Suggested Upper Merged Ontology) interesting is that one of its aims, from the outset, was to facilitate automated natural language understanding in the WWW (Niles and Pease, 2001). With the birth of the Semantic Web, SUMO was seen as one of the most promising tool for semantic web agents for utilizing and making sense of their environments (Pease et al., 2002, Subrata et al., 2002) and also has motivated this research.

As its name suggests, SUMO is designed to act as an upper ontology for more specific domain ontologies, thus enabling the unification of many disparate domain ontologies that may emerge in SW. SUMO, indebted greatly to SUO (Standard Upper Ontology), is itself a result of merging many types of foundation ontologies including John Sowa's (2000) upper level ontology, Russel and Norvig's (1995) upper level ontology, James Allen's (1984) temporal axioms, and many others.

SUMO, together with its Mid-Level Ontology (MILO) and domain ontologies, forms probably the largest freely publicly available formalized ontology today. It uses a version of KIF (Genesereth, 1991), SUO-KIF (http://suo.ieee.org/SUO/KIF/suo-kif.html), as its knowledge representation language.

The SUMO hierarchy was found to be reasonably compatible with the WordNet hierarchy and has mappings to the current and earlier versions of WordNet. In fact, the main authors behind SUMO (Niles and Pease, 2003) among others (Ahrens et al., 2004) have been active in promoting this nexus. While SUMO seems more comfortable with WordNet than Mikrokosmos, its authors have levelled criticism against another well-known ontology, Cyc, to counter the charge that there is no need for another fully fledged comprehensive ontology (Niles and Pease, 2001). According to them, Cyc retains proprietary rights to most of its ontology, no extensive peer review has taken place as a result, and only a small part of its ontology has been released to the public.

Cyc. Like SUMO, Cyc (Lenat, 1995) has integrated or mapped several existing ontologies to its knowledge base, including SENSUS, FIPS 10-4, pharmaceutical thesauri, WordNet, MeSH/Snomed/UMLS, and CIA World Factbook, and like SUMO, Cyc community has shown interest in the Semantic Web. Therefore, some rivalry between these two is understandable. Disparate ontologies in SW, often poorly equipped with semantic and other relations, need a semantically enriched knowledge base and an ontology that can unify and make sense of them. This is what can be accomplished with Cyc, according to its creators (Lenat, 1995). However, the aims of Cyc are much more ambitious than those of SUMO, and apart from SW they deal with natural language understanding, distributed AI, and intelligent search, among others. For this reason, the team behind it prefers to refer to it as a vast common sense knowledge repository rather than ontology, although the meaning of the word "common sense" is still far from being clear as pointed by Sowa (2002). The ambitious approach, the money and time spent on the project, and expectations not realized in projected time-span have formed the main point of objection among Cyc's opponents. It is fair to say, however, that there has been steady progress, and many of the problems perceived might be addressed in the foreseeable future.

Cyc contains now over 300 000 concepts (http://suo.ieee.org/email/msg08310.html) which are language independent, although there is an English lexicon mapped to the knowledge base. The relations within this common sense ontological structure employ facts (about a million of them) and explicitly stated assertions (three and half a million) which, although locally consistent with the help of microtheories, may remain globally inconsistent until (if ever) consolidated. Cycl (http://www.cyc.com/cycdoc/ref/cycl-syntax.html), a higher order language based on predicate calculus, is used for knowledge representation. With Cycl, any reasoning about the properties of collections, relations, and Cycl sentences can be declaratively represented. The complexity of structures this approach creates and the necessary training and skills needed to enter facts and make use of Cyc were long seen as its weak points. This has improved with time, and one can now augment the Cyc knowledge base with a simple question-answering system, in a form of a game of trivia. As from April, 2004, Cyc has announced the release of its latest OpenCyc version containing the full Cyc ontology, thus doing away with some of the criticism levelled against for its proprietary policies.

2.4.2.2 Corpus-based methods

Most of the corpus-based methods are supervised, i.e., learning from previously sense annotated data. The supervised learning sources include tagged and disambiguated raw corpora, and bilingual parallel corpora, or a combination of monolingual corpora and bilingual dictionary, where the senses can be identified from their translations (Gale et al., 1992b). Pseudowords have also been used to avoid manual tagging (Gale et al., 1992c; Schütze, 1992). Recently, the use of WWW as a huge distributed corpus has drawn wide attention: Meaning-project (http://www.talp.upc.es/TALPAngles/ index.html) enriches EuroWordNet and uses its tools to collect examples from the Internet in order to create a Multilingual Web corpus, a semantically annotated corpus for each wordnet word sense, containing concept and domain labels. This will also address data scarcity that is evident even in bigger, conventional type corpora. Moldovan and Mihalcea (2000) are some of the researchers who have used Internet for WSD. Of the many search engines on the Internet, Google has been used by Klapaftis and Manandhar (2005), among others, for the same task.

2.4.2.3 Selectional restrictions -based disambiguation

Jurafsky and Martin (2000) consider selectional restriction-based disambiguation to be separate from the other types of disambiguation in that the disambiguation using selectional restrictions is integrated in semantic analysis while in the other types the semantic analysis is performed separately.

2.5 Cognitive Linguistics and its Potential for WSD

It may be argued that all the approaches to WSD are cognitive to a certain extent. No-one can deny that there is a very strong connection between our language faculty and cognition. We could argue, just as well, also that all WSD is statistical by nature: we learn by repeated occurrences of the word, and knowledge bases are built on the knowledge of those occurrences. Or, equally, we could generalize by saying that all WSD is knowledge-based – to a certain extent, that is. The difficulty of defining the word "WSD", demonstrates the task dependency of sense definition that Kilgarriff (1997) describes.

Cognitive linguistics is relevant to ontological enrichment in WSD in that it calls for encyclopedic knowledge to be incorporated into lexical knowledge. According to Geeraerts (1988), it is necessary to study lexical concepts as a part of human cognition in general, and not as an autonomous language structure within human cognition as Chomsky (1975) and other structuralists have maintained. This view was supported by Lakoff (1987) and Langacker (1987), although Lakoff's Idealized Cognitive Models do not seem very closely related to Langacker's Cognitive Grammar, and vice versa. Other theories related to cognitive linguistics abound, the most notable of which are Jackendoff's (1990) Conceptual Structures, Fauconnier's (1985) Mental Spaces and Gärdenfors's (2000) Conceptual Spaces. Although these and other cognitive linguistics theories are hard to integrate, there are some commonalities and influences between them: Langacker's conviction of the existence of encyclopedic semantics was reluctantly embraced even by Fillmore (1982) who resisted it until having completed the formulation of his own theory of Frame Semantics. Fillmore's frames, together with Langacker's and Fauconnier's ideas, influenced then Lakoff's Idealized Cognitive Models.

In spite of its rapid acceptance the application of Cognitive Linguistics to WSD has been rather muted, implicit rather than explicit. One of the reasons for this might be that unlike structuralists' theories which are more straightforward to code for computers to understand, the concepts of Cognitive Linguistics are rather hard to put into codeable structures. One needs to look at the basics of cognitive theories in order to discover those psycholinguistic structures that may be useful for WSD. This is what we have done in the Articles IV and V of this thesis.

2.5.1 Basic-Level Categories and the prototype theory

Underlying many of the theories in cognitive linguistics or contributing to them are the basic-level categories. They are thought to be situated in the middle of hierarchically structured levels of human conceptual system as follows (Rosch, 1973, 1988):

1. Superordinate level (*furniture*)
2. Basic-level (*chair, table, lamp*)
3. Subordinate level (*kitchen chair, living-room chair / kitchen table, night table / floor lamp, desk lamp*)

Robert Brown (1958) referred to categories falling somewhere between the most general and most specific level as "first-level" categories. He noticed that these are the categories that allow children to learn object categories and name them. Later Rosch et al. (1976) and Rosch (1988) experimented with basic-level categories, as they became known, and were able to find out some of their specific properties. Murphy and Medin (1985), explaining why people prefer to use basic-level concepts, offer what they call a differentiation explanation, which has two main components:

1. *Informativeness*: when compared with subordinate and basic-level categories, superordinate categories are less informative. They contain some common features but lack details. Basic-level and subordinate categories supply the major bulk of the detailed information.
2. *Distinctiveness*: On the basic-level, the categories sharing the same superordinate level features, do not share many of the features emerging

on the basic-level (*tables* and *lamps*, for example). Sub-ordinate categories, on the other hand, add some extra features but, at the same time, share many of these basic-level features, and are, therefore, harder to distinguish from each other (*floor lamp*, *desk lamp*).

Superordinate categories may, according to Rosch (1973) be represented by a stereotypical prototype, which can be demonstrated as a tendency for selecting a certain representative of the whole group within a basic-level category. For example, when asked to pick up a typical representative of the category *furniture*, most of the people select *table* rather than lamp. Rosch's ideas are based on Wittgenstein's (1953) notion of *family resemblances* and Labov's (1973) observation that the classical Aristotelian theory of categorization requiring *necessary and sufficient features* does not always seem to work: a cup can be classified either as a cup or as a bowl depending on its size and/or whether it has a handle (Smith and Medin, 1981), but a prototypical cup is a cup the classification of which is not seen as controversial.

Archambault et al. (2000) selected the following as the most important issues to note about basic-level categories:

- Basic-level categories can be verified fastest.
- Naming of objects is faster at the basic than at the subordinate level.
- Objects are preferentially named with their basic-level names.
- Basic-level names are learned before subordinate names.
- Basic-level names are usually shorter.

Although well-known now for about a half-a-century, basic-level categories have received scant attention in WSD research. Some of the lexical ontologies such as WordNet (Miller et al., 1993a) and Microkosmos (Nirenburg et al., 2004) have incorporated some of the features of basic-level categories in their ontological structures. However, it is hard to use these incompletely coded structures in WSD, and some modifications are needed.

2.5.2 Idealized Cognitive Models

Lakoff (1987) believes that linguistics categories show prototype effects and can be demonstrated to have basic-level categories, but neither he nor Rosch (1973, 1988) advocate the view that basic-level categories would explain any structural or procedural properties of cognition. Instead, they both regard basic-level categories as a mere surface phenomena related to cognition, i.e., below that surface there may be some other more interesting structures and processes to be found.

According to Lakoff (1987) knowledge is organized by means of structures to which he refers to as idealized cognitive models, or ICMs, and that category structures and prototype effects are their by-products. Each ICM is regarded as a structured whole, a gestalt, employing four structuring principles:

- propositional structure (Fillmore's (1982) frames)
- image-schematic structure (Langacker's (1987) cognitive grammar)
- metaphoric mappings
- metonymic mappings

Lakoff's examples of ICM's include a Balinese calendar system with three different "week" structures superimposed (Geertz, 1973), and the category defined by the word *bachelor* (Fillmore, 1982). Fauconnier (1985) describes how these ICMs structure the mental space.

Lakoff (1987) how, by extending the basic-level categories to the linguistic domain, we can end up with novel categorical structures. This is of great importance to this thesis, because these particular categories might have a crucial role to play in WSD. It is still to be seen whether it is for this reason why conventional ontologies alone have so far proved inadequate for linguistics tasks such as word sense disambiguation.

2.5.3 Conceptual Spaces

Gärdenfors (2000) proposes a three-level framework for representing information for a human cognition, in which the conceptual level would act as a bridge between symbolic and connectionist approaches:

1. Symbolic level (propositional representation)
2. Conceptual level (geometric representation)
3. Sub-conceptual level (connectionist representation)

Conceptual spaces around the conceptual level are supposed to bridge the symbolic and sub-conceptual levels, allowing concept learning and aiding communication about concepts. These conceptual spaces are depicted as multidimensional geometric structures representing concepts and properties in their dimensions. Some of these dimensions, like hue, chromacity and brightness for color, would be grounded on human perception, some would be more abstract. An integral set of dimensions, such as color, in which the dimensions are co-dependent, would form a domain. Concepts, in turn would be grounded on these domains.

For example, some of the dimensions of an *apple*, which is a point in a conceptual space, would be color, texture, shape, weight, etc. The domains' salience when dealing with a concept would be dependent on the task or purpose the concept was invoked. If we played with the apple, the salient domain would be *shape*; if we ate it, the *taste* domain would be the most prominent, and when peeling the apple *texture* might appear the most important domain.

This has got important consequences for WSD on two counts. First, as Gärdenfors (2002) himself points out, although the concept itself might not represent the exact same point in the conceptual space in the minds of two people, it resides within the same domain so that communication becomes possible. As the conceptual spaces do not exactly correspond either, the result is an approximation comparable to prototype in the prototype theory.

Second, the preference for the use of adjective to differentiate a category member from the prototype instead of subcategorizing it gives some support to the idea of basic-level categories (see Section 2.5.1) and to their memory economizing function, and a possible mechanism for how to tell the difference between a basic-level category concept and subordinate concepts.

2.5.4 Conceptual structures

According to Jackendoff (1990), conceptual structures are structures of mind that, during evolution, developed in organisms in interaction with their environmental conditions. Conceptual structures are supposed to find their expression in natural language through semantics and partially also through syntax, and play an equally important function in vision, hearing and in other faculties. These structures are thought to be universal; even though languages vary, all human beings are the end result of a broadly similar evolutionary development. The presumed universality of these structures gives some hope that they can be found in all the spoken languages, and could be of immense value in many natural language processing tasks such as translation.

Conceptual structures (Figure 1) form an interface between linguistic structures and other autonomous structures such as vision etc. with the help of the rules of inference (including all sorts of heuristics among other things). Each of the three levels of structure (phonological, syntactic and conceptual) has its own primitives, and follows its own combinatory principles and organization

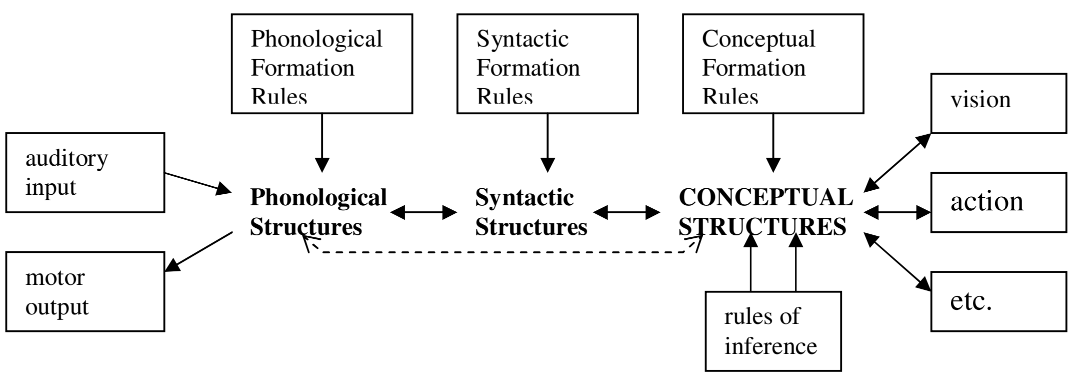


FIGURE 1: Organization of mental information (after Jackendoff , 1990)

into subcomponents. Each level is governed by a set of appropriate well-formedness rules. The structure also has correspondence rules linking the levels: for example, a re-segmentation of a sentence between phonological and syntactic structures is affected by these correspondence rules.

According to Jackendoff (1983), each major syntactic constituent of a sentence is mapped into a conceptual constituent in the meaning of that sentence. For example, in *Jane walked to the store*, *Jane* and the *store* are mapped to THING constituents, *to the store* to a PATH constituent and the sentence as a whole to an EVENT constituent. The main function-argument categories are Thing, Event, State, Action, Place, Path, Property, and Amount.

Although disputing Schank's (1972) "conceptual cases" in their representational aspects while sketching his frame theory, Minsky (1974) seem to have understood well the implications of evolutionary development of conceptual structures in his seminal paper on knowledge representation. His criticism of the inadequacy of predicate logic in semantics is echoed by Jackendoff's (1983) use of Preference Rules, which are assumed to incorporate features from not only one but many logic systems. It seems that conceptual structures may incorporate redundancies and unnecessary parts comparable to those in a living organism (the appendix, fishy characteristic of a human embryo, etc.) reflecting their evolutionary development. Minsky applied his theory to vision as well as to language understanding. Jackendoff (1997) found compatible structures in music, and treats linguistics as just one of the interfaces to the conceptual system in his Tripartite Parallel Architecture.

2.5.5 Cognitive Grammar

Cognitive grammar (Langacker, 1987) assumes that grammatical structures are not autonomous and that language is not self-contained but always needs to refer to cognitive processes. This is in clear contrast with the Chomskian tradition (Chomsky, 1975), which regards the language faculty as an autonomous entity, notwithstanding the subsequent amendments to the theory. Instead of relying on deep structures, cognitive grammar uses construals of grammatical units to explain variations in sentence structures. Langacker (1987) speaks of an inventory of conventional units to indicate that grammar is nongenerative and nonconstructive, and not an algorithmic device outputting well-defined classes of expressions.

The existence of some basic domains such as *time*, *color*, and others is posited. To be able to completely analyze meaning, a complete account of developmental cognition is needed. A full characterization of a semantic structure must include a description of its domain, and, further, the description of the entire hierarchy of more fundamental conceptions on which it depends. For example, to be able to conceive and understand the notion of *Monday*, one must be able to account on its domain *Week*, and so on. To be able to understand the word *elbow*, one need to consider its domain *arm*, then a higher domain, *human body*, etc. Most predications need a complex matrix of domains for their full description.

Predication-domain relationship gives rise to profile and base in conventional imagery. The base, arm, would profile the predication, elbow, and the semantic value would reside in their inter-relationship. Predicates are usually represented as relations between trajectory and landmark the dimensions being time and space in the case of verbs (Langacker, 1987). Three types of basic units are proposed: semantic, phonological, and symbolic. The symbolic unit is said to be bipolar, with the semantic unit (of indefinite semantic complexity) in one pole, and the phonological in the other.

As an example, for a lexical unit *table* a symbolic unit [[SEM]/[PHON]] would be formed as a combination of a semantic unit (TABLE) and phonological unit (table) as follows: [[TABLE]/[table]]. A grammatical rule or construction would be represented by a symbolic unit that is at the same time complex and semantic: *driver* would be represented as [[[PROCESS]/[Y]]-[[ER]/[er]]] composed of the verb schema [[PROCESS]/[Y]] and the grammatical morpheme [[ER]/[er]]. Simpler expressions can be used as building blocks for more complex expressions to cover the whole field of English grammar.

2.6 Self-organizing maps (SOM) and WSD

For humans, visualizing and, subsequently, understanding high-dimensional data is not easy. SOM is a grouping and visualization technique that reduces the number of dimensions enabling the display of grouped data in two dimensions. Today there are many fields of science in which SOM is used as a standard tool for solving various problems. Its use in physics, finance, medicine, chemistry and statistics, to mention only a few, indicates its adaptability to a great variety of tasks in many different fields. It is still considered as one of the best models of the function of the brain, and that may partially explain its adaptability.

SOM (Kohonen, 2000) may be thought of as a model of unsupervised learning (see Section 2.4.1.3) and an adaptive knowledge representation scheme at the same time. Basically it works by adaptation of the model vectors, in which the weight vector of a neighbourhood of neurons is updated at each iteration based on a unique sample.

2.6.1 SOM in linguistic tasks

SOM has been used in linguistic tasks also, but is less known among the linguists than many other techniques. Ritter and Kohonen (1989) experimented with two datasets: one with the features added and another one with the features obtained from a local context. They called the resulting two-dimensional maps *self-organizing semantic maps*. These maps clearly show that with SOM it is possible to cluster words into distinct cohesive two-dimensional regions within which the items show psychological and linguistic affinity. Probably because the publishing platform of their paper was oriented towards biology they sought and discussed parallels between the topographies of the maps and the brain. Their results were qualified with a caution that any realistic semantic brain maps would need a much more complicated, probably hierarchical model.

In this thesis we demonstrate a hierarchical model, in which an ontology is used to supply category labels to unlabeled SOM clusters and to position those clusters, after labeling, in a hierarchy (Article III). We have managed not to increase the complexity of the SOM model but have kept it unchanged. We do not believe that there is enough evidence to show that the hierarchy, at least the one created by our method, thus obtained necessarily reflects the arrangements of the so-called semantic brain. By obtaining the category labels automatically from an ontology, we aimed to keep the SOM unsupervised and thus suitable for applications such as those needed in the Semantic Web where manual tagging of concepts is not feasible in a larger scale.

Apart from the work of Teuvo Kohonen, whose work has inspired this thesis, there have been some other noteworthy applications of SOM in the field of linguistics. Perhaps the best known of these is WEBSOM of Honkela (1997) which is based on the work of Ritter and Kohonen (1989) described above. The WEBSOM method organizes documents into two-dimensional maps in such a way that related documents are clustered together. The word forms in the documents are first organized into a word category map enabling the documents to be encoded based on the similarity of the word meanings. These encoded documents can then be processed by SOM to produce a document map, which provides a general view of the document collection. The labels of the obtained clusters are derived from the documents and thus the classification is not disciplined in the way that it could be used in WSD tasks as such, a problem that our use of sense related ontology for cluster labeling addresses (Article III).

A further interesting development in the use of SOM for linguistics is that of Linden (2005), who in his thesis presents a WSD method based on a WEBSOM map of a patent abstract collection (Kohonen et al., 2000). In Linden's THESSOM method single words are treated as small documents. A previously created WEBSOM map is seeded with word sense vectors, i.e., manually tagged words in an ideal context to express their sense. Then the distance from these ideal word sense locations to the word to be disambiguated can be used to determine the correct sense. The location on the semantic space for both the seeds and the words to be disambiguated is determined by their context. As the idea was tested with a WEBSOM map created for another purpose and domain, it makes the relatively modest results achieved in these non-optimal conditions quite remarkable. Whether this method should be called supervised or unsupervised is a moot point and depends on the definition.

There are now over 5000 publications dealing with SOM (Cottrell and Verleysen, 2006), quite a few of them relevant at least indirectly (cognition, AI) to linguistics.

2.7 WSD evaluation

What makes the evaluation and comparison of WSD results problematic is the heterogeneity of knowledge sources and methods among others. Some of the research deals with a single word, some use a 400 million word corpus as material. The multiplicity of linguistic and cognitive theories within which WSD can be placed often guides the methodology. As a scientific discipline WSD has been subsumed by other disciplines including AI, and it is very recently that it has established itself as a separate discipline within Computational Linguistics.

Recently, there have been strong signs that the research community wants to establish some standard ways to make result comparisons possible.

The principles that Cruse (1986) formulated for deciding whether a word is ambiguous (see Subsection 2.1.3.2) was one of the most notable attempts to establish a standard way to evaluate ambiguity, among others (Zwicky and Sadock, 1975; ten Hacken, 1990, Geeraerts, 1993). Nevertheless, a standard sense inventory for ambiguity resolution seems an eluding goal: sense distinctions depend on tasks and purposes (Kilgarriff, 1997), and new word meanings appear daily. Kilgarriff critizised these tests for dealing mainly with word senses that are clearly distinguishable from each other to demonstrate the contrasting outcomes. According to him, many ambiguity criteria could be criticized on these same grounds, and what was really needed was a criterion which would address the cases where the intuitions were not clear. Also, even widely used sense inventories such as WordNet have been found wanting for WSD due to its too fine sense distinctions (Mihalcea and Moldovan, 2001).

Gale et al. (1992a) proposed lower and upper bounds to evaluate WSD systems' performance and to overcome the problems by human judges due to subjectivity in their assessment. The estimation of the lower bound should ignore the context and assign the most frequent sense to all occurrences of the word. The upper bound, in which there is considerable disagreement between human judges, should always take into account that disagreement.

Senseval (Kilgarriff, 1998; Kilgarriff and Rosenzweig, 2000), which partially resulted from the observations by Resnik and Yarowsky (1997) on the state of WSD, was established to evaluate the performance of WSD systems on manually tagged Gold standards designed for the purpose. Up to the time of writing, three Senseval events have taken place providing a great deal of valuable new research material. The datasets used are available for public use, so that whosoever is interested in testing a WSD system against them can get an idea where the application stands in comparison with the others tested.

3 RESEARCH PROBLEM AND METHODOLOGY

This chapter describes the research problem and the objectives of the research. The research methods are also described here.

3.1 Research Problem

The accuracy of WSD applications based on mainly statistical solutions is leveling out, and new ideas are needed to improve WSD accuracy for applications such as the Semantic Web where the SW agents rely heavily on the use of ontologies for making sense of their environment. Ontologies of various kinds, including those in the SW, are considered as a ready source for background real-world knowledge. This state of affairs gives rise to the general reseach question of this study:

> *How can we improve unsupervised WSD with the use of background knowledge incorporated in ontology relations?*

As the information in the SW is scattered and often unorganized, one needs first to make sure that it is possible to group and organize pieces of information in such a way that they can be matched against their correspondent ontologies, the relations of which the background real-world information can be derived from. We must be able to do that in an unsupervised manner, because the SW agents must be able to act independently. The research question posed as the result of this consideration was:

How could one organize information that is scattered and unorganized, grouping and labeling it so that it could be matched against a suitable ontology for WSD or other purposes?

All background and real-world knowledge does not reside in conventional ontologies. These are after all meant mainly for computational applications. Apart from internalized ontologies, human beings can derive that knowledge from many other sources including pragmatic situations or cognitive predisposition. One such source, embodied knowledge demonstrated in basic-level categories, is based on the related psychological and cognitive theories. Taking thus into account that conventional ontologies might prove insufficient for WSD in the SW, the following research question was formulated as:

Is it possible to improve WSD with the help of background knowledge incorporated in basic-level categories?

The answer to this question would give indication whether the background knowledge necessary for WSD contained in ontologies could be augmented by other sources. There are many types of expressions in natural language which are very hard to disambiguate, foremost among them metaphorical expressions. The background knowledge potentially available in conventional ontologies might prove insufficient for the purpose in these cases, and other sources for background knowledge might be needed. However, if metaphoric expressions could be structured in such a way that they were amenable to WSD with basic-level categories, they could yield useful background information for the purpose. The resulting research question would then be:

How to structure metaphoric expressions to make them amenable to WSD with basic-level categories?

We further need to find out whether it would be possible to use these cognitive theories in WSD and perhaps subsequently structurize them into a format that could be more suitable for conventional ontologies that the agents in the SW were better equipped to deal with.

3.2 Research Methods and Design

This chapter considers the research methods and experimental design of this thesis. The research methods used are considered first (3.2.1). Then the model of the application constructed is introduced in more detail (3.2.2). Finally, the experimental design and the datasets used in the experiments are considered and information about datasets used in the experiments is given (3.2.3).

3.2.1 Research methods

A constructive approach with iterative design is mainly used. Naturally, there are elements of other methodologies involved, including an explorative approach to find out more about the problem, which usually involves a fair amount of literary review and theory building, giving directions where to go next in the construction of the artifact.

The constructive research method aims (Järvinen and Järvinen, 2000) to build an artifact, which may be a new theory, model, algorithm or a framework. The resulting artifact should solve a domain problem and help in determining:

1) how the problem can be resolved
2) how the solution improves on possibly existing solutions.

According to Nunamaker et al. (1991) an artifact is based on a concept roughly comparable to a concept in a business idea or activity, whereas Järvinen and Järvinen (2000) see it as utilization of some resource be it technical, human or knowledge related, or of their combination. The latter does not exclude creation of new knowledge, which is why it is seen as an adequate definition of artifacts created in this thesis. The artifact(s) created here, moreover, combine constructs from across the disciplines of cognitive science and computational linguistics as explained in the introduction (cognitive, ontology relations).

Constructive research is probably the most common research method in computer science. Even though computer science is based on sound theoretical principles and theories, these are easiest to demonstrate through practical applications such as computer programs, languages, and other similar artifacts. This linkage between theory and practice must be explicitly taken into account in constructive research. To embark on constructive research it is often necessary to do some exploratory research prior to or during the constructive phrase precisely to create the linkage between the theory and practice. Part of this exploratory research can be explicitly undertaken, part of it may implicitly incorporated in the researcher's experience and background.

As most of the sciences one way or other become dependent on computer science, the constructive approach is bound to extend to other areas of science, including Computational Linguistics, the area of research of this thesis. Computational Linguistics, like computer science in general, also advances with the creation of new kind of software, algorithms, ontologies, corpora etc. For example, lack of suitable corpora has, on occasions, stifled advancement in the area. On other occasions, the availability and easy usage of some kind of ontology has concentrated the research around the best available ontology rather than around more suitable possible ontologies, resulting in using an inappropriate tool for a particular research purpose. These examples abound, and demonstrate that the advancement in Computational Linguistics, and by extension, in other fields reliant on computation, is often accompanied by a product or a construct of constructive research. This also, to a degree, justifies the use of constructive approach in this research.

The constructive approach itself can also be seen as a part of an iteration cycle between theory and practice: the constructive approach can create new knowledge, which, in turn can be used in theory creation, the theory created can be tested with a new artifact and so on. What is referred to by iterative design here concerns mainly the constructive phase: the artifact is developed to a certain point to test a theory, giving rise to a new idea how to test the theory better or combine with another theory or artifact to improve the artifact under construction.

It should be kept in mind that although we talk about 'design' here, it actually is a kind of decision framework where pre-planning is possible only to a certain degree: many of the decisions on how to continue can only be made after the results of the previous cycle are known. For example, in this research, the decision to extend the disambiguation methodology tested with SOM and a lexical ontology on homonyms was extended to metaphoric expressions only after the results on homonyms indicated that the use of basic-level concepts in disambiguation could be fruitful. Also the decision to extend the SOM based classification to whole WSD systems was taken only after it had proved useful in classification and labeling of lexical items. Later on we decided to combine selection of WSD system and WSD itself as a result of the previous experiments, but its implementation was left for future research.

Through the constructive approach, it is possible to check the assumptions and hypotheses motivating the research and to modify or extend them as necessary to create an application to experimentally demonstrate the viability of ontology based disambiguation. The end result is then incorporated into a WSD system selection application that will further augment the WSD application's disambiguation power.

3.2.2 Specific considerations for the research

Although we have many well designed recipes for research methods that can be applied in most of the scientific fields, one should never lose sight of the research-specific considerations. Here we need to use a multidisciplinary viewpoint as the research unites some of the prevalent theories in computational linguistics and cognitive science. The aim is to utilize the well-known theory of basic-level categories from cognitive science to improve the accuracy of current WSD methods. Secondly, as discussed in Article I, we stress the importance of combining different WSD methods and knowledge sources to optimize that accuracy, and do not solely rely on any particular method or knowledge source. We therefore advocate hybrid combination of WSD methods and knowledge sources to reach our aim.

3.2.2.1 Multidisciplinary viewpoint

Even though it is not necessary to model a word sense disambiguation system on a system based on human cognition, it would probably be unwise to ignore this aspect. This research, focusing on the use of background real-world knowledge in ontologies in WSD, was strongly motivated by the fact that long term memory and relational structures that cognition uses somewhat resemble the interplay between contextual knowledge and ontologies, although the underlying mechanisms most probably are different. The relations incorporated in current ontologies form a still very rudimentary set and need further enhancements and additions to make them more suitable for WSD. Some of the basic relations, such as incorporated in basic-level categories (Rosch et al., 1976) that have received very little attention in the WSD community have been at the centre of interest for cognitive scientists for many years now.

Le Ny (1995) presents an array of relations that, interestingly, can be shared both by machine and mental word representations (Table 1). Most of the basic properties (A) and derived properties (B) have already been used, in various degrees, in WSD (Ide and Veronis, 1998) and many of them are

TABLE 1: Properties common to natural and artificial semantic units (based on Le Ny (1995))

A. Basic Properties	Expression	Interpretation	Example
1.Denotation / categorization	D(L,a)	a is an instance of category L	Peter is a man.
2.Superordination / subordination	Z(L,S)	L is a subcategory of S	A mouse is an animal.
3. Attribution	A(L,p)	L has a feature of p	Apples are round.
	A(L,U,v)	L has the attribute U with the value v	Apples have a form which is round.
4.Filler in case roles / actantiation	C(L,G,H,I)	G= agent, H = patient, I = object	The boy gives the girl an apple.
5.Semantic / conceptual structures	Restaurant (food, tables, to eat, noisy….)	a script or a semantic unit of any type	Restaurant.
B. Derived properties			
6.Entailment	E(L,B) i.e., $\forall x$: L(x) → B(x)	L implies/entails b	If x is a sparrow, it is a bird.
7.Inheritance	I(L,S{U})	L inherits a set of attributes {U} from S	A Sparrow has all the features of a bird.
	I(L,S(t1,t2,t3…))	L inherits a list of features (t1, t2, t3…)	A Canadian apple is green (but has no other color value from apple).
	I(L,S(fi,f2,f3,…,with p1,p2,p3…))	L inherits a list of features (f1, f2, f3…) with respective probabilities (p1, p2, p3…)	A Canadian apple is usually grey, sometimes yellow, and seldom green.
C.Well-known psychological properties.			
8. Similarity	S(L,B)	L is similar to B	A cat is similar to lion.
9. Typicality	T(L,S)	L is typical of S.	Sparrows are typical of birds.
10. Basic-level categories	BL(L,S,I)	L is at the basic level in the hierarchy where S is the superordinate and I the subordinate level.	Dog is at the basic level between Animal (superordinate) and Irish Setter (subordinate).

standard relations in ontologies. Of the well-known psychological properties (C), similarity has been used extensively in WSD, less so typicality, and very little is known about the effect of basic-level categories to WSD.

The intention was to find out whether the use of novel relations incorporated in psychological properties as one of the factors of background information would improve WSD, and for this reason we selected basic-level categories for our system. The inadequate incorporation of these basic-level categories in knowledge sources, such as WordNet, was a further motivating factor for their selection.

3.2.2.2 Hybrid combination of WSD methods and knowledge sources

As noted, in Article I, a multiplicity of different grammars, ontologies, taxonomies and computational systems are currently being used in WSD. The systems are hard to compare, as the datasets used, tasks selected, and other factors vary greatly. None of the systems have been shown to be clearly superior to others, and the latest testbed trials in Senseval 2 and Senseval 3 (http://www.senseval.org/senseval3) have shown that the gains in accuracy are now quite marginal. To improve the accuracy of these disambiguation methods, it is highly desirable not only to look beyond the conventional computational linguistics methods, but also vary system components to see whether additions from the cognitive science or application of various ontologies are component dependent or whether their applicability might be more general.

This in mind, it was decided, from the beginning, to vary the systems components during the development. As there is a huge number of different WSD systems and there would be a myriad of various combinations available, it was decided not to do any exhaustive tests in this dimension, just introduce enough variability to ensure that the application of a basic-level theory from the cognitive science or the nature of ontologies used would not be component-dependent. The decision was also made, in the beginning, that the third author of Article I would pursue more closely the issue of the effect of various systems used on particular words for his PhD thesis.

The main knowledge sources used in the planning and processing stages were SUMO (Standard Upper Merged Ontology) (Niles and Pease, 2003) and WordNet (Miller et al., 1993), the grammars FDG (Functional Dependency Grammar) (Järvinen and Tapanainen, 1997) and HPSG (Head-Driven Phrase Structure Grammar) (Pollard and Sag, 1994). The processing part consisted of a modified SOM (Self-Organizing Maps) application together with InfoMap preprocessing system and our own processing application utilizing the theory of basic-level categories. The modified SOM application was later on used in combining classificatory systems to optimize their WSD capabilities and in view of future combination of the WSD system selection to WSD processing (Article VII).

The WSD system proposed here has two basic components that mirror, to a certain extent, how human cognition deals with natural language processing:

1) Knowledge source: an ontology of some sort that acts as a long-term memory for the WSD system incorporating real-world background knowledge in its concepts and relations.

2) WSD processor: a natural language processor that disambiguates a text using both the context and the ontology employed to help in the task.

The problem with the artificial knowledge sources is that they are still quite poor in types of relations between concepts, and therefore a pale reflection of human long-term memory. These associative relations are very hard to code in artificial knowledge sources. It is expected that the more of these relations can be employed the more useful the knowledge source will be for WSD. This in mind, it was decided to enhance the WSD system with basic-level relations. Naturally, we also needed a WSD processing system that can use these knowledge sources for the task intended.

3.3 Experimental design, datasets and validation

This section describes the design and the techniques used in the experiments of this dissertation. Of the articles published (Figure 2), Articles III and IV are concerned with the experiments proper related to this thesis. Article III concentrates on the use of a lexical ontology as a source for real-world knowledge for WSD and IV the use of basic-level relations for the same task. The experiments conducted in Article VI form a background for the experiments in III, and the experiments in Article VII point a way for further application and development of the concept developed in Article III. The datasets used in Article VI consisted of data generally used in benchmark testing, and one of these datasets was also used in Article III together with other data. Of the other publications, I takes a bird-eye view on the field, II proposes an approach, and V advances the method developed in IV.

3. 3.1 Experiment with the use of SOM and ontology in WSD

a) Prior experiment (Article VI)

This part considered how to identify ontology components using self-organizing maps (SOM). The application first produced a document space with individual vector spaces. Then SOM was constructed and trained with the document space and presented visually as a two-dimensional space. This resulted in the grouping of the concepts into fairly well defined domains.

Two related datasets were used: the Animals dataset (Ritter and Kohonen, 1989) which contains 16 animals with 13 boolean features that describe appearance and activities including the number of legs and ability to swim, and the Zoo dataset which is available from the UCL Machine Learning Repository (2006) and consists of 101 instances with 18 features. The numerical feature that identifies the animals was deleted. The modified Zoo dataset was used later on in the main experiment below.

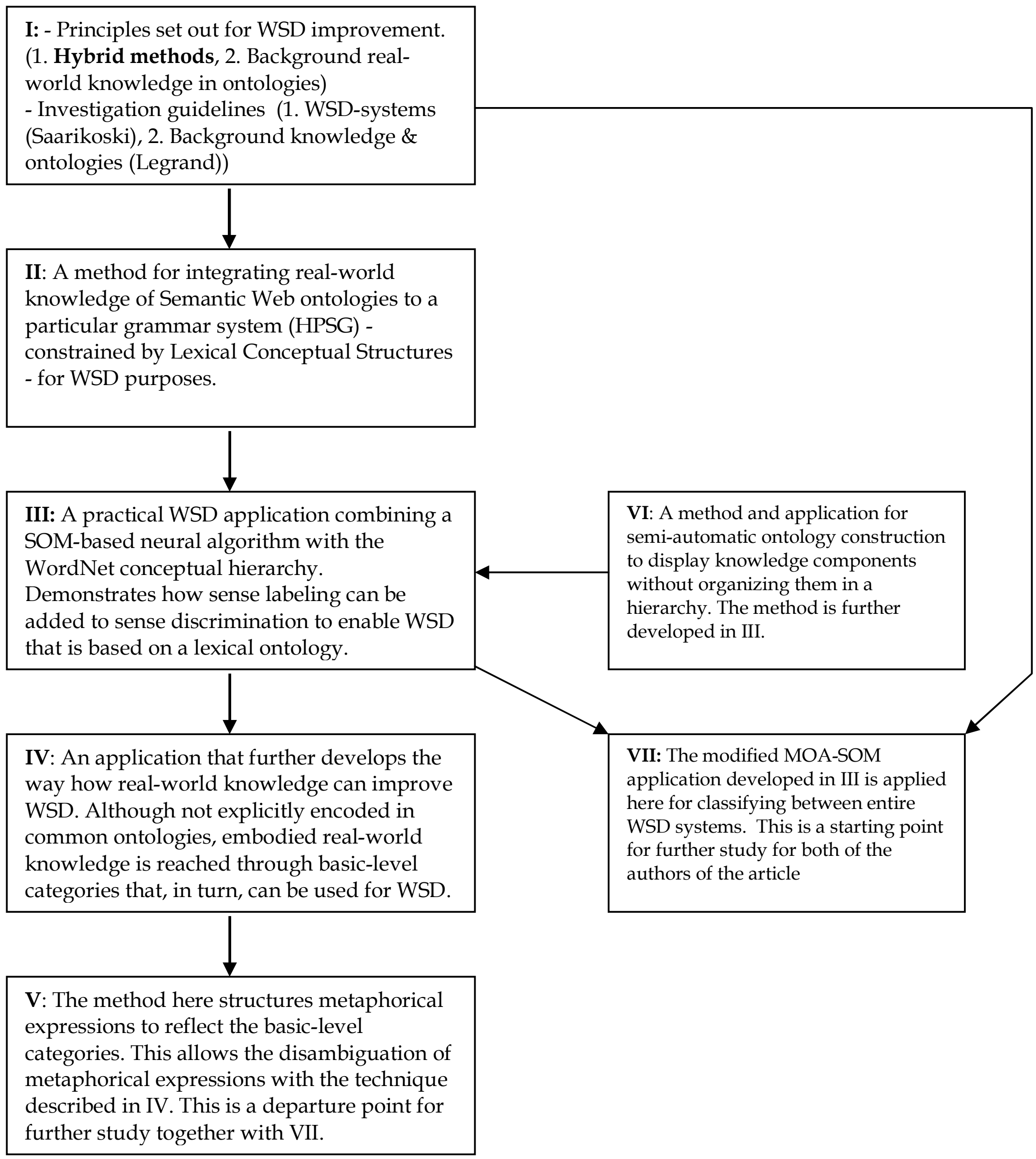


FIGURE 2: Publications roadmap.

b) Main experiment (Article III)

In this part, a feature based ontology component selection process (Article VI) was augmented to allow the labelling of these components automatically using an external ontology. Instead of a usual two-dimensional representation used in most SOM based concept maps, a hierarchical representation was used with the help of a modified SOM based DGSOT algorithm. The hierarchical representation could then be used to align it with an external ontology, WordNet, to get an access to real-world information contained in the ontology and in order to validate the hierarchy created and its correct labelling.

The Zoo dataset used in Article VI was modified for this experiment: the 1st attribute (name) and the 18th attribute (category) were not used for obvious reasons, as we wanted to find these automatically. Only 96 of the 101 animal instances were used: the 5 omitted could not be found in WordNet, and could not have been validated with it, either.

The results of the experiments were validated by comparing the classification obtained with SOM and labelled with a voting algorithm using the actual structure of the WordNet hierarchy. Apart from confirming the found category, its position in the created hierarchy could thus be verified.

c) Follow-up experiment (Article VII)

As the quest were for a hybrid system that could select the best WSD components for each specific word, apart from having access to ontology based real-world knowledge, it was decided to see whether the same components and methods used in Article VI and III would be useful in this task also, keeping in mind the future combination of these aspects in the application level. The modified SOM algorithm used in III was further modified (MOA-SOM) to allow its use here for clustering publicly available WSD system scores by calculating the amount of correlation between systems and words. MOA-SOM indicated the optimal classifier, feature and configuration for each target word. For predictors for the best system, we additionally employed the Weka toolkit (Witten and Frank, 2005). This experiment was planned as a stepping-stone for further research arising from the main subject of this thesis.

The WSD systems that were compared and assessed came from the participating systems from the Senseval 2 and Senseval 3 (http://www.senseval.org/) WSD competition, and, as systems, formed part of the dataset inputted to MOA-SOM. In the case of Senseval 2 systems, the predictors were trained with 39 words and all supervised systems were considered as candidates for best systems which were decided by the number of wins. We tested the model(s) on 19 words and three possible two-system combinations of the three top wordwinning systems as well as an ensemble of all three systems. In the case of Senseval 3 systems, the predictors were similarly trained with 39 words, 15 top systems were considered, and the three top wordwinners were selected for candidate base systems. We tested that model on 19 words and three two-system combinations of the three word-winning systems as well as an ensemble of all three systems.

3.3.2 Experiment with the use of basic-level categories in WSD

a) Prior testing and main experiment (Article IV)

The idea in this experiment was to see whether the accuracy of WSD could be improved with the help of basic-level categories. The reason for the use of basic-level categories in the disambiguation task in this experiment is that some of the background real-world knowledge cannot be readily encoded in ontologies. So-called embodied knowledge is encoded in basic-level categories, which is explained in more detail in Articles IV and V. Figure 3 depicts the data sources and application components used in the experiment.

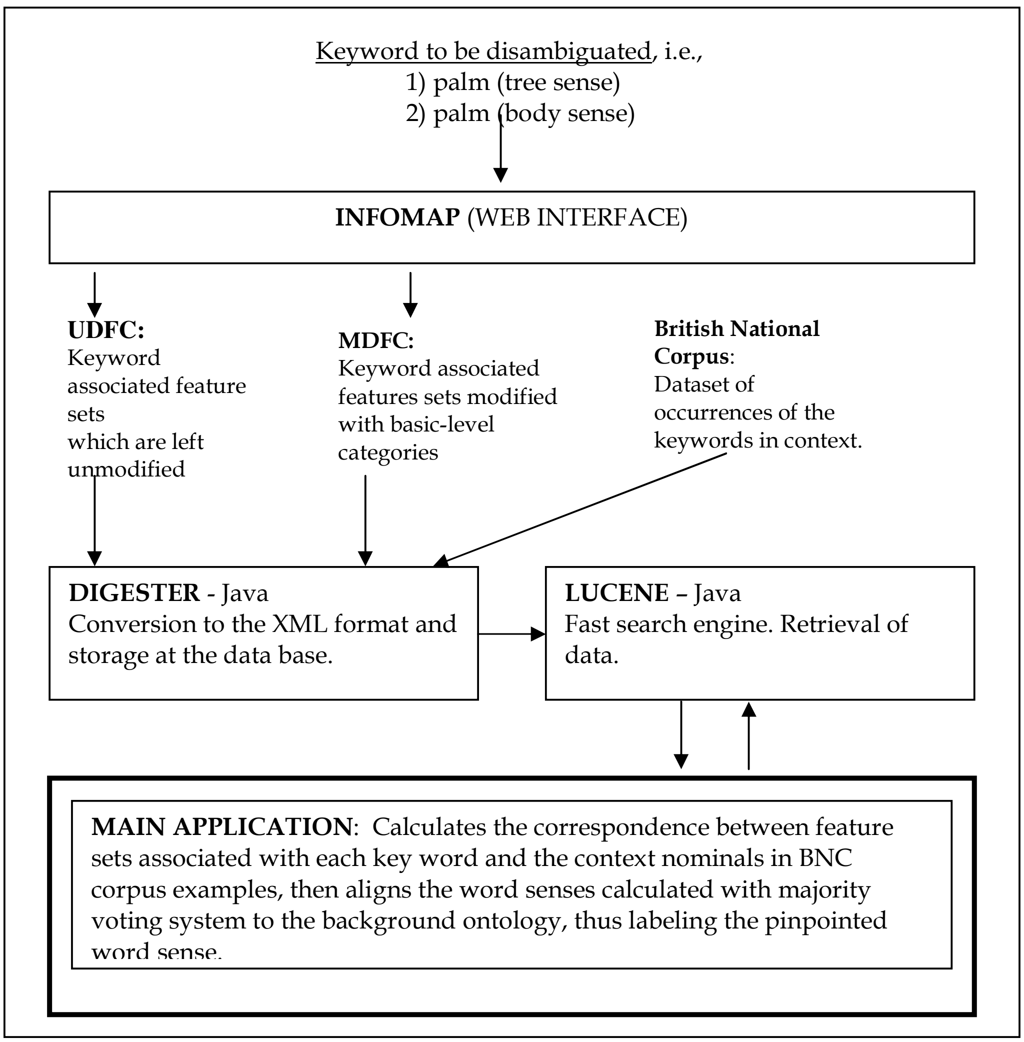


FIGURE 3: WSD application using basic-level categories in disambiguation.

Infomap (http://infomap.stanford.edu/) web interface was used to provide basic-level category words and other words that are related to the sense related keyword pairs such as *bat* and *animal* or *bat* and *sport*. These related words are retrieved from a so-called Wordspace (Schütze, 1997) created with the help of Latent Semantic Analysis applied on the British National Corpus. After converting the word clusters (feature sets) obtained by InfoMap to XML format using Java Digester, the application created for the purpose stored them in a fast Java Lucene search database.

We pre-tested our application with Mihalcea's (2003) TWA sense tagged data for six words with two-way ambiguities in order to select a suitable word sense pair for our experiment. The word "palm" was finally selected, because it had an adequate number of hand-tagged contexts (201) in the TWA dataset used by Mihalcea, and the disambiguation accuracy (75.1%) achieved with the raw Infomap feature set by us was judged sufficient, but not too high, for the purpose of the experiment. This feature set (UDFC) had all the original content from the InfoMap. The set was then modified, by pruning and by adding mainly basic-level information, to create a modified dataset (MDFC) for later use. Many of the words retrieved from InfoMap are basic-level items, but many are not, so pruning and modification is required to solidify the basic-level groups, i.e., parts, functions, and attributes. We matched the two feature sets against the stored BNC paragraphs to disambiguate the keywords, using a simple majority voting scheme as the decision factor. The idea was partially to see what the actual words that played role in disambiguation were and what was their number, in order to be able to roughly categorize the words participating in disambiguation for future modifications and generalizations.

For the main experiment we extracted 1000 paragraphs from the BNC containing the word *palm*. There are several other senses for the word apart from the two main senses, but these other senses were pruned from the dataset leaving 749 instances. This granularity was selected consciously as we were aware of the problems that too fine granularity in sense selection can cause to disambiguation (Mihalcea and Moldovan, 2001).

There were two parts to the main experiments. First the unmodified and the modified cluster sets obtained from the pre-testing experiments were compared for their disambiguation accuracy against the instances extracted from BNC. Then, based on the second set of experiments, an estimate was made about the contribution that the features sets (Parts, Functions) linked to basic-level information made towards the overall disambiguation accuracy.

The results were also validated with a 5-fold verification method, but the results of these were not yet available by the time of the publication. This was noted in the publication. More details about the experimental setting and results can be found in the corresponding sections of the related publication. The structure of some of the dataset is also described there.

b) Follow-up (V)

Article V basically builds on the experiment in Article IV and prepares a stage for an experiment in which basic-level concepts obtained from metaphoric expressions are to be used for WSD in the manner described in the main experiment above (a). For this purpose a theoretical ontological framework is constructed in which basic-level expressions picked from metaphorical contexts form the building blocks.

4 SUMMARY OF THE ORIGINAL ARTICLES

4.1 Article I: "Bridging the Word Disambiguation Gap with the help of OWL and Semantic Web Ontologies"

Reference: Legrand, S., Tyrväinen P., Saarikoski, H. Bridging the Word Disambiguation Gap with the help of OWL and Semantic Web Ontologies. In C. V. Pallotta, A. Todirascu (Eds.): *Ontologies and Information Extraction International Workshop held as part of the EUROLAN 2003,* Bucarest 28th July – 8th August 2003, pp. 29-35

This paper basically outlines the task ahead: how to improve the current word sense disambiguation accuracy, and what do we need to know to be able to accomplish that?

To start our investigation we define the so-called disambiguation gap which is the gap in WSD accuracy between human beings and artificial systems in WSD. The main factors (application domain, information sources, and knowledge types) influencing that gap and how to deal with it are considered. Among the sources and types currently under-utilized in WSD due to their poor availability, costly acquisition or insufficient appreciation, we identify ontologies.

Ontologies may contain a lot of background real-world knowledge that is useful in WSD. The use of ontologies is topical also because they form the core of the Semantic Web now under construction. The role of NLP applications in the Semantic Web, and the inference capabilities that the now de-facto web ontology language OWL brings in to WSD, are therefore explained.

At the moment, there is a multiplicity of different grammar systems, ontologies, taxonomies and computational applications that are being used in word sense disambiguation (WSD). None of these have been shown to be clearly superior to others, and the results are hard to compare due to variety of methods, datasets, and application domains involved. For this reason, we propose the use of hybrid multilevel disambiguation, which would include a mix of knowledge types and information sources including ontologies with their inference capabilities. In our example, we use the combination of Functional Dependency Grammar (FDG), which provides a mix of knowledge types (morphology etc), Suggested Upper Merged Ontology (SUMO) adding background real-world knowledge, and OWL inferencing language.

Finally, a construction of a WSD Knowledge base is proposed: it would be related to various domain ontologies through a set of clearly defined terms. Knowledge acquisition bottleneck is spotted as one of the most important problems to overcome in the task, and various possibilities to test the knowledge base are entertained. At this stage the idea of the WSD Knowledge base is still quite sketchy. The subsequent articles will contribute towards this idea, although WSD Knowledge base is not explicitly mentioned in them.

On the whole, the article clarifies the research area, identifying the so-called disambiguation gap as one of its most central issues to be addressed and hybrid multilevel disambiguation with real-world ontologies as a probable solution to it. Behind the scenes it also sets up the author's responsibilities in the research area (the results to be combined later in Article VII): while the first author (Legrand) embarks on the research of the role of background real-world knowledge in WSD tasks with different grammar systems and ontologies and the coding of that knowledge for the use of WSD, the third author (Saarikoski) starts investigating different disambiguation systems and their contributions to WSD on a word-by-word basis applying his methods on Senseval data. The second author (Tyrväinen) is responsible for guiding that process. The article was written mainly by the first author.

4.2 Article II: "Connecting Lexical Knowledge to Distributed Real-World Knowledge"

Reference: Legrand, S., Tyrväinen, P. Connecting Lexical Knowledge to Distributed Real-World Knowledge, In T. Cameron, C. Shanks, K. Holley (Eds.): *High Desert Linguistic Society V Conference,* 1-2 November 2002, Albuquerque, New Mexico, pp. 109-120.

In this article we investigate, with the help of HPSG (Head-driven Phrase Structure Grammar) how to incorporate real-world knowledge in a grammar for linguistic processing. Using HPSG as an example grammar fits well with our overall strategy of hybrid knowledge sources. HPSG uses a kind of a hybrid representation of language. It is one of the unifying grammars, with a multidimensional representation of language, which usually include information about orthography, morphology, syntax and semantics. Nevertheless, no detailed mechanism had been proposed as regards how to connect this grammar to real-world knowledge sources. Our research was embarked upon to provide a remedy for this. The article further illustrates that adding background real-world knowledge to lexical entries is feasible in more than one type of grammar (another grammar, FDG, was dealt with in Article I).

On the other hand, we also investigate how to connect distributed real world ontologies in the Semantic Web to linguistic knowledge. How to utilize the growing repository of ontological real world knowledge in the Semantic Web for the purpose of WSD through a parsable grammar (HPSG) is thus the other main topic of the article.

It is often hard to separate semantic and real-world knowledge entirely, for example, entailments when mapped from semantics to syntax can be viewed as world knowledge being implications about certain classes of actions in the world. Partially for this reason the article considers first how semantic knowledge coded in Jackedoff's Lexical Conceptual Structures (LCSs) can be integrated to HPSG using its semantic sorts. These conceptual structures, although not related to basic-level categories (the topic in Articles IV and V) are relevant to this thesis in the sense (as shown in Articles IV and V) that where the real-world knowledge coded in conventional ontological structures may prove inadequate for WSD, other more natural structures, from a human viewpoint, containing semantic and real-world knowledge could be considered.

The avenues available for integrating real-world knowledge to HSPG are then briefly considered, and the avenue looking most promising in the light of the Semantic Web development is then chosen, i.e., "ontological enrichment". Semantic sorts in HPSG, specifically the CONTEXT's BACKGROUND attribute, are proposed to be used to create a top ontology that is then aligned directly with the appropriate Semantic Web distributed ontology. This can also be done indirectly using SUMO top ontology to coordinate the alignment. With the development of agent technology (this aspect is further developed in Article VI), the communication and coordination between the grammar and the ontologies in the Semantic Web is not seen as a problem.

The article set out to find a mechanism to integrate real-world knowledge in Semantic Web ontologies to the structure of HPSG, a grammar that can be parsed and be used in WSD. A feasible way was found and demonstrates (together with Article I), on the general level, that connecting lexical knowledge in a grammar to various types of ontologies can be achieved with suitable modifications to the grammar in question. The article was written and the method presented conceived by the first author; the second author acted as the adviser, contributed to its contents, and checked the article.

4.3 Article III: "A Hybrid Approach to Word Sense Disambiguation: Neural Clustering with Class Labeling"

Reference: Legrand, S., Pulido J.R.G. A Hybrid Approach to Word Sense Disambiguation: Neural Clustering with Class Labeling. In P. Buitelaar, J. Franke, M. Grobelnik, G. Paaß, V. Svatek (Eds.): *Knowledge Discovery and Ontologies (KDO-2004) workshop in15th European Conference on Machine Learning (ECML)* Pisa, Italy, September 24, 2004, pp. 127-132.

We wanted to conduct a practical experiment in order to show that word sense disambiguation can benefit from real-world knowledge available in ontologies. The article shows how by combining a neural algorithm based on SOM (Self-Organizing Maps) with the WordNet lexical database it is possible to label groups of items clustered in a multi-branched hierarchy, paving way for the use of neural algorithms together with ontological knowledge in word sense disambiguation tasks.

The problem with SOM when using it in WSD tasks has been in its representation of concepts: it discriminates the senses into groups, but it does not label the groups or make the relations within and between each group explicit. We were able to create a hierarchical division and to label, in an unsupervised manner, the discriminated groups with the help of a WordNet conceptual hierarchy - a necessary step in order to use the combination for word sense disambiguation.

Another important component in the application is the DGSOT algorithm, based on SOM and related to the SOM application in Article VI. DGSOT was connected to WordNet and modified to allow labeling, in addition to discrimination. The application thus created was later on used in Article VII for discrimination and labeling of entire WSD systems. Once the modified DGSOT has grouped the items on the basis of selected features (these features may be, for example, related to co-occurring words in a text), we can automatically label them using the WordNet's hypernym-hyponym conceptual hierarchy. The data set used (zoo) was the same as in Article VI.

The algorithm we used for this is simple but powerful. First it finds the most likely common ancestry, according to WordNet, for the items in a group, and then seeks for the location of the item in the WordNet hierarchy branching from the calculated common ancestry. For the calculation of the common ancestry we use a simple majority voting scheme: first we count the total number of times each ancestry node occurs in all of the group item's ancestries, then we compare the accumulated total points of ancestries in the group in question and select the ancestry with the maximum points as the winner for that group.

On the whole, this article shows that it is possible to have a system, which combines real-world/background knowledge in taxonomies/ontologies such as Wordnet with a neural algorithm and can be used for word sense disambiguation. In particular, we effectively demonstrate that neural processing of datasets to categories does not need to stop at sense discrimination, but that word sense disambiguation through class labeling can be attempted.

The first author contributed the bulk of this article, which includes planning, building the application, devising the algorithm and conducting the experiment. The second author, provided the dataset and advised about the peculiarities and benefits in its use. He also provided the use of his own application to confirm the partial results before labeling was attempted.

4.4 Article IV: "Word Sense Disambiguation with Basic-Level Categories"

Reference: Legrand, S. Word Sense Disambiguation with Basic-Level Categories. In *Advances in Natural Language Processing*. A. Gelbukh (Ed.): Research in Computing Science. Vol.18, IPN, Mexico, 2006, pp. 71-82.

The aim of this article is to find out whether so-called basic-level categories could help in word sense disambiguation by supplying a part of the real-world knowledge that is not normally coded in conventional ontologies' relations of various kinds. This kind of real-world knowledge consists of concepts such as color, size, appearance etc., which are experienced as a Gestalt phenomenon. These types of concepts are not normally associated with conventional ontologies, which usually rely on subsumption and meronymic relations, mainly because the Gestalt phenomenon based on basic-level categories is not easy to fit in. One kind of real-world knowledge that is known as embodied knowledge, and often available in basic-level relations, is especially hard in this respect. Nevertheless, it plays a fundamental part in human language use, and therefore has great potential for WSD.

The article first considers the theoretical basis of basic-level categories from the viewpoint of its suitability to linguistic tasks. Lakoff's challenge to the classical view of categorization, in the form of idealized cognitive models (ICMs) is the key in selecting the basic-level categories as a knowledge source for the experiments in this article. Lakoff advocates the view that linguistic categories have the same character as other conceptual categories, i.e., they show prototype effects and can be demonstrated to have basic-level categories. Both he and Rosch regard the basic-level categories as a surface phenomena related to cognition, and expect there to be other interesting structures and processes below the surface. This particular aspect is further developed in Article V, which deals with structuring of metaphorical expressions. Lakoff's research indicates thus that by using the basic-level categories in the linguistic domain, we can discover novel categorical structures, not available in conventional ontologies. As stated, this may be one of the reasons why these conventional ontologies may prove inadequate for linguistics tasks such as word sense disambiguation, and this is the motivating factor for this article.

The WordNet ontology is used here in a role similar to that in Article III, i.e., to confirm the correct word sense selection. As an ontology, WordNet is organized in the manner of a conventional ontology, although it also contains many other relations that are useful for linguistic tasks. To aid in finding out the correct sense we use InfoMap web-based software from the Stanford University. In InfoMap a profile of the words usage in a context is created out of the distribution of word co-occurrences between a word and sets of content-bearing words. By comparing the profiles of the words in question the similarity between words can be calculated. In effect, a list of word groups, which are related to the keywords entered and which are arranged in the order of affinity of meaning, is returned by the browser. These keywords, in our case, consisted of the word "palm" and "hand" on one domain, and "palm" and "tree" on the other domain. The idea was to disambiguate between the two senses.

After converting the clusters obtained by InfoMap to XML format using Java Digester, the application created for the purpose stored them in a fast Java Lucene search database. First we pre-tested our application with Mihalcea's TWA sense tagged data for six words with two-way ambiguities. Our own experimental data sets were also extracted from BNC, which could provide an adequate number of instances. We used two sets of clusters for comparison. One of them was moderately modified, by pruning and additions, to contain mainly basic-level information (parts, functions, attributes), while the other set kept all the original content from the InfoMap. The idea was partially to see what were the actual words that played role in disambiguation and what was their number, in order to be able to roughly categorize the words participating in disambiguation for future modifications and generalizations.

The experiment had two parts. In the first part the unmodified and the modified cluster sets were compared for their disambiguation accuracy. In the second part, an estimate was made, on the basis of a second set of experiments, about the contribution that the feature sets (Parts, Functions) linked to basic-level information made towards the overall disambiguation accuracy.

The results of the experiments are encouraging. When the two major senses (part-of-hand, tree) of the word palm in 749 contexts were disambiguated with the unmodified feature cluster set, the correct choices formed 79.6% in a paragraph-wide context (wide context), 73.4% in a sentence-wide context (narrow context). For the feature cluster set modified by basic-level concepts the correct choices formed 93.7% and 92.4%, respectively, of the same sets - a clearly better result. The positive effect of basic-level concepts to WSD was confirmed with some further experiments, where the contribution towards the overall disambiguation accuracy of feature sets (Parts, Functions) linked to the basic-level information was measured. When the parts and functions were excluded from the modified feature cluster, the correct choices formed only 62.7% and 52.8% of the wide and narrow context, respectively, whereas, when only parts and functions were used in disambiguation the corresponding figures were 77.2% and 74.6%.

We expect that by augmenting the relations of WordNet with basic-level relations and idealized cognitive models WordNet could be made more suitable for disambiguation purposes. The results also provide motivation for more fine-grained study of metaphorical expressions with the help of basic-level relations (Article V).

4.5 Article V: "Structuring metaphors with basic-level concepts for word sense disambiguation"

Reference: Legrand, S., Structuring metaphors with basic-level concepts for word sense disambiguation. In J. Šķilters (Ed.): *Baltic International Yearbook of Cognition, Logic, and Communication.* Vol.2: Complex Cognition and Qualitative Science, 2007, pp. 171-188.

While from the point of view of real-world knowledge, the hard thing was how to capture, for WSD purposes (Article IV), embodied real-world knowledge that may not be encoded in the ontology used, from the viewpoint of language itself one of the hardest thing is to disambiguate metaphorical expressions. The article aims to show how this could be done by structuring metaphorical expressions to reflect the basic-level categories in source expressions for the metaphor. These basic-level expressions could then be used in WSD as described in Article IV. It turns out that the results in Article IV may support the task we set out here.

The model we present in the paper is based on basic-level categories by Rosch (1988), Contemporary Theory of Metaphor by Lakoff and Johnson (1980), and Conceptual Mapping Model by Ahrens et al. (2004). These concepts are explained in detail in the article. Also, there is a section on current research where we draw a distinction between research that attempts to discover metaphorical knowledge from ontologies without modifying them and research which aims to modify ontologies with metaphorical knowledge. The model proposes a way that may mimic a human cognitive system in building abstract concepts with the help of metaphors and basic-level schema. It does not pretend to be a faithful model of human abstraction mechanism, although it uses some widely respected results from the research field in its construction.

The idea behind is related to the results in Article IV in which we used basic-level concepts in WSD. If we can use basic-level categories to help in WSD, then, if metaphors can be reduced to basic-level concepts by restructuring them, that would help in their automatic disambiguation. In the article, the structural transformation of metaphors to their basic-level primitives faithfully mirrors the transformation of more concrete concepts to these primitives usually referred to in the literature. What is striking is that one can find almost exactly the same categorical structure which was found in the earlier research for more concrete concepts, once the metaphor is reduced to its basic-level primitives.

The example used in the article is the abstract concept, "*argument*", which can be expressed through many different metaphors. The Argument concept itself is deemed to belong to the basic-level, sandwiched between the superordinate level concept, "*conflict*", on the one hand, and numerous subordinate concepts, such as "*sparring*", "*fight*", "*polemic*" etc., on the other. As in the case of more concrete subjects when reduced to their basic-level primitives, here too we can distinguish the all-important triad: functions-parts-attributes. For example, if the metaphor used is ARGUMENT IS STRUCTURE (building) the parts would include "*basis*", and "*framework*", the functions would include "*construct*" and "*demolish*", and the attributes would include "*strong*" and "*weak*", among others. Many other parts, functions, and attributes could be found. Thus a paragraph such as

> "*Your argument is going to fall down. Its framework is weakly constructed, and it can be easily demolished*."

could be restructured to its basic-level primitives using the concepts occurring in the paragraph context. Other metaphors could also be transformed in a similar way. In the case of argument, ARGUMENT IS CONTAINER or ARGUMENT IS NAVIGATION metaphors would be amenable to a similar treatment. As each metaphor seems to be associated with a certain domain (building , navigation) it is conceivable that the background real-world knowledge in specific domain ontologies when connected to basic-level context words in metaphors could then help in WSD. These ontologies could come from the Semantic Web as discussed in Articles I and II, or the specific domains could form part of more general ontologies.

The theories on which the research here is based has its critics and their objections to these theories are exposed and discussed in the article. The discussion part points out that the intention is not to prove or justify any of these theories but to show some practical use they can be put into. Some weak points in these objections are pointed out. Also, various qualifications to some of these theories by their proponents are brought up to bring the theories to a more reasonable light. Due to some of these objections the work in this article is also examined more critically in the discussion part.

The article set out to create a novel method to structure metaphors with the help of basic-level categories for WSD purposes. The resulting model will help scientists dealing with natural language disambiguation, whenever metaphors are used to construct abstract concepts. The work here also forms a base for connecting metaphor domains to domain ontologies in order to utilize real-world knowledge contained in them for WSD. This is an area of further study and will be combined with Article VII and other results obtained later to allow the proposed comprehensive System selection/WSD application to also cater for harder cases such as metaphorical expressions by including embodied real-world knowledge in addition to real-world knowledge in conventional ontologies. Simply put, the planned application, when selecting a domain ontology to connect to, would subject the rejected domains to a second level of scrutiny, in which possible links between metaphoric context words and other domains for the keyword would be mapped prior to any new attempt to connect to a potentially relevant ontology.

4.6 Article VI: "Identifying Ontology Components from Digital Archives for the Semantic Web"

Reference: Pulido, J.R.G, Herrera, R, Aréchiga, M., Block, A, Acosta, R, Legrand, S. Identifying Ontology Components from Digital Archives for the Semantic Web. In S. Sahni (Ed.): *The IASTED Conference on Advances in Computer Science and Technology.* Puerto Vallarta, Mexico, January 23-25, 2006, pp. 7-12

The article describes an approach that contributes towards semi-automatic construction of ontologies for web sites. The idea of combining ontologies and semantic maps motivated the work. One of the aims here is to provide solution to the problem of semi-automatic ontology construction. Although not explicitly stated, this is related to the present thesis in that these ontologies would provide useful real-world knowledge for linguistic applications in the Semantic Web. Also, browsable representation of ontology components, another of the aims, is relevant to the present thesis. Although published later, this work went on before and simultaneously with Article III.

The system presented here consists of two applications: Spade and Grubber. Spade pre-processes HTML pages and creates a document space, after which Grubber uses the document space to produce knowledge maps with the help of a modified SOM algorithm. These maps then allow visualization of ontology components on a two-dimensional plane. The components can be organized more formally into Entities, Relations and Functions. New knowledge can be inferred from these by problem solvers and other inference mechanisms. In the experiment, two datasets (animals, zoo) from the University of Central London repository were employed due to their easy availability and comparability of the results of other experiments. The zoo dataset was also used in Article III.

The experiments conducted in this article show that Self-Organizing Maps are an efficient software tool to analyze domains. At this stage, the application discriminated between concepts, grouping them in such a way that a human domain expert can label the groups with a suitable label from a concept ontology. This is still inadequate for unsupervised word sense disambiguation. The idea is further developed in Article III where the labeling is done, in an unsupervised manner, with the help of an existing concept hierarchy, WordNet, allowing word sense disambiguation. The work in this article, also helped in selecting the data set used in the experiment in Article III, and similarly, in the development of the application used in Article III. In addition the work accomplished here offers insight to the employment of software agents in the Semantic Web for linguistics applications.

The contribution of the author of this thesis, in this article, consists of the preparation of the data set and the experiment, participation in the experiments and the analysis of the results. The author also participated in the preparation of the final text.

4.7 Article VII: "Building an Optimal WSD Ensemble Using Per-Word Selection of Best System"

Reference: Saarikoski, H., Legrand, S., Building an Optimal WSD Ensemble Using Per-Word Selection of Best System. In J. F. Martínez-Trinidad, J. A. C. Ochoa, J. Kittler (Eds.): *Progress in Pattern Recognition, Image Analysis and Applications*. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2006, pp. 864-872.

The results of the evaluations of Senseval 2 and Senseval 3 - de-facto evaluation testbeds for WSD systems - indicate that progress in disambiguation methods has become very slow indeed. This seems to be due, firstly, to that different disambiguation methods result in different performance bias, and, secondly, that each word poses a different set of learning problems. This article presents a method to resolve these biases.

To do that we first needed to find a definition of a system that is best equipped to handle a particular target word. Literature shows that there are three important factors that most influence WSD system performance: word grain (number of senses distinguished), amount of training, and most frequent (dominant) sense bias in training data. On the basis of these three word features we then set out to develop a method that could predict the strength of each system when dealing with a particular word. The longer-term intention is to develop a hybrid system that could disambiguate text using the strongest component for a particular word to handle it.

Based on the application in Article III we developed a meta-classifier MOA-SOM for the task. The tool clusters WSD system scores stored in database based on correlation between features defining the systems (e.g. classifier algorithm, feature sets) and target words (e.g. PoS, training, word grain). As an output, optimal classifier, feature and configuration for that target word are obtained. The feature matrix can be input to SOM using either system names as labels and words as data points or vice versa.

On the basis of the results obtained from MOA-SOM, we trained the most promising predictors using both manual rules and machine-learning algorithms implemented in the Weka toolkit. Then we tested selected base systems and the ensemble (according to the best predictor for that ensemble) on test words. Next the ensemble was evaluated by comparing it to the better of the base systems. Also the predictor was evaluated using a gain measure we created for the purpose: ((PredictionAccuracy - (1.0 / NumberOfSystems))*2)* GrossGain. The method was finally applied to the two Senseval evaluations.

Best predictors varied according to the base system pair, both in terms of learning algorithm and input features. Machine-learning models seemed to work better than manual rules. A combination of factors tended to work better than individual factors. This indicated, to us, that system prediction task , like the WSD task itself, is affected by the details and the difficulty of the task, and that a customized predictor may need to be developed for a given system pair. If two such 'opposite' systems that together optimally cover the word space are combined that might give a rise to a more general hybrid system.

This is a departing point for future study for the authors of the article. The idea is, now that it has been shown, in Article III, that the neural algorithm can be used - together with a suitable ontology - for word sense disambiguation, to modify the MOA-SOM application to make it: 1. Select the best WSD system for the words in question, and 2. Combine the system with a background ontology, as in Article III, to maximize the WSD accuracy in the subsequent disambiguation. Combining system selection with ontology selection and a subsequent WSD is a topic to be pursued in future studies.

The first author is responsible for the design of the research in general, and, in particular, for the design and conduct of the WEKA experiments. The second author contributed with the MOA-SOM application and its modification for the purpose. The article was written together in close consultation.

5 CONCLUSIONS

This chapter summarizes the main contributions of this thesis, their relevance to the study, and discusses its limitations. It also considers possible future directions of the study, for which the thesis itself has already contributed in preparing the stage.

5.1 Contributions of the thesis

The main contribution of this thesis is to improve the present state-of-art in unsupervised WSD accuracy by the use of background real-world information contained in ontologies and basic-level categories. This is beneficial in environments such as the Semantic Web, which relies on ontologies in its functioning and where, therefore, the use of unsupervised WSD in agent-based AI applications is needed. This research is based on theoretical work in computational linguistics and cognitive science both and combines them in a novel way. The main research question

RQ1: *How can we improve unsupervised WSD with the use of background knowledge incorporated in ontology relations?*

subsumes the other research questions, and should be kept in mind when finding answers to them. As noted in Article VI and Article VII, WSD has reached a standstill and might benefit of background knowledge in ontological relations. We set out to find out, by the experiment in Article III, how to do that.

Prior to finding an answer to the main research question, we however needed to find out – keeping in mind that our research was motivated by possible applications in the Semantic Web – whether it would be possible to collect the ontological information automatically from an environment resembling the environment in SW:

RQ2: *How could one organize information that is scattered and unorganized, grouping and labeling it so that it could be matched against a suitable ontology for WSD or other purposes?*

We needed to find an unsupervised method that would be robust enough for the environment and also tested and tried in linguistic applications and ended up with SOM (see Kaski, 1997), which we used to group feature based information into relevant sets (Article III). These features were obtained from an artificial dataset, but can also be obtained from any text with suitable co-occurrence technologies. After classification into sets, our modification to the SOM-based DGSOT algorithm together with our own majority voting type algorithm allowed us to find out the correct labels for each word when connected to WordNet with a Java based interface. Thus we ended up with an application prototype that could be used, with some refinements, in an environment such as SW, to organize scattered and unorganized information, grouping and labeling it to allow it to be matched against a suitable ontology for WSD purposes. This finding also satisfactorily answers RQ2.

Another result from this particular research was the discovery that the same modification to the SOM based DGSOT algorithm could also be used in finding out the best WSD system for a particular word to be disambiguated, opening up the possibility of combining a WSD system finder with the WSD itself. The research is now continuing in this area.

The main research question, RQ1, does not specify where the ontological background knowledge comes from, though. Background ontological knowledge can be coded, inherently, in other types of systems and relations that do not resemble conventional ontologies. We decided to test whether our assumption might hold for relations like these also. The idea to use so-called Basic-level relations for our second experiment came directly from our knowledge of the Miller et al's (1993a) initial plans to include the basic-level relations into WordNet which was never, however, implemented. There was a strong suspicion that basic-level categories are crucial for language understanding. The third research question was formulated to bring some light to this aspect:

RQ3: *Is it possible to improve WSD with the help of background knowledge incorporated in basic-level categories?*

To answer this question we had to find a good source for basic-level information. Still, keeping in mind our SW oriented approach, that source should be available through a programmable interface that could be automated.

We found a partial answer to that, InfoMapping software. We manually modified the result set and compared it with the unmodified result set on disambiguating word senses from the paragraphs extracted from the BNC.

The results of the experiment in Article IV clearly show that the use of basic-level relations, and therefore the background knowledge inherently encoded in it, does improve WSD. For the original feature set extracted with InfoMapping the disambiguation accuracy for a wide context was 79.6% and in a narrow context 73.4 %, whereas for the modified feature set consisting mainly of basic-level information the corresponding percentages were 93.7% and 92.4% respectively. Excluding parts and functions from the modified feature set gave the accuracy of 62.7% for a wide context and 52.8% for a narrow context, whereas using only parts and functions of the modified feature set improved the corresponding figures to 77.2% and 74.6%, respectively.

Another result from this experiment is that it confirms the suspicions of the makers of WordNet that basic-level relations are, indeed, important constituents of a lexical database - moreover, if used for WSD, perhaps a crucial one. The implication is that conventional ontologies could be improved in respect to WSD uses by augmenting them with basic-level relations.

However, even supposing that we augmented conventional ontologies with basic-level and other important relations currently missing, not all linguistic constructs yield, at least directly, to these ontological manipulations that may form part of WSD. One such, and very important, linguistic concept, is metaphor.

Metaphors are notoriously difficult to disambiguate due to their somewhat unpredictable appearance in a text, and their gradual conventionalization. Also, new metaphors are constantly being created. During Experiment 2, we noticed that quite a few of the errors in disambiguation were due to metaphorical expressions (these expressions were subsequently pruned away from the BNC set), and this gave the idea of structuring them in such a way that they could be disambiguated easier. This question to research became:

RQ4: *How to structure metaphoric expressions to make them amenable to WSD with basic-level categories?*

This question is answered with a theoretical construct that arises from the experience gained from the two experiments above. It is based on theories and work of Rosch (1988), Lakoff and Johnson (1980), and Ahrens et al (2004).

The idea here was to reduce metaphoric expressions to their constituent parts which are related to the basic-level category structure used in WSD in Article IV. For this an ontological structure accommodating the constituents of a metaphorical expression and relating them to basic-level expressions was constructed.

The result forms a promising base for connecting metaphor domains to domain ontologies in order to utilize real-world knowledge contained in them for WSD. It can be combined, in theory, with the earlier proposed comprehensive WSD system finder/disambiguator to also cater for harder cases, among them metaphorical expressions. The resulting application would include embodied real-world knowledge in addition to real-world knowledge in conventional ontologies.

5.2 Limitations of the thesis

What we have shown in this thesis is how to benefit, with the help of ontologies, from background real-world knowledge for WSD in environments such as SW. We have also shown that it is possible to restructure cognitive constructs into a form of conventional ontologies in order to enable them to be used for similar effect. However, there are two main things that must be kept in mind when evaluating the overall research goal and the results related to it.

First, although ontologies seem to be the best structured source for background real-world knowledge, this type of knowledge is hard to structure and exist in many other forms where it is not directly available for unsupervised machine WSD. Most kind of pragmatic knowledge such as situational knowledge belongs to that category. A lot of work is still to be done in this area.

Second, apart from metaphors dealt with in here, there are many other special cases in the linguistic phenomena that do not readily yield to conventional or novel ontological structuring. Therefore, other types of adaptations or modifications might be needed to use them in WSD. These are the main reservations related to the present study, and limit its findings to the use of ontologies as a source of background real-world knowledge for WSD, excluding most of the special cases of linguistic phenomena.

Apart from the limited focus, there are other more specific limitations within that focus. In the first experiment (see Articles VI and III), in order to make the results easier to compare with the results of other research, the main datasets used were artificial. The main dataset consisted of nouns, excluding other syntactic categories such as verbs and adjectives, and it also originated from a single domain. It seems very unlikely that the 100 percent matching obtained in the test with nouns could be maintained with the other categories or domains; it is well known that the categories of verbs and adjectives in WordNet are much shallower and harder to use for labelling, and some domains in the noun category are not as well structured as others. Therefore, the algorithm used needs to be applied on larger and more heterogeneous datasets and modified if necessary. Further experiments are required in this area. The related article contains a qualifying note to this effect.

In the second experiment (see Article IV) the context data came from the British National Corpus (http://www.natcorp.ox.ac.uk/) and was thus more natural. It would have been better, though, if the pre-testing data had come from a source other than Brown Corpus (http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/corpora/list/private/brown/brown.html), which is a subset of the BNC, but as it was the only annotated part of it available to us, we used it. Also the results of the 5-fold verification we conducted confirming the results of the experiment were not yet available at the time of the publication of the article. This was due to the fact that our research for Articles VI and III took place in two distant countries and there were some problems with coordination. As in the case of Experiment 1, this experiment should be extended to syntactic categories and domains other than nouns.

The theoretical construct in Article V, the structuring of metaphorical expressions with the help of basic-level categories, should be extended to include as many metaphorical structures as feasible and then apply it on WSD with experimental data. This is going to be a huge effort, however, and was defined out of scope of this thesis.

5.3 Future Work

At this point it is useful to be reminded of the fact that language understanding by machines in the realm of artificial intelligence since 1950s and computational linguistics more recently has been a very elusive goal. New approaches are needed to reach that goal, especially now that the emergence of the Semantic Web is round the corner. The main point of this thesis was to contribute towards this goal by combining research in the fields of cognition and computational linguistics.

As discussed in the previous section, the area of linguistics research is vast, and it is necessary to apply the techniques used in this thesis for other syntactic categories, other types of ontologies and for special cases, apart from metaphors, in linguistics to enable a machine to understand natural language. Other sources for background real-world knowledge need to be investigated as well, including pragmatic knowledge. This could be called the general future goal.

More specific goals would include the application of different inference mechanisms with the help of existing and future SW standards, formats and languages on domain and general ontologies used in SW in order to help in unsupervised natural language disambiguation and understanding. The author has contributed to the idea of how this could be done, and the work in this thesis is also to be seen in the framework of the Semantic Web.

A very specific goal that is currently under investigation is combining WSD system finder for any particular word with WSD proper of the word concerned to maximize the accuracy.

REFERENCES

Agirre, E., Edmonds, P. (Eds.) 2006. Word Sense Disambiguation. Algorithms and Applications. *Text, Speech and Language Technology,* Vol. 33. Berlin/Heidelberg: Springer

Ahrens, K., Siaw-Fong C., Chu-Ren H. 2004. From Lexical Semantics to Conceptual Metaphors: Mapping Principle Verification with WordNet and SUMO. In J. D. Hong, L. K. Teng, W. Hui (Eds*.) Recent Advancement in Chinese Lexical Semantics: Proceedings of the 5th Chinese Lexical Semantics Workshop (CLSW-5), June 14-15*, 2004, Singapore: COLIPS, pp. 99-106

Allen, J. 1984. Towards a general theory of action and time. *Artificial Intelligence, 23*(2), pp. 123-154.

Anderson, J. 1983. A Spreading Activation Theory of Memory. *Journal of Verbal Learning and Verbal Behavior* 22(3), pp. 261-295.

Archambault, A., Gosselin, F., Schyns, P. 2000. A natural bias for the basic level? In L. R. Gleitman, A. K. Joshi (Eds.): *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, Mahawah, NJ: Lawrence Erlbaum, pp. 585-590.

Arnold, D., Balkan, L., Meijer, S., Lee Humphreys, R., Sadler, L. 1994. *Machine Translation: an Introductory Guid*e. London: Blackwells-NCC, pp. 132-137. Available in: http://www.essex.ac.uk/linguistics/clmt/MTbook/

Ayto, J. 1983. On specifying meaning. In R. R. K. Hartmann (Ed.): *Lexicography: Principles and Practice*, London: Academic Press, pp. 89-98.

Bloomfield, L. 1933. *Language*. New York, NY: Holt, Rinehart & Winston.

Boguraev, B. 1979. *Automatic resolution of linguistic ambiguities.* Computer Laboratory, University of Cambridge. Doctoral dissertation.

Brown, R.1958. How Shall a Thing be Called? *Psychological Review* 65, pp. 14-21.

Chomsky, N. 1975. *Reflections on Language*. New York: Pantheon Books.

Collins, A., Loftus, E. 1975. A spreading activation theory of semantic processing. *Psychological Review,* 82(6), pp. 407-428.

Cottrell, M., Verleysen, M. 2006. Advances in Self-Organizing Maps. *Neural Networks*, 19(6-7), pp. 721-722.

Cruse, D.A. 1986. *Lexical Semantics*. England: Cambridge University Press.

Cullen, J., Bryman, A. 1988. The Knowledge Acquisition Bottleneck: Time for Reassessment, *Expert Systems,* 5(3), pp. 216-225.

Curtis, J., Baxter, D., Cabral, J. 2006. On the Application of the Cyc Ontology to Word Sense Disambiguation. In G. Sutcliffe, R. Goebel (Eds.): *Proceedings of the Nineteenth International FLAIRS Conference*. Menlo Park, CA: AAAI Press, pp. 652-657.

Dempster, A., Laird, N, Rubin, D. 1977. Maximum likelihood from incomplete data via the EM algorithm*. J.Royal Statist. Soc. Ser. B,* 39(1), pp. 1–38.

Domingos, P., Pazzani, M. 1996. Beyond independence: conditions for the optimality of the simple Bayesian classifier. In L. Saitta (Ed.): *Proceedings of*

13th International Conference on Machine Learning. San Francisco, CA: Morgan Kaufmann, pp. 105-112.

Domingos, P., Pazzani, M. 1997. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning* 29 (2,3), pp. 103-130.

Dorow, B., Widdows, D. 2003. Discovering corpus-specific word senses. In A. Copestake, J. Hajic (Eds.): *Proceedings of EACL 2003*, Budapest, Hungary, pp. 79–82.

Eisner, J., Karakos, D. 2005. Bootstrapping without the boot. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT-EMNLP),* Vancouver, October. East Stroudsburg: PA, pp. 395-402.

Escudero, G., Marquez, L.,Martinez, D., Rigau, G., 2002. Machine-Learning Methods for WSD. *Working Paper 6.8 (First Year Reports and Deliverables).* Meaning project. Available in http://www.lsi.upc.es/~nlp/meaning/documentation/WP6.8.pdf.gz. Retrieved 16. 12. 2007.

Fauconnier, G. 1985. *Mental Spaces*. Cambridge, Mass: MIT Press.

Fensel, D., 2001. *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*. Heidelberg: Springer-Verlag.

Fillmore, C., J. 1982. Frame semantics. In Linguistic Society of Korea (Ed.) *Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Co., pp. 111-137.

Gale, W.A., Church, K.W., Yarowsky, D., 1992a. Estimating upper and lower bounds on the performance of word-sense disambiguation programs. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics,* 28 June-2 July 1992, University of Delaware, Newark, Delaware. Morristown, NJ: ACL, pp. 249-256.

Gale, W.A., Church, K.W., Yarowsky, D., 1992b. Using bilingual materials to develop word sense disambiguation methods. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation,* Montreal, Canada, pp. 101-112.

Gale, W.A., Church, K.W., Yarowsky, D., 1992c. Work on statistical methods for word sense disambiguation. In R. Goldman et al. (Eds.): *Working Notes of the AAAI Fall Symposium on Probabilistic Approaches to Natural Language.* Menlo Park, CA: AAAI Press, pp. 54 -60.

Gärdenfors, P. 2000. *Conceptual Spaces.* Cambridge, MA: MIT Press.

Gärdenfors, P., 2002. Cooperation and the Evolution of Symbolic Communication. *Lund University Cognitive Studies* 91.

Geeraerts, D. 1988. Cognitive grammar and the history of lexical semantics. In B. Rudzka-Ostyn (Ed.) Topics in cognitive linguistics. *Amsterdam Series in the Theory and History of Linguistic Science. Series IV. Current Issues in Linguistic Theory* 50. Amsterdam/Philadelphia: John Benjamins, pp. 647 - 677.

Geeraerts, D. 1993. Vagueness's puzzles, polysemy's vagueness. *Cognitive Linguistics,* 4(3), pp. 223–272.

Geertz, C. 1973. Person, time, and conduct in Bali. In C. Geertz (Ed.): *The interpretation of cultures.* New York: Basic Books, pp. 360–411.

Genesereth, M. 1991. Knowledge Interchange Format. In J. Allen et al. (Eds.): *Proceedings of the Second International Conference on the Principles of Knowledge Representation and Reasoning (KR-91).* Cambridge, MA: Morgan Kaufman Publishers, pp. 238-249.

Gruber, T.A., 1993. A Translation approach to portable ontology specifications. *Knowledge Acquisition,* 5 (2), pp. 199-220.

Harabagiu, S.M., Miller, G.A., Moldovan, D.I. 1999. WordNet 2 - A Morphologically and Semantically Enhanced Resource, In *SIGLEX 99: Standardizing Lexical Resources Workshop*, June 21-22, University of Maryland, USA.

Harris, Z. S.1954. Distributional Structure. *Word,* 10, pp. 146-162.

Hayes, P. 1977. On semantic nets, frames and associations. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, Cambridge, Ma: MIT, pp. 99-107.

Hirst, G. 1987. Semantic interpretation and the resolution of ambiguity. *Cambridge Studies in Natural Language Processing*. Cambridge, United Kingdom: Cambridge University Press.

Honkela, T. 1997. *Self-Organizing Maps in Natural Language Processing*. Helsinki University of Technology, Neural Networks Research Centre. Doctoral Dissertation.

Ide, N., Veronis, J. 1998. Word Sense Disambiguation: The State of Art, *Computational Linguistics, 24(1)*: 1-40.

Jackendoff, R. 1983. *Semantics and Cognition*. Cambridge, Mass: The MIT Press.

Jackendoff, R. 1990. *Semantic Structures*. Cambridge, Mass: The MIT Press.

Jackendoff, R. 1997. *The Architecture of the Language Faculty*. Cambridge, Mass: The MIT Press.

Järvinen, P. , Järvinen, A. 2000. *Tutkimustyön metodeista*. Tampere: Opinpaja Oy.
Järvinen, T., Tapanainen, P. 1997. A dependency parser for english,
Department of General Linguistics, University of Helsinki. Technical report

Java, S.., Finin, T., Nirenburg, S. 2006. Text understanding agents and the Semantic Web. In R. Sprague (Ed.): *Proceedings of the 39th Hawaii International Conference on System Sciences (HICSS'06),* Washington: IEEE Press, p. 62b.

Jurafsky, D., Martin, J. 2000. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. New Jersey: Prentice Hall.

Kaski, S. 1997. Computationally efficient approximation of a probabilistic model for document representation in WEBSOM full-text analysis method. *Neural Processing Letters*, 5(2), pp. 69-81.

Kilgarriff, A. 1997. “I don't believe in word senses.” *Computers and the Humanities* 31(2), pp. 91-113.

Kilgarriff, A. 1998. SENSEVAL: An Exercise in Evaluating Word Sense Disambiguation Programs. In *Proceedings of LREC*, Granada, May 1998, University of Granada, pp. 581 - 588.

Kilgarriff, A., Rosenzweig, J. 2000. Framework and results for English SENSEVAL. *Computers and the humanities*, 34(1-2), pp. 15 - 48.

Klapaftis, I.P, Manandhar, S. 2005. Google & Wordnet based Word Sense Disambiguation. In *Proceedings of the 22nd International Conference on Machine Learning (ICML05): Workshop on Learning and Extending Ontologies by using Machine Learning Methods*, Bonn, Germany, 7 – 11 August, 2005, pp. 23-27.

Kohonen, T. 2000. *Self-Organizing Maps* (Third edition). Berlin, Heidelberg, New York: Springer-Verlag.

Kohonen, T., Kaski, S., Lagus, K., Salojärvi, J., Paatero, V., Saarela, A. 2000. Organization of a massive document collection. *IEEE Transactions on Neural Networks, Special Issue on Neural Networks for Data Mining and Knowledge Discovery* 11(3), pp. 574–585.

Labov, W. 1973. The boundaries of words and their meanings. In C-J. Bailey, R. J. Shuy, (Eds.): *New Ways of Analyzing Variation in English.* Washington, D.C.: Georgetown University Press. pp. 340-373.

Lakoff, G. 1987. *Women, Fire, and Dangerous Things*. Chicago and London: The University of Chicago Press.

Lakoff, G., Johnson, M. 1980. *Metaphors We Live By*. Chicago and London: The University of Chicago Press.

Langacker, R. 1987. *Foundations of Cognitive Grammar*. Vol. 1, Stanford, CA: Stanford University Press.

Lenat, D. 1995. Cyc: A Large-Scale Investment in Knowledge Infrastructure. *Communications of the ACM,* 38 (11), pp. 32-38.

Lenat, D., Miller, G., Yokoi, T. 1995. CYC, WordNet, and EDR: Critiques and Responses. *Communications of the ACM,* 38(11), pp. 45-48.

Le Ny, J-F., 1995. Mental lexicon and machine lexicon: Which properties are shared by machine and mental word representations? Which are not? In P. Saint-Dizier, E. Viegas (Eds.): *Computational Lexical Semantics, Studies in Natural Language Programming,* Cambridge, UK: Cambridge University Press, pp. 50-67.

Linden, K. 2005. *Word Sense Discovery and Disambiguation*. University of Helsinki, Department of General Linguistics, Publications No: 37. Doctoral Thesis.

Malakhovski, L. V.1987. Homonyms in English dictionaries. In R. W. Burchfield (Ed.): *Studies in Lexicography*. Oxford, United Kingdom: Oxford University Press, pp. 36-51.

Manning, C., Schütze, H. 1999. *Foundations of Statistical Natural Language Processing*, Cambridge, MA: MIT Press.

Masterman, M. 1957. The thesaurus in syntax and semantics. *Mechanical Translation*, 4 (1-2), pp. 45–49.

McClelland, J., Rumelhart, D. 1981. An interactive activation of context effects in letter perception: part 1. An account of basic findings. *Psychological review,* 88(5), pp. 375-407.

Meyer, D., Schvaneveldt, R. 1971. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology,* 90(2), pp. 227- 234.

Mihalcea, R., 2003. The role of non-ambiguous words in natural language disambiguation. In *Proceedings of the Conference on Recent Advances in Natural Language Processing* (RANLP), Borovets, Bulgaria.

Mihalcea, R., Moldovan, D., 2001. EZ.WordNet: Principles for Automatic Generation of a Coarse Grained WordNet. In I. Russell, J. F. Kolen (Eds.): *Proceedings of the Fourteenth International Florida Artificial Intelligence Research Society Conference*, May 21-23, 2001, Key West, Florida, USA, Menlo Park, CA: AAAI Press, pp. 454-458.

Miller, G.A., Beckwith, R., Fellbaum, C.D, Gross, D., Miller, K. 1993a. *Five Papers on WordNet*. Technical report, Princeton, N.J: Princeton University.

Miller, G., Leacock, C., Randee, T., Bunker, R. 1993b. A semantic concordance. In *Proceedings of the 3rd DARPA Workshop on Human Language Technology*, Princeton, New Jersey, March 21 – 24, Morristown, NJ: ACL, pp. 303-308.

Minsky, M. 1974. A Framework for Representing Knowledge: MIT-AI Laboratory Memo 306. In P. Winston (Ed.): 1975. *The Psychology of Computer Vision*, New York, NY: McGraw-Hill.

Moldovan, D. Mihalcea, R. 2000. Using WordNet and Lexical Operators to Improve Internet Searches, *IEEE Internet Computing,* 4 (1), pp. 34-43.

Mooney, R.J. 1996. Comparative experiments on disambiguating word senses: an illustration of the role of bias in machine learning. In E. Brill, K. Church (Eds.): *Proceedings of the Conferences on Empirical Methods in Natural Language Processing (EMNLP-96),* Philadelphia, PA, pp. 82-91.

Murphy, G., Medin, D. 1985. The role of theories in concept coherence. *Psychological Review*, 92(3), pp. 289–316.

Nakov,P., Hearst, M. 2003. Category-based Pseudowords. In *Proceedings of HLT-NAACL'03 (Companion Volume),* Edmonton, Canada, May 2003, pp.67-69.

Niles, I., Pease, A. 2001. Towards a Standard Upper Ontology. In C. Welty, B. Smith (Eds.): *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001),* Ogunquit, Maine, October 17-19, 2001, New York, NY: ACM, pp. 2-9.

Niles, I., Pease, A. 2003. Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In *Proceedings of the 2003 IEEE International Conference on Information and Knowledge Engineering (IKE '03),* Las Vegas, Nevada, June 23-26, 2003, pp. 412 -416.

Nirenburg, S., McShane M., Beale, S. 2004. The Rationale for Building Resources Expressly for NLP. In *Proceedings of LREC-2004*.

Nirenburg S., Raskin V. 2004. *Ontological Semantics*. Cambridge: MIT Press.

Nunamaker, J.F., Chen, M., Purdin, T.D.M. 1991. Systems development in information systems research, *Journal of Management Information Systems* 7(3), pp. 89-106.

Pantel, P., Lin, D. 2002. Discovering word senses from text. In *Proceedings of ACM SIGKDD*, Edmonton, New York: ACM, pp. 613–619.

Pease, A., Niles, I., Li, J. 2002. The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. In A. Pease (Ed.): *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web*, Edmonton, Canada, July 28-August 1, 2002, Menlo Park, CA: AAAI Press.

Pollard, C., Sag, I.A., 1994. *Head-Driven Phrase Structure Grammar*. Chicago: The University of Chicago Press.

Quillian, M. R. 1968. Semantic Memory. In M. Minsky, (Ed.): *Semantic Information Processing*, Cambridge, MA: The MIT Press, pp. 216-270.

Resnik, P., Yarowsky, D. 1997. A perspective on word sense disambiguation methods and their evaluation. In M. Light (Ed.): *ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How?* April, Washington, D.C., pp. 79-86.

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

Robins, R. H. 1987. Polysemy and the lexicographer. In R. W. Burchfield (Ed.): *Studies in Lexicography*. Oxford, United Kingdom: Oxford University Press, pp. 52-75.

Rosch, E. 1973. Natural categories. *Cognitive Psychology* 4(3), pp. 328-350.

Rosch, E. Principles of Categorization. 1988. E. Smith (Ed.): *Readings in Cognitive Science, a Perspective from Psychology and Artificial Intelligence*. San Mateo, California: Morgan Kaufmann Publishers. pp. 312-322.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., Boyes-Braem, P. 1976. Basic objects in natural categories. *Cognitive Psychology,* 8(3), pp. 382-439.

Russell, S., Norvig, P. 1995 *Artificial Intelligence: A Modern Approach*, New Jersey: Prentice Hall.

Schank, R. 1972. Conceptual Dependency: A Theory of Natural Language Understanding, *Cognitive Psychology* 3 (4), pp. 552-631.

Schütze, H. 1992. Context space. In *AAAI Fall Symposium on Probabilistic Approaches to Natural Language,* Cambridge, MA., Menlo Park, CA: AAAI Press, pp. 113–120.

Schütze H, 1997. Ambiguity Resolution in Language Learning. *CSLI Lecture Notes* 71. Stanford, CA: Center for the Study of Language and Information

Schütze, H. 1998. Automatic word sense discrimination. *Computational Linguistics,* 24(1), pp. 97–123.

Sharman, R., Kishore, R., Ramesh, R. (Eds.). 2006. Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems. *Integrated Series in Information Systems*. Berlin/Heidelberg: Springer.

Simons, P. 1987. *A Study in Ontology*, New York: Oxford University Press.

Slator, B. M., Wilks, Y. A.1987. Towards semantic structures from dictionary entries. In *Proceedings of the 2nd Annual Rocky Mountain Conference on Artificial Intelligence*. 17-19 June 1987, Boulder, Colorado, pp. 85-96.

Smith, E., Medin, D. 1981. Categories and Concepts. *Cognitive Science Series* 4, Cambridge, MA: Harvard University Press.

Sowa, J. 2000. *Knowledge Representation*, Pacific Grove, CA: Brooks/Cole.

Sowa, J. 2002. Available in: http://suo.ieee.org/email/msg08310.html. Retrieved 16. 12. 2007.

Staab, S., Studer, R. (Eds.), 2004. *Handbook on Ontologies*, International Handbooks on Information Systems, Berlin/Heidelberg: Springer.

Studer, R., Benjamins, V. R., Fensel, D. 1998. Knowledge Engineering: Principles and methods. *Data & Knowledge Engineering,* 25 (1-2), pp. 161-197.

Subrata D., Shuster K., Wu, C. 2002. Ontologies for Agent-Based Information Retrieval and Sequence Mining. In S. Cranefield, T. Finin and S. Willmott (Eds.): Pr*oceedings of the Workshop on Ontologies in Agent Systems (OAS02), held at the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems,* Bologna, Italy, July 15-19, 2002. Tilburg, The Netherlands: CEUR Publications.

ten Hacken, P. 1990. Reading distinction in machine translation. In H. Karlgren, (Ed.): *Coling-90: Papers Presented to the 13th Conference on Computational Linguistics*, Helsinki University, Vol. 2, pp. 162-166.

UCL Machine Learning Repository. 2006. Available in: http://www.ics.uci.edu/~mlearn/MLRepository.html. Retrieved 16.12. 2007.

Wilks, Y. 1975. Preference semantics. In E. L. III Keenan (Ed.): *Formal Semantics of Natural Language*. Cambridge, UK: CambridgeUniversity Press, pp. 329-348.

Wilks, Y., Stevenson, M. 1996. *The grammar of sense: Is word sense tagging much more than part-of-speech tagging?* University of Sheffield, Sheffield, United Kingdom. Technical Report CS-96-05.

Witten, I.H., Frank, E. 2005. *Data Mining: Practical machine learning tools and techniques*, 2nd Edition, San Francisco: Morgan Kaufmann.

Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Blackwell.

Yarowsky, D. 1994. Decision lists for lexical ambiguity resolution: Application to accent restoration in Spanish and French. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*. Somerset, NJ: ACL, pp. 88–95.

Yarowsky, D. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*. Somerset, NJ: ACL, pp. 189–196.

Zwicky, A., Sadock., J.M. 1975. Ambiguity tests and how to fail them. In J. Kimball (Ed.): *Syntax and Semantics* 4. New York: Academic Press, pp. 1–36.

ORIGINAL PAPERS

Article I

BRIDGING THE WORD DISAMBIGUATION GAP WITH THE HELP OF OWL AND SEMANTIC WEB ONTOLOGIES

by

S. Legrand, P. Tyrväinen, H. Saarikoski 2003

Ontologies and Information Extraction International Workshop held as part of the *EUROLAN 2003, Bucarest 28th July – 8th August 2003*, pp. 29-35.

Bridging the Word Disambiguation Gap with the Help of OWL and Semantic Web Ontologies

Steve Legrand, Pasi Tyrväinen
Department of Computer Science
University of Jyväskylä
stelegra@cc.jyu.fi
Pasi.Tyrvainen@jyu.fi

Harri Saarikoski
Department of Linguistics
University of Helsinki
harri.saarikoski@kolumbus.fi

Abstract

Due to the complexity of natural language, sufficiently reliable Word Sense Disambiguation (WSD) systems are yet to see the daylight in spite of years of work directed towards that goal in Artificial Intelligence, Computational Linguistics and other related disciplines. We describe how the goal could be approached by applying hybrid methods to information sources and knowledge types. The overall aim is to chart the shorfalls of the present WSD systems related to the use of knowledge types and information sources in them. Real world ontologies and other ontologies in the Semantic Web will make a useful contribution towards the WSD knowledge base envisaged here. The inference capabilities inherent in Ontology Web Language (OWL) especially will have an important role to play in natural language disambiguation and knowledge acquisition. The emphasis is on ontologies as one of the important information sources for hybrid WSD.

1 Introduction

WSD methods have undergone changes and increased in number and variety in recent times, reflecting the requirements of many different types of uses WSD is put into. New types of information sources have appeared enabling the utilization of various types of knowledge incorporated in them. Nevertheless, there is no such thing as 100% WSD in any domain, however restricted. This, in fact, smacks of a goal that cannot be realized, taking into account that human WSD cannot reach that goal either.

1.1 Disambiguation Gap

However, we can try and get as close to the level of human WSD as possible. For this we need to specify the gap that currently exists between machine WSD and human WSD. This gap varies and is influenced by:

- *application domain* (machine translation, information retrieval, information extraction, knowledge acquisition, textual data mining, and natural language understanding among them).
- *information sources* available for those domains (machine-readable dictionaries, ontologies, corpora and their combinations among others), and
- *knowledge types* that these information sources incorporate (part-of-speech information, morphology, collocations, semantic associations, syntactic cues, sense frequencies, selectional preferences etc.).

Attempts to systemize the variety of sources and types have been made (Agirre and Martinez 2001, Ide and Véronis 1999) with some success, and attempts to unravel the knowledge types used for particular application domains, for example machine translation (Mowatt 1999), have also been made. The field now seems ripe for an approach that takes advantage of the potential of hybridisation in the multiplicity of these methods to optimise the effectiveness of WSD.

1.2 Dealing with the Disambiguation Gap

First, it is important to specify the disambiguation gap for various application domains by using our existing knowledge about information sources and knowledge types, and by experimenting with their combinations. The optimum mix varies from application to application: corpus statistical methods can support manual or semi-automatic knowledge methods. The relative weight of each method and knowledge type may be tested on corpora and defined. Hybrid methods (Ng and Lee 1996) and unsupervised methods (Yarowski 1995) have proved their mettle in comparative studies.

Second, one needs to identify those sources and types whose disambiguation potential is currently under-utilised due to their poor availability, costly acquisition, or insufficient appreciation. In particular, the use of ontologies for NLP tasks must be investigated, as ontologies will play a dominant part in the creation of the Semantic Web. OWL seems like a very good choice for disambiguation in which dense ontologies and disambiguation rules are used. There has been a rapid increase in the inference capabilites of Semantic Web languages with each layer added on RDF (RDFS → OIL → Daml-OIL) (Antoniou, 2002). At the time of this writing, the OWL standard is in its Last Call Working Draft phase (McGuinnes and van Harmelen, 2003), and may become a standard before the publication of this paper, and can then be added on top of the stack of the Semantic Web languages. As an indication of OWL's widespread acceptance at this early stage, one can cite some of the tools and technologies that have already been developed to take advantage of it, among them the knOWLer (2003) information management system and the OWL Converter (2003) converting from DAML-OIL format to OWL.

A further advantage of using the OWL together with the Semantic Web ontologies is in the distributed nature of those ontologies. A useable set of domain ontologies will take a considerable time to create: the task will never be completed, because new words and concepts are entering the vocabulary constantly. Hundreds and thousands of experts are needed to make sense of the world. The Open Source community exemplifies the way that the collaborative potential of likeminded people can be effectively harnessed. Already there are Open Source type development groups active in the Semantic Web.

Before Semantic Web came into being, a WSD system for machine translation, Mikrokosmos, based on knowledge-dense ontologies represented by TMRs (Text Meaning Representation) (Mahesh and Nirenburg, 1995) saw daylight: its word sense disambiguation success for all the words in a corpus, 97%, for both training and unseen text, and 90% for ambiguous words (Mahesh et al 1996) is encouraging, but the application lacks portability.

2 Hybrid Multilevel Disambiguation

In hybrid multilevel disambiguation, the idea is to disambiguate word senses using a mix of knowledge types and information sources, including real world knowledge in ontologies and their inference capabilities.

In the ontology related method, the correct ontology that corresponds to the document or document part domain is first detected. The domain itself can be automatically classified by using a hybrid method, for example a committee approach, as outlined in Hammond et al (2002). The domain ontology thus determined and located in the Semantic Web can then be used as the basis for the subsequent hybrid disambiguation.

For example, a paragraph such as:

"I sat in my old buggy It was very hot, so I turned on the engine, and drove under a tree to get cooler. Then I opened the window. "

can be disambiguated in several levels:

1. Morpho-syntactic level
2. Semantic level
3. World knowledge level

All of these levels overlap to some extent. As the result of morpho-syntactic level disambiguation, a sentence is annotated with POS (Part-Of-Speech) and morphosyntactic feature tags. From these we derive the syntactic function labels indicating whether the word is the subject, object, predicate, modifier or complement (Figure 1).

The semantic level can use these annotations together with the selectional preferences of the words themselves to clarify the meaning further. For example in the example paragraph above, 'buggy' is more likely to denote a kind of car than a horse-drawn carriage or a baby pram, the noun 'engine' regularly co-occurring with the noun 'car' in the same context. The Part-of ontological hierarchy in Figure 2 confirms their

I	i	@SUBJ
sat	sit	@+FMAINV
in	in	@ADVL
my	i	@A>
old	old	@A>
buggy	**buggy**	**@<P**
....	....	....
I	i	@SUBJ
turned	turn	@+FMAINV
on	on	@ADVL
the	the	@DN>
engine	engine	@<P
and	and	@CC
drove	**drive**	**@+FMAINV**
under	under	@ADVL
the	the	@DN>
tree	tree	@<P
......	......	......
I	i	@SUBJ
opened	open	@+FMAINV
the	the	@DN>
window	window	@OBJ

Figure 1. Syntactic functions (column 3) for the example paragraph according to FDG. Morphosyntactic feature tags are not shown here.

<u>*Sumo ontology (in bold):*</u>

- **Entity**
  - **Physical**
    - **Object**
      - **SelfConnectedObject**
        - **ContentBearingObject**
          - **LinquisticExpression**
            - **Word**
              - **Noun**
              - **Verb**
            - **Phrase**
              - **NounPhrase**
                - Object
                - Subject
              - **VerbPhrase**
                - Predicate
        - **CorpuscularObject**
          - **Artifact**
            - **Device**
              - **TransportationDevice**

<u>*Domain ontology fragments :*</u>

- Vehicle
  - MotorVehicleType
    - bus
    - car
      - Jeep
      - sedan
      - *buggy*
      - ......
  - MotorlessVehicleType
    - bicycle
    - horse carriage
      - buggy
      - ....
    - baby carriage
      - pram
      - buggy
      - .....
    - ....
  - MotorVehiclePart
    - engine
    - transmission
    - electrical system
    - body
      - roof
      - bumper
      - floor
      - door

Figure 2: SUMO top ontology (bold) subsuming the syntactic function and transportation domain ontologies (lighter color).

close relationship. Naturally, a rule specifying this would need to span over the sentence boundaries.

Even though we are aware now that the protagonist has started the engine of his car, it would be difficult by morpho-syntactic and semantic disambiguation alone to reason that the sentence following: *"Then I opened the window..."* would necessarily mean the car window. In this a real world ontology would be of great help. Apart from confirming that 'buggy' can be a type of 'car', it would confirm that the window, in this case, is a part of a car and not of a house, and that horse-drawn carriages or baby prams do not have engines. One could reason further that the car was a type of a vehicle etc, the selectionally preferred noun for the verb 'to drive' was 'a car' etc. If, however, the context of the paragraph was established to be that of golf, then of course the selectional preference for the verb 'drive' in 'I drove under a tree' would most likely change. These types of inferences could be drawn from ontologies built with the help of OWL, combined to minimize the word sense ambiguity.

The above is to give a rough idea of the way word sense disambiguation can be handled if OWL's inferencing capabilities are combined with traditional means of disambiguation. The example is just to illustrate the idea, and does not provide enough details for a complete disambiguation of the paragraph. It is easy to find fault with it by insisting, for example, that 'buggy', according to a definition found in many dictionaries is a small vehicle without windows and doors and with a roof mounted on the chassis. This would contradict what is said about the window above, unless one classified some off-road vehicles such as converted VW's (with windows) as buggies, which is also quite common. This further illustrates the importance of not relying excessively on any single information source or disambiguation method in trying to reduce the disambiguation gap.

Once the correct domain ontology and the position of the word in it denoting the concept are determined, the word can be matched with the foreign word in the same ontological structure if the purpose is to translate it. It may be that a house 'window' and a car 'window' in another language are denoted by two entirely different words, unlike in English, and therefore it is important to select the correct ontological concept.

3 FDG and Ontological Approach

We use a morpho-syntactic Functional Dependency Grammar (FDG, 2003) analyser as the baseline setter on which to found our research. The FDG analyser is based on ENGCG parser which, when combined with Xerox tagger, reached 98.5% structural disambiguation accuracy, outperforming all the other parsing combinations tested in the study of Tapanainen and Voutilainen (1995). FDG was selected for the present research mainly due to its accuracy. However, although its disambiguation error rate seems very small, it is still significant when considering natural language applications. Using the same formula as Abney (1996) it is easy to show that this word-based disambiguation rate, when applied to sentence-level, still needs some improvement to satisfy the requirements of natural language processing applications. If we assume that a sentence consists of 20 words on an average, the 98.5% word disambiguation accuracy is transformed into 26% error rate on a sentence level ($1-.985^{20} = 26\%$). For the purposes of machine translation this is clearly not yet adequate (1/4 of all the sentences erroneous even in ideal cases where the domain is restricted).

The current morpho-syntactic word sense disambiguation in FDG will soon be augmented with a semantic disambiguation module, which is likely to further improve the parser's accuracy. This is not sufficient, however. In addition to morphological, syntactic and semantic word sense disambiguation, real world knowledge is required for optimal understanding of a natural language. In our approach, the gap remaining in the disambiguation that cannot be bridged using the mix of currently available methods and their modifications is subjected to ontological disambiguation using real world distributed domain ontologies and SUMO upper ontology (Pease et al 2002) in the Semantic Web. Currently, most of these ontologies are in the RDF-based DAML-OIL format, but can be converted to OWL, the standard that is expected to replace DAML-OIL in the near future.

Farrar et al (2002) have suggested the addition of a general ontology for linguistic description (GOLD) to the SUMO upper ontology and published a draft version of it in the OWL format. They see it useful as a part of an expert system reasoning about language data, or as a part of an interlingua for machine translation system. We envisage being able to use their linguistic ontology to hold disambiguated morphological and syntactic data. Syntactic functions for nouns could indirectly be indicated in the case system portion of GOLD. However, these can also be plugged directly to the SUMO upper ontology (Figure 2), although their positioning under the SUMO's Phrase category may prove unsatisfactory in the long run.

```
<owl:Class rdf:ID="Car">
  <rdfs:subClassOf
rdf:resource="#MotorVehicleType" />
</owl:Class>

<owl:Class rdf:ID="Buggy">
  <rdfs:subClassOf rdf:resource="#Car" />
</owl:Class>

---------------------------------------------
<owl:Class rdf:ID="Engine">
  <rdfs:subClassOf
rdf:resource="#MotorVehiclePart" />
</owl:Class>

<owl:Class rdf:ID="Body">
  <rdfs:subClassOf
rdf:resource="#MotorVehiclePart" />
</owl:Class>

<owl:Class rdf:ID="Window">
  <rdfs:subClassOf rdf:resource="#Body"/>
</owl:Class>

---------------------------------------------

<owl:Class rdf:ID="Predicate">
  <rdfs:subClassOf
rdf:resource="#VerbPhrase" />
</owl:Class>

<Predicate rdf:ID="Drive">
   <selectionalPreferenceObject
   rdf:resource="#Car" />
</Verb>
```

Figure 3. Owl fragments connecting ontologies with subsumption and relational properties. Namespace declarations, superclass definitions, and property definitions are omitted.

The domain ontology holding the real world data and relations for the words can then be aligned to the SUMO ontology (Figure2) mainly through OWL subsumption and to the linguistic ontology using OWL property relations (Figure 3) as the glue: inferences can be drawn from the real world data to increase the disambiguation power of the linguistic ontology. For the purpose of alignment, SUMO also need be formatted to OWL. An agent-based application can then be used to manipulate the structures created for linguistic disambiguation.

The coding and property and class relations in Figure 3 are grossly simplified fragments forming part of an OWL document. The idea here is to show that the verb 'drive' selects a car rather than a baby carriage as its preferred noun phrase object. The word 'buggy' may subsequently be matched with 'car', and 'window' to the car body through their part-of relations. Similarly, both 'body' and 'engine' would be identified as parts of a motor vehicle. A comprehensive OWL statement about the verb 'drive' would have a set of preferentially weighed selectional preference entities to select from and a set of restrictions applied to it. Contextually closest (shortest arc distances in the ontology) selectional preferences would have the greatest preference weighing.

4 WSD Knowledge Base

Ontologies to be tested and designed for our optimal WSD include new and existing ontologies in the Semantic Web suitably modified and contain, for each concept, a dense network of subclass/superclass (eg. car is-a motorVehicle) relationships, property rules (eg. selectional preferences) and associative relations. Essentially, the WSD Knowledge Base will contain the differentiating factors between two senses of a word, which will disambiguate the sense of the target word. Synonym sets may be thought of as a differentiator, the sense's place in the Knowledge Base hierarchies and categories as another. Selectional preferences of the senses, and, of course, context word statistics - among other differentiators -, can also be used for disambiguation.

It is in this density and multiplicity of knowledge types and "sub-atomicity" (concepts are defined rigorously and adequately from within) that it contrasts the traditional, atomic (concepts are only defined in terms of their few external relations to related terms in network) ontologies. The Mikrokosmos ontology holds 5000 concepts with an average of 16 attributes and relations per concept (Mahesh et al. 1996). Our WSD knowledge base starts from that density and increases/decreases density until WSD is optimised. The result will be referred to as the *WSD Knowledge Base* for which we define each aspect of its construction and functioning. We will rigorously define the principles of designing such knowledge base, both in terms of quality (knowledge types required) and quantity (number of concept-internal definitions = information from knowledge types). As such, this research will also provide a feasible requirements specification for eventual implementation of the WSD system described.

One important application for our optimal WSD system is knowledge acquisition. It is precisely the lack of knowledge, and the high cost of acquiring dense knowledge bases and ontologies, that stands as the bottleneck in the way of knowledge-based NLP systems becoming more useful. WSD Knowledge Base may deliver a solution to both structural and semantic disambiguation tasks, and can as such be utilised in a multitude of NLP applications.

The WSD system could then be tested using corpora and test cases (disambiguable target words) from earlier research. Such starting points, and also points of comparison, could be Ng and Lee (1996) who tested their hybrid system on the senses of a single noun 'interest', Bruce and Wiebe (1994) who worked with the same noun, or Towell and Vorhees (1998) who tested some highly disambiguous words such as 'line' (noun), 'serve' (verb), and 'hard' (adjective). Another possibility, offering an equal amount of comparability to previous research, would be to examine the systems from the SENSEVAL-2 (2001) competition to see what knowledge types and information sources would most naturally and effectively disambiguate the target words.

5 Conclusion

This paper outlines the three main aspects in bridging the current disambiguation gap in WSD: application domain, information sources and knowledge types. There is a multiplicity of different domains, sources and types. Methods dealing with them have their limitations and can be partially overcome by combining the best of them in hybrid methods. It is important to determine the part of the disambiguation gap for language understanding that is dependent on knowledge aqcuisition.

Ours is an attempt to quantify the disambiguation potential of each information source and their contained knowledge types for each target word type. For example, if we find that what differentiates a word from another is synonym sets, points are added to the knowledge type and information source involved. The idea is to get an overall view on the most useful differentiating and disambiguating factors, knowledge types, and information sources in each particular case. Efforts in the KA and NL communities can then be better directed toward acquiring these information sources and knowledge types and developing more reliable hybrid WSD systems.

The FDG parser that we use in our morpho-syntactic and semantic disambiguation provides a mix of knowledge types (POS, morphology etc) to which we add selectional preferences and other types for the purpose of semantic / world knowledge disambiguation.

Ontologies as information sources are gaining momentum thanks to the emerging Semantic Web language specifications such as RDFS, DAML-OIL, and the most recent arrival, OWL, with its enhanced inference capabilites suitable for knowledge-based NLP. The use of OWL ontologies further reduces the disambiguation gap by allowing word sense disambiguation with the help of real world knowledge contained in Semantic Web domain ontologies.

It is still the early days. However, the OWL will become a standard soon, the SUMO upper ontology will be translated to OWL in a due course, and linguistic ontologies and ontologies from other domains (knowledge-saturated and knowledge-optimized ontologies) will be added and aligned with it. It is our hope that this paper can offer a glimpse of how Semantic Web, saturated ontologies, and OWL can contribute as one of the disambiguation methods used in hybrid WSD.

References

Abney, S., Part-Of-Speech Tagging and Partial Parsing, In: Church, K., Young, S., Bloothoof, G., Methods in Language and Speech. An ELSENET book, Kluwer Academic Publishers, Dordrecht, 1996.

Antoniou, G., Nonmonotonic Rule Systems on Top of Ontology Layers, *Lecture Notes in Computer Science*, 2342, Online publication: May 29, 2002. Available in: http://link.springer.de/link/service/series/0558/bibs/2342/23420394.htm

Agirre, E., and Martinez, D., Knowledge Sources for Word Sense Disambiguation, *Lecture Notes in Computer Science* 2166, Springer 2001.

Bruce, R., and Wiebe, J., Word-Sense Disambiguation Using Decomposable Models, In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, Las Cruces, New Mexico, 1994.

Farrar, S., Lewis, W.D., Langendoen, D.T., A Common Ontology for Linguistic Concepts, *Proceedings of the Knowledge Technologies Conference*, March 10-13, Seattle, 2002.

FDG. Conexor Functional Dependency Grammar. In: http://www.conexoroy.com/fdg.htm, Last accessed: June 2003

Hammond, B., Amit, S., Kochut, K, Semantic Enhancement Platform for Semantic Applications over Heterogenous Content, To appear in *Real World Semantic Web Applications, V.Kashyap and L.Shklar, Eds, IOS Press, 2002*. Available in: http://lsdis.cs.uga.edu/lib/download/HSK02-SEE.pdf

Ide, N., and Véronis, A., Word Sense Disambiguation: The State of Art, *Computational Linguistics*, Vol.24, No.1, March 1998, p.1-40

knOWLer. Ontology based information management system. In: http://taurus.unine.ch/GroupHome/knowler/wordnet.html. Last accessed: June 2003.

Mahesh, K., Nirenburg, S., Beale, S., Onyshkevych, B., Viegas, E., and Raskin, V., Word Sense Disambiguation: Why Statistics When We Have These Numbers? 1996.

Mahesh, K., and Nirenburg, S., A Situated Ontology for Practical NLP, in *IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing*, Aug. 19-21, Montreal, 1995.

McGuinnes, L.D., van Harmelen, F., Eds., OWL Web Ontology Language Overview, W3C Working Draft 31 March 2003, Available in: http://www.w3.org/TR/owl-features/

Mowatt, D., Types of Semantic Information Necessary. In *Machine Translation Lexicon, Conférence TALN 1999*, Cargèse, 12-17 July 1999

Ng, H. T. and Lee, H.B., Integrating multiple knowledge sources to disambiguate word sense: An exemplar-based approach, 1996. *In Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, Santa Cruz, California, 1996.

OWL Converter. In: http://www.mindswap.org/2002/owl.html. Last accessed: June 2003.

Pease, A., Niles, I., Li, J., (2002) The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. In *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web.* Available in: http://reliant.teknowledge.com/AAAI-2002/Pease.ps

Rigau, G., Magnini, B., Agirre, E., Vossen, P., and Carroll, J., MEANING: a Roadmap to Knowledge Technologies, 2002.

SENSEVAL-2. *Second International Workshop on Evaluating Word Sense Disambiguation Systems*. 5-6 July 2001, Toulouse, France. In: http://www.sle.sharp.co.uk/senseval2/

Tapanainen, P. and Voutilainen, A., Tagging accurately: don't guess if you don't know. Technical Report, Xerox Corporation, 1994.

Towell, G., Voorhees E.M., Leacock, C., Disambiguating Highly Ambiguous Words In *Computational Linguistics* Volume 24, Issue 1 / March 1998, p. 125 – 145

Yarowsky, D., Unsupervised word sense disambiguation methods rivaling supervised methods, *ACL95 - 33rd Annual Meeting of the Association for Computational Linguistics* 26-30 June 1995,Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, 1995.

Article II

CONNECTING LEXICAL KNOWLEDGE TO DISTRIBUTED REAL-WORLD KNOWLEDGE

by

S. Legrand, P. Tyrväinen 2003

In T. Cameron, C. Shanks, K. Holley (Eds.): *Proceedings of the Fifth Annual High Desert Linguistic Society Conference, 1-2 November 2002, Albuquerque, New Mexico*, pp. 109-120.

CONNECTING LEXICAL KNOWLEDGE TO DISTRIBUTED REAL-WORLD KNOWLEDGE IN THE SEMANTIC WEB

Steve Legrand and Pasi Tyrväinen
University of Jyväskylä, Finland

1. INTRODUCTION.

Often, something that might be regarded as obvious may sound amusing or completely counter-intuitive, especially if the listener has not yet completely internalised the language. For example, it is not self-evident for a novice language learner how to conceptualise a sentence such as (1):

(1) The author, *The Racing Times* contributing editor, Shirley Smith, is writing a book on the horse.

Nor is it obvious to a translation application of the current standards. The ambiguity arises in trying to decide whether Shirley Smith is writing a book about the horse, or whether she is writing a book while seated on the horse, or whether she is writing a book that rests on a horse. A native English speaker wouldn't have any problem in interpreting it the way intended, nor would he or she hesitate in rejecting (2) as incorrect, although a translation application might process it without a further ado:

(2) The horse, *The Racing Times* contributing author, Shirley Smith, is writing a book on the editor.

In (1), the context plays a prominent part in its acceptance, whereas (2) should be rejected on the basis of real-world knowledge; i.e., horses do not write books. Real-world knowledge is made up of myriads of everyday facts and their relationships. Without this knowledge it is practically impossible to make sense of the world when we communicate with each other. However, unlike syntax, pragmatics, and semantics, real-world knowledge is not linguistic knowledge. For this reason, it has proved problematic for linguists attempting to incorporate the required knowledge about language use into lexical entries.

Various attempts to create real-world databases have been made and are underway to address this and other problems. Lenat's (1995) CYC, a massive database of real-world knowledge under painstaking construction for the last two decades, has been criticised by Yuret (1997), among others, as being too explicit in its representation of knowledge in a single uniform framework and for using deduction as its main inference method. Locke (1990) warns about the dangers of creating systems that are accessible only to experts. To avoid the threat of 'ontological imperialism' the Semantic Web with its distributed ontologies and technologies seems a better alternative. XML-coded and RDF-Schema-based knowledge representation languages such as OIL and DAML-OIL and their future extensions (Fensel et al. 2001), with their increasing inference capabilities, are suitable for domain-specific, distributed ontology representation of real-world knowledge. If a larger centralized database is eventually needed, then Suggested Upper Merged Ontology (SUMO) could be used to unify disparate Semantic Web ontologies as discussed in (Pease et al. 2002).

This paper investigates the possibility of connecting distributed, real-world ontologies in the Semantic Web to linguistic knowledge (syntactic-pragmatic-semantic). RDF-based, real-world knowledge in the Semantic Web and elsewhere is divided into many distinct domain ontologies. How to utilize this growing repository of ontological, real-world knowledge for the purpose of disambiguation is the main topic of this article.

One way to go about utilizing this growing repository is by directly aligning lexical ontologies formed by semantic sorts (Dölling 1995) with the ontologies in the Semantic Web. Here, we use the Head-Driven Phrase Structure Grammar (HPSG) formalism to exemplify this. Jackendoff's (1983, 1990) two-tier lexical conceptual structure (LCS) is mapped to HPSG semantic sorts using methodology derived from Androutsopoulos and Dale (2000). These semantic sorts, forming an upper ontology, can then be mapped to distributed, real-world ontologies in the Semantic Web. This mapping is based on semantic similarity measures similar to those used by O'Harà et al. (1998). As a result, real-world knowledge, together with semantics, syntax, and pragmatics, can be integrated to constrain the structure-shared lexical entries.

Another way to accomplish the integration of real-world knowledge with lexical entries would be to use SUMO as a go-between to align semantic sorts with distributed real-world ontologies. This approach is considered only briefly: depending on the development of the Semantic Web, this approach might prove more viable in the long run.

The main motivation behind this research is to improve the accuracy of linguistic parsers to benefit linguistic applications used in translation, language learning, and other tasks that use parsers for disambiguation. Current parsing applications might seem adequate for these purposes having reached accuracies close to 100 % as demonstrated by Tapanainen and Voutilainen (1994). However, a word-based disambiguation error rate as small as 4 % is high enough to completely change the meaning of an average-length sentence, translating into a 56% per-sentence error rate (Abney 1996). Deployment of real-world knowledge together with linguistic knowledge in disambiguation will help to bridge this gap. As the expanding repository of that knowledge, the Semantic Web should be exploited.

2. HEAD-DRIVEN PHRASE STRUCTURE GRAMMAR (HPSG).

HPSG (Pollard and Sag 1995) is an integrated theory of natural language syntax and semantics drawing upon theories such as Categorial Grammar (CG) and Generalized Phrase Structure Grammar (GPSG) among others. Situation semantics and computer science have also contributed to its formulation. Unlike the transformational government-binding theory (GB) (Chomsky 1982), HPSG is NONDERIVATIONAL: its attributes of linguistic structure are related by STRUCTURE SHARING and not by transformational operations as they are in GB. The DECLARATIVE SYSTEM OF CONSTRAINTS employed in developing grammars based on HPSG ensures process neutrality; i.e., comprehension and production models are order-independent, because the constructs of these grammars can be applied in any required order. The principal type of object in HPSG is a SIGN (a word or a phrase) represented as a feature structure in an attribute-value matrix as shown in Figures 1 through 4.

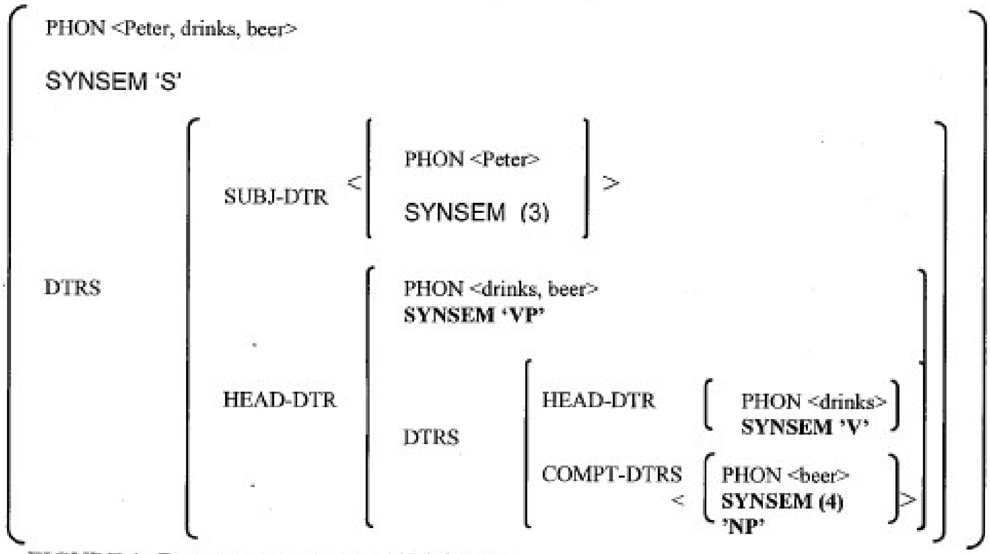
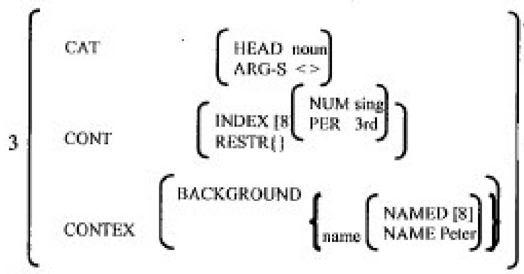


FIGURE 1: BASIC STRUCTURE OF HPSG SIGN.

The SYNSEM structures (DTRS: 3NP, 4NP, and VP) of Figure 1 are shown in more detail in Figures 2 through 4.



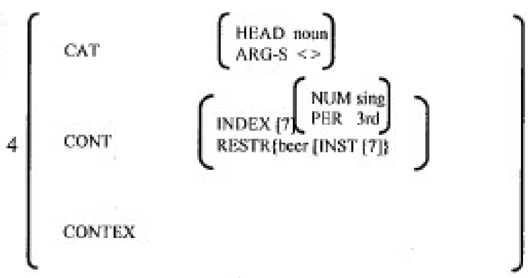


FIGURE 2: SUBJ-DTR 'PETER' IN SYNSEM.

FIGURE 3: COMPT-DTRS 'BEER' IN SYNSEM.

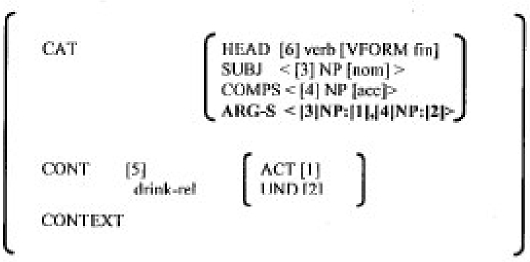


FIGURE 4: HEAD-DTR 'DRINKS' IN SYNSEM [V].

FIGURE 4: HEAD-DTR 'DRINKS' IN SYNSEM [V].

Here we will give just a brief account of the Figure 1 details relevant to this article. For a more detailed explanation, refer to (Pollard and Sag 1995) and (Davis 1997). Daughters (DTRS) are used in combinatorial saturation with the head as specified in the head's CATEGORY, which is part of the local syntax-semantics (SYNSEM) attribute. SYNSEM (Figures 2 through 4) consists of (1) CATEGORY, (2) CONTENT, (3) CONTEXT.

Roughly speaking, CATEGORY might be regarded as the syntactic component with its grammatical arguments, CONTENT as the semantic component, and CONTEXT as its pragmatic component/interface with a semantic dimension.

The verb *drinks* in Figure 1 acts as the HEAD of the phrase and forms part of the following CATEGORY composition:

HEAD	verb [fin]	(drinks)
SUBJ	< [3]NP[nom] >	(→Peter)
COMPS	< **[4]**NP[acc] >	(→beer)
ARG-S	< **[3]NP:[1],[4]NP:[2]** >	(→Peter, → beer)

SUBJ (subject list) and COMPS (complements list) are known as valence features, combining syntactically with the head in a combinatorial saturation. ARG-S (argument structure list), a concatenation of the SUBJ and COMPS lists, corresponds to the hierarchical argument structure of the predicate. The order of the arguments in ARG-S is related to their relative obliqueness with the least oblique argument occupying the leftmost position as follows:

subj < dir_obj < indir_obj < oblique_comp

To be of any use in cross-linguistic applications, the ARG-S structure in HSPG would have to be adapted to this hierarchy accordingly, for example, by using Schema 3 as detailed in (Pollard and Sag 1995:40). In Finnish, for example, the argument obliqueness does not usually conform to this positional hierarchy but is determined morphologically by the case endings. In the specification of ARG-S, the number after the colon indicates the content of the arguments, which, in this case, are shared with ACT and UND (Figure 4). ACT and UND are Davis's modifications to HPSG and to Jackendoff's (1983, 1990) conceptual structures. We take a closer look at these and the semantical nature of CONTENT in the following section, and deal with the pragmatics of CONTEXT after that.

3. LEXICAL CONCEPTUAL STRUCTURES.

CONCEPTUAL STRUCTURES can be understood as those structures of mind that have developed in living organisms during their evolution in interactions with the changing environmental conditions. These structures are reflected in the semantics and are partially captured in the syntax of a natural language. However, natural language is, by no means, the only expression of those conceptual structures: all of the other senses (hearing, vision, etc.) employ the same structures. The value of these structures lies in their universality: languages may vary, but as all the human beings have presumably similar evolutionary development behind them, conceptual structures should vary very little from region to region and between individuals. This gives us hope that *some* universal semantic structures encoded in syntax may, in fact, be found in all languages and could be employed productively in many natural language processing tasks such as language learning and translation.

Jackendoff's theory of semantic structures (Jackendoff 1990) or Lexical Conceptual Structures (LCS) being connected with conceptual structures is testable in the sense that if conceptual structures follow the universality constraint as claimed, then, for example, mapping the model of his semantic structures from one language to another should be possible, at least at a coarse level. The concepts and primitives he uses, although well structured, are still malleable for further elaborations as shown by Davis (1995), Verspoor (1997), and Wiese (2001) among others. Even though it might be possible to use only a part of these assumed universal conceptual structures, one could use ad hoc semantic struts and prostheses to temporarily replace those structures still not discovered or found to be erroneous. Jackendoff maps each major syntactic constituent of a sentence and the sentence itself into a conceptual constituent in the meaning of that sentence. His function-argument constituent categories include THING, EVENT, STATE, ACTION, PLACE, PATH, PROPERTY, and AMOUNT. This is illustrated by example (3):

(3) *Peter walked toward the sea.*

In (3) *Peter* and *the sea* are mapped to THING constituents, *toward the sea* to a PATH constituent, and the whole sentence to an EVENT constituent.. Jackendoff (1983:166) exemplifies another aspect of his theory with (4). This ambiguous sentence, then,

(4) *The mouse ran under the table.*

needs three different representations for its differing senses to make it unambiguous:

1. $[_{\text{Path}}$ TO $([_{\text{Place}}$ UNDER$([_{\text{Thing}}$ TABLE$])])]$ (goal)
2. $[_{\text{Path}}$ VIA $([_{\text{Place}}$ UNDER$([_{\text{Thing}}$ TABLE$])])]$ (path)
3. $[_{\text{Place}}$ UNDER$([_{\text{Thing}}$ TABLE$])]$ (location)

This sentence is ambiguous in English and has three interpretations. In the first interpretation the mouse goes under the table (goal). In the second, the mouse passes under the table and to the other side of it (path). In the third, the mouse moves around under the table (location). In Finnish, these ambiguities are resolved with the help of case endings as shown in (5 a-c):

(5)	a.	*Hiiri juoksi pöydän alle.*	(goal)
	b.	*Hiiri juoksi pöydän alta/ali/alitse.*	(path)
	c.	*Hiiri juoksi pöydän alla.*	(location)

Apart from lending support to the validity of Jackendoff's basic ideas, these sort of comparisons might help to hone the structures more suitable for cross-linguistic use. As far as the case endings and other grammatical constructs are concerned, they can be taken care of within the functions. For example, *alla* 'under' could be modified within the TO or VIA function to receive the correct case ending. Other functions affected would be called to modify their own arguments if necessary. As the syntax, in effect, disambiguates semantics, and lexical conceptual structures make this disambiguation more explicit, it is possible to take advantage of this in translation and other cross-linguistic pursuits.

Davis (1995) makes use of Jackendoff's ideas when he proposes mapping between HPSG's semantics and syntax in his thesis on multiple-inheritance lexical semantics. He incorporates Dowty's (1991) lexical entailments in Jackendoff's framework. For example, the attributes ACT and UND of CONTENT in Figure 4, corresponding to 'Peter' (actor) and 'beer' (undergoer), would express the following entailments:

ACT
- Volitional involvement in the event or state
- Perceives or has a notion of the other participants or events
- Causes an event or changes the state in another participant

UND
- Undergoes a state of change
- Incremental theme
- Causally affected by another participants

ACT and UND correspond to Dowty's Proto-agent and Proto-patient, respectively. Dowty's argument selection principle states that the argument for which the predicate entails the greatest number of Proto-Agent properties will be lexicalised as the subject of the predicate, and the argument having the greatest number of Proto-Patient entailments will be lexicalised as the direct object. Davis detects difficulties in tying syntactic argument assignments directly to the number of entailments and introduces these proto-roles into lexical representations as a mediating level between semantic entailments and syntactic arguments to remedy the situation. Davis also organises all his proto-role (ACT, UND, FIG, GRND, EFFECT, MEANS, ACC-EV) relations influenced by Jackendoff's thematic layer into sort hierarchies with monotonous inheritance.

Davis's approach demonstrates that semantics can be interfaced with syntax in the HPSG framework with the help of conceptual structures similar to those advocated by Jackendoff and incorporating proto-roles advocated by Dowty and Wechsler (1991) with some modifications. However, the problem of how to integrate real-world knowledge and pragmatics is not dealt with to any depth in his thesis. Davis (1995:58), nevertheless, implies that entailments, when mapped from semantics to syntax, can be viewed as real-world knowledge, being implications about certain classes of actions in the world.

4. Ontological enrichment.

As we have seen, the constraint-based HPSG framework can be used to produce extremely rich lexical entries for extensive linguistic manipulation. Syntax, semantics and pragmatics can all be used to constrain the lexicon. Davis (1995) has shown how to use multiple linguistic hierarchies to connect the semantics of verbs through modified conceptual structures to syntax, and Verspoor (1997) has shown how to connect noun semantics to syntax and pragmatics with the help of qualia. Copestake's (1992) view of the necessity to isolate linguistic knowledge from the real-world (encyclopaedic) knowledge has been observed in both and for a good reason: although any two languages vary, the underlying real-world knowledge remains the same for both languages no matter what words the speaker uses. Nevertheless, because real-world knowledge is often necessary to disambiguate the language, we need a mechanism to connect it somehow to lexical entries.

As noted in Section 2, in HSPG there is a rough correspondence between CATEGORY and syntactic structure, CONTENT and semantics, and CONTEXT and pragmatics, but there doesn't seem to be a similar component for real-world knowledge.

There are at least the following avenues one could use to introduce real-world knowledge systematically to HSPG framework:

1. Incorporate all lexical entries into a hierarchically organized, inheritance lexicon to form a super-ontology or a semantic net of ontologies.
2. Use the BACKGROUND attribute in CONTEXT to connect with real-world knowledge.
3. Use the conceptual attributes in CONTENT to connect with real-world knowledge.

Including real-world information in lexical entries themselves would enable the arrangement of lexical entries along the lines of hierarchies in real-world ontologies. However, as previously noted, it is better to keep the domain knowledge separate from linguistic knowledge in lexical entries. There are ontologies aiming at comprehensive coverage of common sense, real-world data, such as CYC (Lenat 1995), which may have potential if a comprehensive, common sense ontology connection to HSPG lexical entries is attempted.

There are also hundreds, even thousands, of real-world ontologies distributed around the net and created for specific domains with no ambitions to CYC-like common sense or semantic knowledge like that of WordNet (Miller et al 1990). For example, there are many RDF- and DAML-based ontologies available for the use in the context of the Semantic Web (http://semanticweb.org). The number of these ontologies is growing, and their development is undertaken in a distributed fashion. The results of this development could be taken advantage of by the HSPG framework.

If we follow the principle that real-world knowledge incorporated in HSPG lexical entries should be minimized, then we need some sort of interface or pointer system to connect real-world knowledge to linguistic knowledge. In practice, it is possible to include any relevant information in CONTEXT's BACKGROUND attribute. Another way to accomplish this might be to fine-grade either the nouns's qualia structure or the verbs' conceptual structures to such an extent that the distinction between linguistic and real-world knowledge is erased. However, this is undesirable as already pointed out.

Our proposal is to connect distributed, real-world ontologies to linguistic knowledge (syntactic-semantic-pragmatic) while keeping the two separate from each other. As explained in Section 2, the selectional restrictions imposed by the verb's ARG-S structure constrain the types of arguments that the head verb can accept. This is shown in Table 1.

A) Sentence:	'Peter' **[3] drinks** 'beer' [4]
B) Verb's Synsem's ARG-S:	< **[3]**NP:**[1]**,**[4]**NP:**[2]** > (See Figure 4)
C) Lexical Conceptual Structure:	[Event CAUSE ([Thing]**1**, [Event GO ([Thing LIQUID]**2**, [Path TO ([Place IN ([Thing MOUTH OF ([Thing]**1**)])])])])]

TABLE 1: VERB'S ARG-S RELATION TO THE PHRASE AND LEXICAL CONCEPTUAL STRUCTURES.

If Jackendoff's lexical primitives in the Action tier (Table 1: C) had a finer-graded internal structure, a semantic sort could be constructed. *Beer*, for example, could then be subsumed under Jackendoff's THING → PHYSICAL→ SUBSTANCE → LIQUID in a semantic sort having the structure shown in Figure 5 (grossly simplified and lacking in detail).

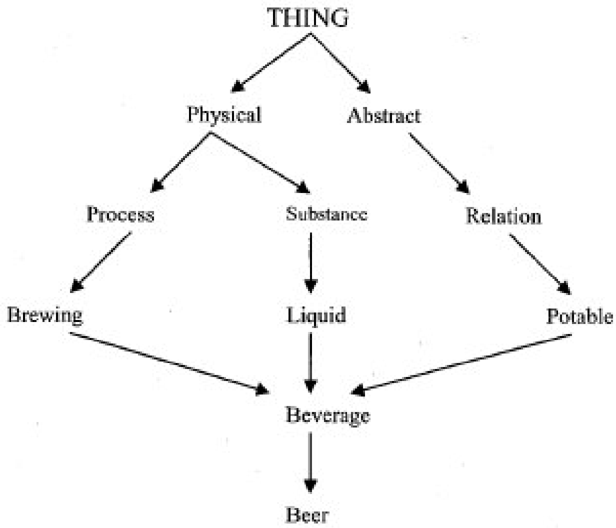


FIGURE 5: SIMPLIFIED, PARTIAL HPSG SEMANTIC SORT ONTOLOGY FOR NOUNS.

Selectional restrictions imposed by the verb's ARG-S (Table 1:B) could then help in determining whether a particular leaf in the semantic sort was acceptable. For example, *Peter drinks motorbikes* would not be an acceptable phrase because *motorbikes* is not a potable liquid.

HPSG's semantic sort forms an upper ontology, which, in our approach, is directly aligned with the suitable Semantic Web ontology (i.e., Foods or Drinks ontology of some sort in this particular case) with the help of Semantic Web Agents (Pease et al. 2002). In aligning the minimal noun ontology with the Semantic Web, similarity heuristics similar to those used in the MikroKosmos project to align MikroKosmos with WordNet can then be used (O'Hara et al. 1998).

Pollard and Sag's (1994) footnote in their Appendix supports this view of interfacing linguistic and real-world knowledge:

> Such knowledge should probably not regarded as strictly linguistic, but rather part of a distinct module of encyclopaedic knowledge with which linguistic knowledge interfaces. Likewise we do not declare here what features ('semantic roles') are appropriate for various subsorts of qfpsoa; but we assume that for each subsort of qfpsoa, the only sorts of values that are available for these features are either REF or PSOA.

In the present case, Index's ref attribute could be used to refer to the index of the particular entity in the HPSG's semantic sort, while the Background's parameterised state of affairs (psoa) could refer to the Semantic Web ontology with which the partial HPSG's semantic sort ontology would then need to be aligned. Figure 6 shows how *Beverage* is indexed to the HPSG semantic sort, and how the domain in the lexical entry points at the Semantic Web ontology with the ID of 412.

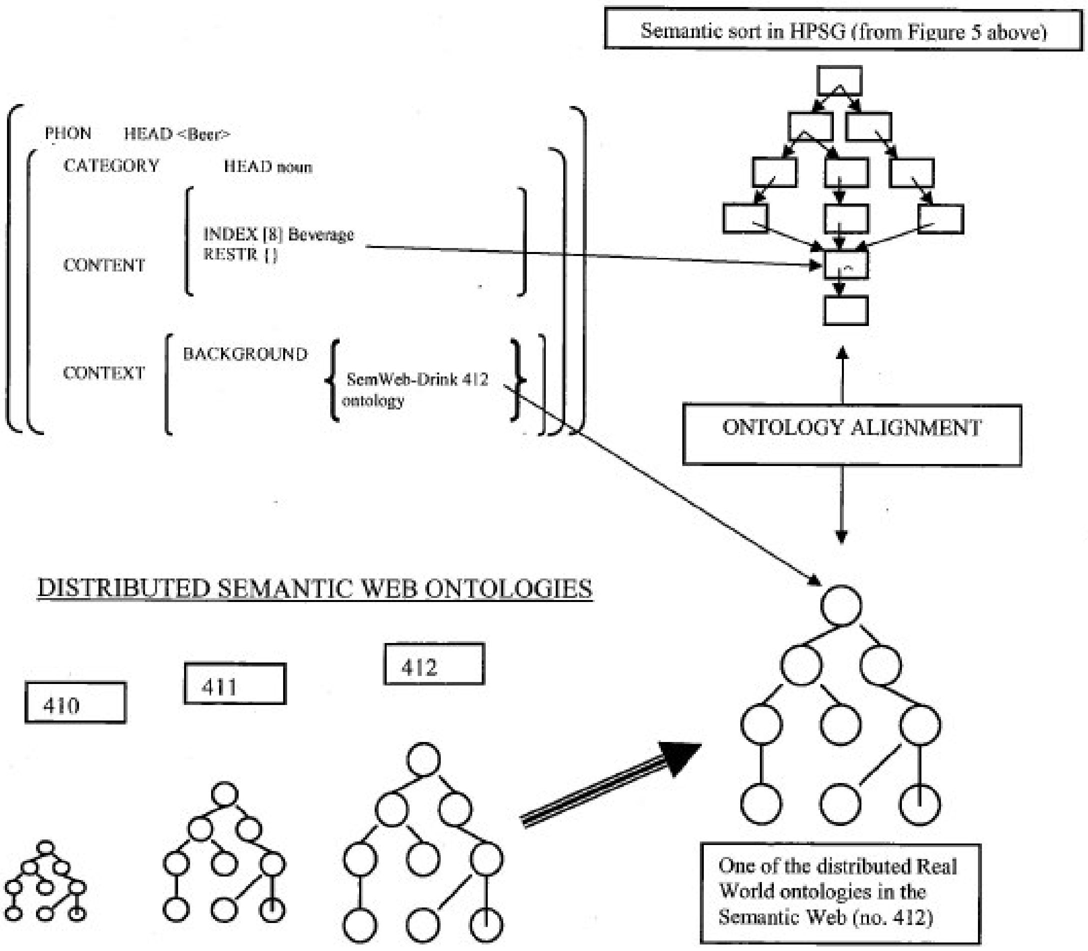


FIGURE 6: LEXICAL ENTRY FOR '*BEER*' AND ITS RELATION TO THE HPSG SEMANTIC SORT ALIGNED WITH THE SEMANTIC WEB ONTOLOGY.

In this way, Jackendoff's LCS's could be used to constrain the lexicon semantically and to connect it to the distributed Semantic Web ontologies for further disambiguation, exploiting world knowledge coded there. This would form the interface between linguistic and real-world knowledge, the real-world knowledge residing partly in an HPSG-based lexicon but mostly in the distributed ontologies. A comparable division in MikroKosmos between its lexicon and ontology was found to be highly desirable (Nirenburg et al. 1996).

When aligning the partial semantic sort ontology with the appropriate distributed Semantic Web ontology, semantic agents and the Agent Semantic Communications Service (ASCS) can be used in the task (Pease et al. 2002). The idea is that each RDF-based semantic ontology in the Semantic Web exhibits its own structure in a way that is readily interpretable by the Semantic Web agents.

Another way to connect real-world knowledge to a lexicon, be it based on the HPSG or another type of grammar, would be to create a semantic/world knowledge interface along the lines suggested above, but, instead of directly aligning it with the Semantic Web ontology deemed most appropriate, align it with SUMO currently under development (Niles and Peace 2001). SUMO is designed to act as a go-between top ontology in the semantic web to unify its disparate ontologies. This would make it easier to design a compatible semantic sort ontology for the lexicon (in this case HSPG-based).

5. CONCLUSION.

Davis (1995) and Verspoor (1997), basing their research on theories of Jackendoff (1983, 1990), Pollard and Sag (1994), Pustejovsky (1995), and others have demonstrated the feasibility of constraining lexical entries with the help of lexical conceptual structures and semantic sorts to facilitate the integration of syntactical, practical, and semantic knowledge. However, although both Davis and Verspoor admit the need for real-world knowledge in linguistic processing, they have not tackled this issue head-on because it falls outside the scope of their respective studies. Androutsopoulos and Dale (2000) have made concrete suggestions on how to use selectional restrictions in HPSG to achieve this.

In this article we have indicated some avenues to follow in order to include real-world knowledge with the constraining elements of HPSG. Nirenburg et al. (1996) have pointed out the importance of including some of the real-world knowledge in the lexicon in addition of having a more extensive ontology added to it. Their MikroKosmos project in a stand-alone environment has inspired and guided us in the attempt to extend this idea to a distributed environment in the Semantic Web. XML-coded, RDF-based, and domain-specific ontologies that are accessible to all can be developed in an extendible fashion in the Semantic Web. One of the latest developments, SUMO, will allow the combination of distributed, domain-specific ontologies under the common top ontology. This, in turn, will facilitate the alignment of lexical ontologies (semantic sorts), be they based on HPSG or any other grammar.

The motivation for this article and our proposal was outlined in the introduction. HPSG could be used in the integration of real-world knowledge as an additional constraint into lexical entries. The resulting lexicon could then be used for disambiguation in order to improve the accuracy of current parsers. This, in turn, would lead to more reliable language learning and translation applications.

Although we have specified the requirements of HPSG and LCS here, at the same time we have tried to avoid getting too deep in their specific details: there are other grammars, conceptual frameworks, and techniques that could be used instead. The important point we have tried to raise is that the distributed ontologies in the Semantic Web can be used for the purpose of adding real-world knowledge to linguistic knowledge as a constraint for the purpose of disambiguation.

REFERENCES

Abney, S. 1996. Part-Of-Speech Tagging and Partial Parsing, In: Church, K., Young, S., Bloothoof, G., 1996. *Methods in Language and Speech*. An ELSENET book, Dordrecht: Kluwer Academic Publishers. Available in: <http://www.sfs.nphil.uni-tuebingen.de/~abney>.

Androutsopoulos, I., Dale, R. 2000. Selectional Restrictions in HPSG. *Proceedings of the 18th International Conference on Computational Linguistics (COLING)*, Saarbrücken, Germany, 31 July – 4 August 2000, 15-20.

Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Copestake, A. 1992. The Representation of Lexical Semantic Information. PhD Thesis. Sussex University. Cognitive Science Research Paper CSRP 280.

Davis, A. 1995. Lexical Semantics and Linking in the Hierarchical Lexicon. PhD Thesis. Stanford University.

Dorr, B.J. 1992. The Use of Lexical Semantics in Interlingual Machine Translation. *Journal of Machine Translation* 7 (3): 135-193.

Dölling, J. 1995. Ontological domains, semantic sorts and systematic ambiguity. *International journal of human-computer studies* 43: 785-807.

Dowty, D. 1991. Thematic Proto-Roles and Argument Selection. *Language* 67 (3): 547-619.

Fensel, D., van Harmelen, F., Horrocks, I., McGuinness, D.L., Patel-Schneider, P. 2001. OIL: An Ontology Infrastructure for the Semantic Web, *IEEE Intelligent Systems*, March/April 2001: 38-45.

Jackendoff, R. 1983. *Semantics and Cognition*. Cambridge, Mass.: The MIT Press.

Jackendoff, R. 1990. *Semantic Structures*. Cambridge, Mass.: The MIT Press.

Lenat, D.B. 1995. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38:33-38.

Locke, C. 1990. Common Knowledge or Superior Ignorance? Robotics Institute. Carnegie Mellon University. A modified version of the paper that appeared in *IEEE Expert*, Dec., 1990. <http://www.panix.com/~clocke/ieee.html>.

Miller, G., R. Beckwith, C. Fellbaum, D. Gross and K. Miller. 1990. Five Papers on WordNet. *CSL Report 43*. Cognitive Science Laboratory. Princeton University.

O'Hara, T., Mahes, K., and Nirenburg, S. 1998. Lexical acquisition with WordNet and the Mikrokosmos ontology. In Sanda Harabagiu, editor. *COLING-ACL'98 Workshop on Usage of WordNet in Natural Language Processing Systems*. Université de Montréal, 94-101.

Niles, I., & Pease, A. 2001. Toward a Standard Upper Ontology, in *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*.

Nirenburg, S., Beale, S., Mahesh, K., Onyshkevych, B., Raskin, V., Viegas, E., Wilks, Y., Zajack, R. 1996. *Lexicons in the MikroKosmos Project, in Proceedings of the AISB'96 Workshop on Multilinguality in the Lexicon, UK, April 1-2.*

Pollard, C. and Sag, I.A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: The University of Chicago Press.

A. Pease, I. Niles, and J. Li. 2002. The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. Presented at *AAAI-2002 Workshop on Ontologies and the Semantic Web.* <http://reliant.teknowledge.com/AAAI-2002/Pease.doc>.

Pustejovsky, J. 1995. *The Generative Lexicon*. Cambridge, Mass.: The MIT Press.

Tapanainen, P. and Voutilainen, A. 1994. Tagging accurately: don't guess if you don't know. Technical Report, Xerox Corporation.

Verspoor, C.M. 1997. Contextually-Dependent Lexical Semantics, PhD Thesis. The University of Edingburgh.

Wechsler, S. 1991. Argument Structure and Linking. PhD Thesis. Stanford University.

Wiese, H. 2001. Modelling semantics as a linguistic interface system. Unpublished manuscript. <http://www2.rz.huberlin.de/linguistik/institut/wiese/publications/index.htm.>

Article III

A HYBRID APPROACH TO WORD SENSE DISAMBIGUATION: NEURAL CLUSTERING WITH CLASS LABELING

by

S. Legrand, J.R.G. Pulido 2004

In P. Buitelaar, J. Franke, M. Grobelnik, G. Paaß, V. Svatek (Eds.): *Knowledge Discovery and Ontologies (KDO-2004) workshop in15th European Conference on Machine Learning (ECML) Pisa, Italy, September 24, 2004,* pp.127-132.

A Hybrid Approach to Word Sense Disambiguation: Neural Clustering with Class Labeling

Steve Legrand[1] and JRG Pulido[2]

[1] Department of Computer Science, University of Jyväskylä, Finland,
stelegra@cc.jyu.fi,
[2] Faculty of Telematics, University of Colima, Mexico,
jrgp@ucol.mx

Abstract. By combining a neural algorithm with the WordNet lexical database we were able to semi-automatically label the groups of items clustered in a multi-branched hierarchy, paving way for the use of neural algorithms together with ontological knowledge in word sense disambiguation tasks.

1 Introduction

When classifying documents on the web into different categories, one needs to have some knowledge about the domain of each document, i.e., what the document is about. For the purpose of web page categorization this may be adequate. When disambiguating words in a document, however, document classification into rough domain categories of various kinds, although often helpful, is clearly not adequate for the task on its own. For example, if the document deals with computing and contains the word *mouse*, one could presumably make use of the computing domain ontology to conclude that *mouse* refers to a computer peripheral and not to a rodent. However, it should be kept in mind that the document may deal with several domains and different contexts within the domains: the word *mouse* in the document in question may refer to an animal in one part of the document and to a computer mouse in another.

The premise of our investigation is that once we have found the precise concept that corresponds to a word in a conceptual hierarchy, then the word sense is deemed disambiguated. This statement, however, needs some clarification. The principal qualifications are that the density of the conceptual hierarchy must be of an appropriate granularity for the task and that the hierarchy needs some sort of mechanism to acquire novel concepts that do not yet occur in the hierarchy. To be able to map a word to its concept in the conceptual hierarchy and to label it correctly, we need to know about the concepts' neighborhoods and the labeling of the nodes in those neighborhoods.

2 Concept Discrimination and Labeling

Several problems arise when contemplating the use of document maps, produced by a neural net, in word sense disambiguation (WSD). In conventional self-organized map (SOM) clusters [3], the idea is to present multi-dimensional vector clusters in two-dimensional space, the similarity between concepts presented as the distance between the clusters. This is not the preferred form in the presentation of conventional ontologies: they are usually presented as concept hierarchy trees, although their underlying structure is more complex. If we look at the conventional SOM-produced cluster map (Figure 1a), we see clusters of words with no concept hierarchy visible and no labeling of the various sense groups.

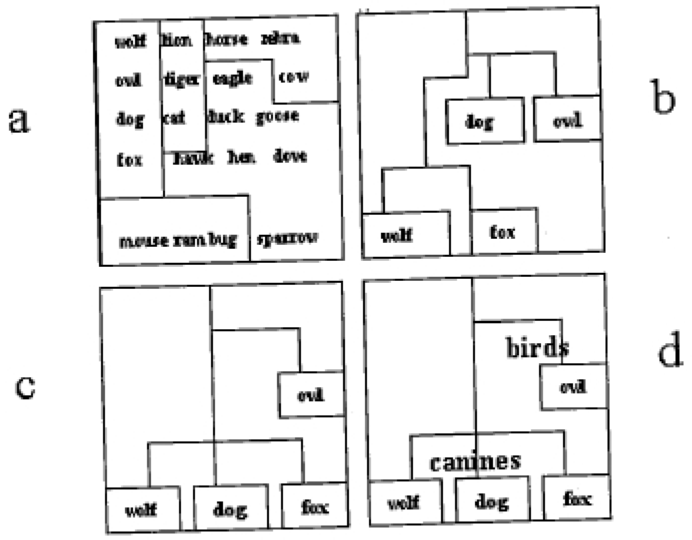


Fig. 1. a) SOM-produced clustering, b) binary hierarchy, c) multi-branching hierarchy, d) labeled multi-branching hierarchy

In Figure 1a it appears that most of the birds are grouped together, apart from *owl*, which is grouped together with canines. Felines are correctly grouped close to each other and the same applies to the suborders of ungulates (*horse*, *zebra* and *cow*). *Mouse*, *ram* and *bug* do not seem to bear much resemblance to each other except for not belonging to any other group. They would not seem out of place in a computing domain either. All the groupings in Figure 1 are based on feature vectors, the features of which can vary depending on the purpose and can affect the groupings. For example, if size and shape attributes figured prominently, then the animals would be grouped according to their size and shape rather than other more commonly used zoological attributes. Hierarchies based on binary division (Figure 1 b) often produce incorrect results as more than two divisions may be required to accommodate all the instances. This can be corrected by employing a multi-branched hierarchy (Figure 1 c). Nevertheless, unless we can reliably name (label) the divisions produced (Figure 1 d) we can only talk about sense discrimination [6] rather than sense disambiguation.

To tackle the task of producing a hierarchy of the type depicted in Figure 1d (labeled multi-branched hierarchy) semi-automatically, we combined a neural algorithm with a lexical database. With the neural algorithm we divided the dataset into groups in a multi-branched hierarchy. The lexical database was used to find the most probable labels for those groups.

3 WordNet Labeling Algorithm

WordNet 2 defines itself as "a machine-readable lexical database organized by meanings". In WordNet, English nouns, verbs and adverbs are organized into synonym sets representing lexical concepts [2]. These sets are linked by relations such as hypernym, synonym, antonym and so forth.

DGSOT [5] is a tree-structured self-organizing neural network that constructs a hierarchy from top to bottom by division, optimizing the number of clusters at each hierarchical level. While DGSOT is impressive in its ability to organize data into their respective groups close to the desired granularity by modification of its parameters, it would also need some kind of sense-labelling mechanism in order to adapt it to word-sense disambiguation.

The idea behind our overall algorithm is to first roughly group the items (animals) and then label them using the WordNet's hypernym-hyponym conceptual hierarchy. Once DGSOT has grouped the items, we find all the ancestries of each item according to WordNet. For example, in a group consisting of *aardvark* and *bear* we have one possible ancestry for *aardvark* in DGSOT results:

1. *entity* <- *object* <- *living thing* <- *organism* <- *animal* <- *chordate* <- *vertebrate* <- *mammal* <- *placental* <- aardvark

For *bear*, however, two different ancestries to choose from are given by DGSOT:

1. *entity* <- *object* <- *living thing organism* <- *animal* <- *chordate* <- *vertebrate* <- *mammal* <- *placental* <- *carnivore* <- bear

2. *entity* <- *object* <- *living thing organism* <- *person* <- *capitalists* <- *investor* <- bear

We have to decide which is the most likely ancestry that *bear* belongs to given the context in which the items are grouped. For this we use a simple majority voting scheme by counting the total number of times each ancestry node occurs in all of the group's ancestries and then counting the total accumulated points for each ancestry. For example, the ancestry node *entity* occurs three times in the group's (*aardvark*, *bear*) ancestry (value = 3), whereas *investor* occurs only once (value = 1). We sum up these points for each individual ancestry and if there are more than 1 ancestries for an item, we select the one having the highest voting score. For *bear* we would have the following values for its two ancestries:

3+3+3+3+2+2+2+2+2+1 = 23 (carnivore)
3+3+3+3+1+1+1 = 15 (investor)

As the result of this calculation the ancestry tree denoting *bear* as a carnivore is retained and added to the labelled hierarchy, whereas the ancestry tree denoting *bear* as an investor is discarded.

Items completely misgrouped by the neural algorithm would retain the sense that would have the highest points in their ancestry count. These misgrouped items would, nevertheless be correctly separated from the rest of the group in the resulting labeled hierarchy.

The example above consists only of two items, but we found it working well for groups of much greater size. The deceptive simplicity of our algorithm bears testimony to the discriminatory powers of DGSOT which makes the downstream labelling with the help of conceptual hierarchy a straightforward matter.

4 The Experiment with the Zoo-dataset

Our long-term aim is to use feature-vectors of free text in word sense disambiguation with hybrid algorithms [4] and for semi-automatic creation of ontology components for the semantic web [1]. We chose a simple Zoo-dataset to test the viability of the concept. Unlike free text, this dataset deals only with the animal domain, and consists of nouns only. Further modifications are necessary to our application to deal with the general WSD issues, and we are currently working on this.

The Zoo dataset is available at UCL machine learning repository [7]. It consists of 101 instances, and in its original form has 18 attributes. However, the first attribute names the data item itself, and was not used as an attribute in our algorithm. The 18th attribute classifies the items in 7 categories and was not used, because our aim was to do the classification of the items ourselves. Nevertheless, as only Boolean values were used by us, this would not have made any difference to the outcome. The 14th attribute (legs) is numeric, but as its numeric values were converted to Boolean it did not have as great effect as the numeric value would suggest. We were left with 16 attributes (hair, feathers, eggs, milk, airborne, aquatic, predator, toothed, backbone, breathes, venomous, fins, legs, tail, domestic, and catsize). Only 96 out of 101 instances in the data sets were used in the experiment because 5 of the instances were not found in WordNet (*fruitbat*, *pitviper*, *sealion*, *seasnake*, and *slowworm*).

Overall, all the animals in the resulting WordNet-based taxonomy were correctly classified. Figure 2 shows a partial taxonomy of some of the groups including parts of the ancestral paths of *aardvark* and *bear*.

When WordNet first senses were selected as a baseline to compare with, we found that only 73% of the items were correctly classified as animals by WordNet. Our approach, correctly classifying all the items of the dataset occurring in WordNet, greatly improves the accuracy of the WordNet baseline classification results.

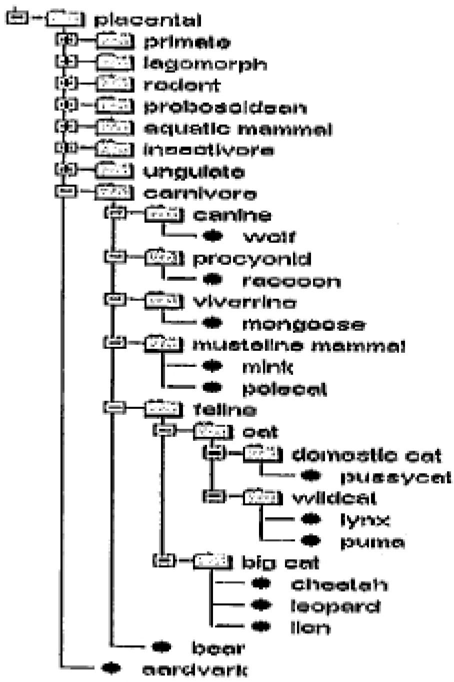


Fig. 2. A multi-branched labeled hierarchy based on WordNet. Note: the groups are labeled.

5 Discussion

Although all the dataset items in our experiment were correctly labeled, we hesitate to claim a 100% general labeling accuracy at this stage. One needs to keep in mind that the data set on which the algorithm was applied is quite small, and consists only of nouns. The algorithm needs to be applied on larger and more heterogenous datasets as our aim is to apply it to free text in the future.

One might think of using the majority voting method on the data directly, by creating a huge filtering matrix and then eliminating categories based on their ancestry as described earlier. However, if DGSOT or a similar neural algorithm were not used to preprocess the data to the approximate categories, the domains of the categories might become confused, i.e., the majority voting approach would not work on a very heterogenous set. Consider, for example, the following line of data:

bug, mouse, ram, house, car

The first three labels could describe either computer parts or animals. It is hard to know which one is which unless associated with its categorical neighbours, i.e:

keyboard, console, mouse or
bug, tiger, dog, swine, mouse, horse

The precategorization by DGSOT in our algorithm works by organizing the data items so that the domains roughly correspond to groups, allowing majority voting within those groups. DGSOT's output discriminates senses, whereas our approach disambiguates them. Class labelling together with WordNet rearranges and allows the labeling of all the nodes in the tree structure. Although WordNet is used here, any other concept hierarchy could be considered.

We have effectively demonstrated that neural processing of datasets to categories does not need to stop at sense discrimination, but that word sense disambiguation through class labeling can be attempted. We have also shown that taxonomies/ontologies, such as WordNet, can be used in combination with neural algorithms to achieve this.

References

1. Elliman, D., Pulido, J.G.R.: 2002. Visualizing ontology components through Self Organizing Maps, 6th International Conference on Information Visualization: 434-438. D. Williams, ed.
2. Fellbaum, C. ed.: 1998. WordNet: An Electronic lexical database, MIT Press, Cambridge, MA.
3. Kohonen, T.: 2001. Self-Organizing Maps, Information Sciences Series, Springer-Verlag, Berlin, 3rd edition.
4. Legrand, S., Tyrvainen, P., and Saarikoski, H.: 2003. Bridging the Word Disambiguation Gap with the Help of OWL and Semantic Web Ontologies. Proceedings of EUROLAN 2003, the Semantic Web and Language Technology, Bucharest, Romania, July 28 - August 8, 2003.
5. Luo, F., Khan, L., Yen, I-L., Bastani, F.: 2004. A Dynamical Growing Self-Organizing Tree (DGSOT) For Hierarchical Clustering, Working Paper, Available at: http://utdallas.edu/ luofeng/DGSOT.doc.
6. Schütze, H.: 1998. Automatic Word Sense Discrimination, Computational Linguistics, Volume 24, Number 1: 97-124.

Article IV

WORD SENSE DISAMBIGUATION WITH BASIC-LEVEL CATEGORIES

by

S. Legrand 2006

Advances in Natural Language Processing. In A. Gelbukh (Ed.): *Research in Computing Science*. Vol.18, IPN, Mexico, 2006, pp. 71-82.

Word Sense Disambiguation with Basic-Level Categories

Steve Legrand

Department of Computer Science, University of Jyväskylä, Finland
stelegra@cc.jyu.fi

Abstract. Research in basic-level categories has provided insights that can be beneficially used in word-sense disambiguation. Human beings favor certain categories over others and this is reflected in their use of language. Parts and functions categories are especially useful in providing contextual clues. The disambiguation effort in this article concentrates on the two main senses of the word "palm." The results are encouraging and indicate that basic-level categories will have a role to play in computational linguistics.

1 Introduction

If a word has only one sense, a non-native speaker can confirm its meaning by a quick look at a dictionary. Most of the words do have, however, more than one sense, and both the native and the non-native speaker need to use the word context in order to find its correct sense. For example, when we look at the sentence,

There was a large blister on the heel of his right palm.

it is obvious to us that the word *palm* refers to a body part rather than to a tree or a handheld computer. The words *blister*, *heel*, *his*, and *right* when combined in a certain way point us towards the correct meaning.

Most of the automated disambiguation techniques, one way or another, are context-based, making use not only of the words themselves, but also of the part-of-speech information, word order, document genre and so on. Generally, we can say that these techniques are justified by our observations that certain words do co-occur quite regularly with each other within certain contexts. This notion has been used somewhat heuristically in automated word sense disambiguation, and often there is no reference to any cognitive disambiguation mechanism that could have been involved. Nevertheless, it is not disputed that context plays a very important part in the word sense disambiguation by our cognitive faculties.

The question arises: what is this human disambiguation mechanism like if it exists, and would it be possible to mimic and exploit it in automated word sense disambiguation? Is it rooted in our biology, and consequently reflected in our cognitive abilities, including our ability to categorize? The classical view of categories is often interpreted as meaning that things belong to the same category only if they have certain properties in common. It might seem that car parts such as a wheel and an engine do not share any properties, therefore should one assume that they cannot belong to the

© A. Gelbukh (Editor)
Advances in Natural Language Proccesing
Research in Computing Science 18, 2006, pp. 71-82

same category? On the closer inspection one can, however, discover, that they have at least one common property, and that is that they are parts of a car. So partonomy can create categories of things that apparently do not have much to do with each other. In fact, we can divide and subdivide our universe in so many different ways that what we know as classical categorization may prove inadequate for many tasks, including word sense disambiguation. Using the Family Resemblance Theory [30], basic-level categories [3] and experiments demonstrating these theories [21], Lakoff [13] challenges the classical view of categorization, proposing to correct it with a move to idealized cognitive models (ICMs) based, to a large extent, on prototype-level categories.

Here we will demonstrate that the type of categorization, "a human view of the world", that Rosch and Lakoff favour, may indeed be reflected in the language that we use to describe things, and, therefore, can benefit word sense disambiguation. The work is still at its preliminary stages, and the purpose of this paper is merely to explain the theoretical basis behind it and illustrate it with a simple example.

In what follows, we will go briefly through the concepts of basic-level categories (Section 2), idealized cognitive models (Section 3) and ontology structures (Section 4) before explaining how to use refined InfoMap [11] results for creating an ontology that may be more suitable for word sense disambiguation (Section 5). Finally, to exemplify our suggested approach, the two major senses of the word *palm* are disambiguated (Section 6). A more extensive study is underway, the results of which will be published shortly.

2 Basic-Level Categories

Roger Brown [3] explained his notion of a “first level” as a kind of category which allows children to learn object categories and name them, but which, as a category, falls somewhere between the most general and the most specific level. Later Rosch [21, 22] designed a series of experiments in which she demonstrated that the basic-level categories, as she started calling them, were somewhat inconsistent with the classical theory of categories, and she explained their specific properties:

From the point of view of human cognition, the categories seem to be divided roughly into three kinds: superordinate (*furniture*), basic-level (*chair, table, lamp*), and subordinate (*kitchen chair, living-room chair / kitchen table, night table / floor lamp, desk lamp*). The basic-level objects have most of the attributes that are common to all members of the category and they share the least number of attributes with other, contrasting categories. Category membership is also influenced by family resemblances [30] to prototypical members. Archambault et al [1] in their brief review of literature selected the following (most of it also investigated by Rosch or based on her research) as the most important issues to note about basic-level categories:

- Categories at the basic-level are verified fastest.
- Objects are named faster at the basic than at the subordinate level.
- Objects are preferentially named with their basic-level names.
- Basic-level names are learned before subordinate names.
- Basic-level names tend to be shorter.

Tversky and Hemenway [27, 26] propose that parts may play a major role in the recognition of basic-level objects, which may have something to do with the so-called Gestalt peception [12] related to part-whole configuration. In their proposal there is a strong suggestion that our basic-level object perception may well be based around this part-whole division. Parts, in turn, are related to functions, shape and interactions of these basic-level objects. This gives rise to an interesting question: is the categorization around parts reflected in the language we use?

3 Idealized Cognitive Models

Lakoff [13] believes that linguistics categories have the same character as other conceptual categories: they show prototype effects and can be demonstrated to have basic-level categories. But he makes it clear that neither he nor Rosch advocate the view that basic-level categories would explain any structural or procedural properties of cognition. Rather, they both regard basic-level categories as a mere surface phenomena related to cognition, and assume that below that surface there may be some other more interesting structures and processes to be found.

Lakoff's main thesis is that our knowledge is organized by means of structures to which he refers to as idealized cognitive models, or ICMs, and that category structures and prototype effects are their by-products. Each ICM is seen as a structured whole, a gestalt, with four structuring principles employed:

– propositional structure (Fillmore's [7] frames)
– image-schematic structure (Langacker's [14] cognitive grammar)
– metaphoric mappings
– metonymic mappings

These ICMs would then structure the mental space as described by Fauconnier [6]. As examples of ICMs, among others, Lakoff refers to a Balinese calendar system with three different "*week*" structures superimposed [9], the category defined by the English word *bachelor* [7] and other examples.

The importance of Lakoff's ICMs to this research is in that he shows how, by extending the basic-level categories to the linguistic domain, we can end up with novel categorical structures, which may have not been considered at all in the creation of ontologies that are widely used today. This, in turn, may be one of the reasons why these conventional ontologies may prove inadequate for linguistics tasks such as word sense disambiguation.

4 Example Ontology — WordNet

WordNet 2 defines itself as “a machine-readable lexical database organized by meanings”. It organizes English nouns, verbs and adverbs into synonym sets representing lexical concepts [8]. The sets are linked by relations such as hypernym, meronym, synonym and antonym. WordNet has been critizised for not providing a useful organisational principle for information retrieval, reasoning, or knowledge management, being based on linguistic rather than encyclopaedic coherence [2]. Concepts likely to occur together in a domain are often found widely separated from each other in the conceptual hierarchy [24].

However, the linguistic principles employed in WordNet's construction have made it a useful tool for word sense disambiguation. WordNet has been used with many different WSD techniques, the resulting disambiguation accuracies ranging from 57% to 92% [4, 16, 19, 20]. To make it even more useful for WSD, some important cognitive principles might need to be explicitly added to its organization. These could be implemented through pointers as ontological relations.

In fact, the authors of WordNet had this in mind when starting to construct it. As an example, Miller pointed out that the word *canary* should be associated with at least three types of distinguishing features: (1) attributes (small, yellow and other adjectives), (2) parts (beak, wings and other nouns), and (3) functions (sing, fly, and other verbs). The addition of the distinguishing features important to basic-level categories was contemplated, but was not implemented explicitly except for the pointers to the parts [18]. Instead, glosses were added which contain some of these features. Many WSD implementations have used these glosses since for sense disambiguation.

In this research, feature sets incorporating these distinguishing features and also other associations and collocations are used. Most of these are not explicitly expressed in WordNet, and here we try roughly to gauge their relative importance to WSD.

5 Use of InfoMap as the First-Stage Disambiguator

To disambiguate with the help of context one needs a set of words that co-occur, more often than would be the case by chance, with the word to be disambiguated. One way to do this would be to collect co-occurrence statistics with whatever software were available for the purpose, but the drawback of this method is that the statistics do not discriminate between the senses. A better way is to use an application that is based on co-occurrences but which, nevertheless, can be made to discriminate, to some extent, between the senses when a judicious selection of the search terms is performed. One such application is InfoMap (http://infomap.stanford.edu), which is freely available from the Stanford University site and is explained in detail in [25, 29].

The principle behind InfoMap was developed by Schütze [23] and implemented and modified principally by the InfoMap team at the Stanford University. The distribution of word co-occurrences between a word and sets of content-bearing words creates a profile of the words usage in a context, and thus a profile of the word meaning itself. A similarity between two words can be calculated by comparing the profiles of the words in question. It is possible to return related documents whose profiles are close to each other even though they may not include the query words themselves. The meaning can be narrowed down by the selection of search words and can thus be used to disambiguate the key search term to some extent at least.

To get a set of word clusters related to the word palm in the sense of a hand (Table 1) we can simply enter the words *palm* and *hand* together as search terms, together with any negative keywords. The web interface allows us to retrieve up to 200

Table 1. Results of an InfoMap-query using *palm* and *hand* as keywords and *tree* as a negative keyword. 10 clusters were specified.

Prototypical Example	Cluster Members
hand	hand wrist elbow finger thumb forearm grasped glove holding squeezed grasping firmly coin ear torch cigarette lever grasp isambard pencil verbal button squeeze candle undone propped superiority mister tapped arm's
palm	palm tapping held knuckles squeezing aloft wrapped knotted cradle shield caf woven loom restraining cloth smacked clips begging salute raffle necklace delights twists cane embroidered
fingers	fingers cheek cupped clutched stroked grip touched stroking brushed gripped lightly kissed gently tenderly fingertips delicately hold flinched knife stretched touch rubbing rested touching blade lifted pins dagger limp slid knelt shake caress razor pressed gasping tip rope raised brush
shoulder	arm shoulder outstretched sleeve fist clasped clutching clamped thigh waved sword knee patted foot hip gripping rein gesture hips trouser knob leg reins swinging breast smoothing bend needle forward
pocket	pocket put picked wallet handed bag tore crumpled briefcase pen drawer pad cardboard paw pockets parchment suitcases handbag putting lend packet
left	left fork hemisphere edge side scars stile pictured
grabbed	grabbed gun pistol snatched fumbled wrenched grab
lips	tightly rubbed chin trembling clenched mouth handkerchief kiss boy's lips breath gasped brow twitched
jar	saucer basket teapot bottle plate crumbs biscuit jar tray
shoulders	forehead shaking bent resting waist arms jerked tugging tilting rolled curled chest palms slapped knees wrists shoulders

words associated with the key search terms and divide them into 1–20 categories as desired. A prototypical example is given for each category. Other search strategies could also be used for the same end including contrasting pairs.

6 Basic-Level Categories and ICM's in WSD

The idea behind using InfoMap is to get a set of terms associated with the word to be disambiguated and occurring together in the same context. InfoMap is based on co-occurrence information and word vector relations and, therefore, seems suitable for the purpose. The public web interface for the application at the Stanford University site was included within the Java-based disambiguating application created for the purpose. The mode of the operation was, shortly, as follows.

The parameters posted to the site were the search term palm + other keywords., (hand), negative keywords (tree), corpus (British National Corpus), command (associate), and parameters specifying clustered results with 200 words divided in 10 clusters. The request to the site was sent separately for both of the major senses of the word to be disambiguated and the results received were combined to form a Disambiguation Feature Cluster set consisting of 20 clusters, the first 10 for the first major sense of the word to be disambiguated (keywords: palm hand, neg. keyword: tree) the second 10 for the second major sense (keywords: palm tree, neg. keyword: hand). The returned information (Figure 1 showing half of the set) was then converted into an XML-format and indexed into a file database using Java Digester Libraries [5]. The context to be disambiguated was indexed to another data base using Digester and Porter Stemming Algorithm. In the process of disambiguation, the context sentences were iterated through and matched against the Disambiguation Feature Cluster set: each time a word in the context sentence matched the clusters 1–10 the first sense increased its score, and when 11–20 were matched the second sense increased its score. The maximum of these scores indicated the word sense. The matches were indicated either as correct, undecidable (no matches), even, or wrong. For the query matching, Java Lucene [10, 17] libraries were used.

First we tested our application with Mihalcea's sense tagged data for six words with two-way ambiguities, previously used in word sense disambiguation research and extracted from BNC [28]. We simply took her Meanings-labels as positive and negative keywords to create the feature-sets with the help of InfoMap and then used these feature-sets to disambiguate her examples. The results are shown in Table 2. As expected, the results were variable, ranging from 47.3 to 82.2 % in accuracy, indicating that the selection of the keywords is significant. Changing *tank*'s "vehicle"-keyword to "military," for example, increased the disambiguation accuracy to 64.7%. Increasing the number of the keywords also had a significant effect on the result.

Table 2. Disambiguation accuracies reached using the TWA dataset's Meanings-labels as keywords

Word	Meanings	Examples	Correct
bass	fish/music	107	82.2 %
crane	bird/machine	95	68.4 %
motion	movement/legal	201	49.8 %
palm	hand/tree	201	72.0 %
plant	living/factory	188	77.1 %
tank	container/vehicle	201	47.3 %

However, our purpose was not to find out the maximum disambiguating power of InfoMap , but to use it as a tool to help in our own experiments. We merely needed a rough set of context words to modify using basic-level category information to see how that information affected the disambiguation accuracy.

For our example, the word *palm* was selected, because it had an adequate number of hand-tagged contexts (201) and the disambiguation accuracy (75.1%) achieved was judged sufficient, but not too, high for our purpose. Moreover, we could extract an adequate number of contexts (1000) with the word *palm* from the BNC against which to test this set. As both sets come from the BNC, they may partially overlap, but, as said, the purpose of the experiment was to test the effect of the basic-level category words on the overall disambiguation against normal context words. More extensive n-way tests will follow based on this experiment. After pruning out some minor senses, 193 contexts remained. The remaining TWA contexts were processed using the un-

Table 3. Results for the two major senses (part-of-hand, tree) of the word *palm* in 193 contexts when disambiguated with the unmodified disambiguation feature cluster set (UDFC).

Wide context (paragraph)			Narrow context (sentence)		
correct:	139	72.0 %	correct:	130	67.4 %
undecidable:	0	0.0 %	undecidable:	0	0.0 %
equal:	24	12.4 %	equal:	22	11.4 %
wrong:	30	15.5 %	wrong:	41	21.2 %
Total:	193	~100.0 %	Total:	193	~100.0 %

Table 4. Results for the two major senses (part-of-hand, tree) of the word *palm* in 193 contexts when disambiguated with the MDFC set.

Wide context (paragraph)			Narrow context (sentence)		
correct:	193	100.0 %	correct:	193	100.0 %
undecidable:	0	0.0 %	undecidable:	0	0.0 %
equal:	0	0.0 %	equal:	0	0.0 %
wrong:	0	0.0 %	wrong:	0	0.0 %
Total:	193	~100.0 %	Total:	193	~100.0 %

modified InfoMap feature set for disambiguation. The results were as shown in Table 3.

Even when disambiguating with the unmodified InfoMap results, the disambiguation achieved is significantly better than what could be expected by chance. Our purpose was to modify the feature set to see what the actual words were that played role in disambiguation and what was their number, in order to be able to roughly categorize the words participating in disambiguation. For this reason the words that had not played any part in disambiguation, were pruned from the Disambiguation Feature Cluster Set. Some words that were judged as missing were added, and to get a 100% disambiguation result for the TWA contexts (Table 4) further 5 collocations ({"his","palm"}, {"read","palm"}, {"her","palm"}, {"my","palm"}, {"palm","tree"}) were added. The number of the words in the modified and unmodified set remained roughly the same. We call the original, unmodified set the Unmodified Disambiguation Feature Clusters (UDFC) set and the modified one the Modified Disambiguation Feature Clusters (MDFC) set. In the MDFC the feature categories were rearranged to create additionally a feature set for a) Parts, b) Objects Affected, and c) Functions in order to roughly isolate the features that might be related to basic-level information.

1000 contexts containing the word *palm* were then extracted from the BNC out of which 749 contained either of the major senses (part-of-hand, tree) and these were then selected for disambiguation. First these contexts were disambiguated with the help of the UDFC set (Table 5) and then with the help of the MDFC set (Table 6).

As these results show, the disambiguation accuracy for the MDFC was considerably higher than for the UDFC. The accuracy of UDFC increased when the number of contexts was increased, whereas the accuracy declined for MDFC. This probably was due to the fact that MDFC was optimized for TWA contexts whereas UDFC was not, i.e., some of the pruned words might have proved useful in new contexts, etc.

Table 5. Results for the two major senses (part-of-hand, tree) of the word *palm* in 749 contexts when disambiguated with the UDFC set.

Wide context (paragraph)			Narrow context (sentence)		
correct:	596	79.6 %	correct:	550	73.4 %
undecidable:	2	0.3 %	undecidable:	0	0.0 %
equal:	71	9.5 %	equal:	76	10.1 %
wrong:	80	10.7 %	wrong:	123	16.4 %
Total:	749	~100.0 %	Total:	749	~100.0 %

Table 6. Results for the two major senses (part-of-hand, tree) of the word *palm* in 749 contexts when disambiguated with MDFC set.

Wide context (paragraph)			Narrow context (sentence)		
correct:	702	93.7 %	correct:	692	92.4 %
undecidable:	11	1.5 %	undecidable:	32	4.3 %
equal:	13	1.7 %	equal:	12	1.6 %
wrong:	23	3.1 %	wrong:	13	1.7 %
Total:	749	~100.0 %	Total:	749	~100.0 %

Then a very rough estimation was made of the contribution that the feature-sets (Parts, Functions) linked to basic-level information (hypernyms, parts, functions) made towards the overall disambiguation. For this the 193 pruned contexts from TWA were used. First, the parts and functions clusters from the MDFC were removed and the remaining clusters only were used for disambiguation. As the word *palm*'s salience varied within the context, being sometimes in the foreground sometimes in the background, it was decided to conflate the part information between adjacent levels: all tree parts were considered together and all body parts were considered together. Similarly, all tree function words were considered together, and all body function words were considered together.

The disambiguation accuracy exceeded the 50/50 (Table 7) with significant results, but a lot of scope was left for improvement, which shows that the inclusion of parts and functions in the clusters used in MDFC is essential for accuracy. This is shown even clearer when we include only the parts and functions clusters and remove all others from MDFC (Table 8).

7 Discussion

Although, as the very first experiment with the unmodified set returned by InfoMap shows, the disambiguating word set needs to be modified for more accurate functioning, the size of the set (200 words for each sense) seems adequate. Our preliminary experiments with other disambiguous words have shown that an ontology relating the words through structures including the novel categories would ideally suit for word sense disambiguation. We have previously successfully used WordNet for disambiguating words based on an artificial taxonomy (animals) [15] and expect that by aug-

Table 7. Results for the two major senses (part-of-hand, tree) of the word *palm* in 749 contexts when disambiguated with the MDFC set. Two MDFC clusters, parts, and functions, are not used in this MDFC.

Parts and functions clusters not included					
Wide context			Narrow context		
correct:	121	62.7 %	correct:	102	52.8 %
undecidable:	49	25.4 %	undecidable:	84	43.5 %
equal:	13	6.7 %	equal:	1	0.5 %
wrong:	10	5.2 %	wrong:	6	3.1 %
Total:	193	~100.0 %	Total:	193	~100.0 %

Table 8. Results for the two major senses (part-of-hand, tree) of the word *palm* in 749 contexts when disambiguated with the MDFC. Only parts, and functions clusters are used in this MDFC.

With parts and functions clusters only					
Wide context			Narrow context		
correct:	149	77.2 %	correct:	144	74.6 %
undecidable:	2	15.0 %	undecidable:	40	20.7 %
equal:	9	4.7 %	equal:	6	3.1 %
wrong:	6	3.1 %	wrong:	3	1.6 %
Total:	193	~100.0 %	Total:	193	~100.0 %

menting the relations within WordNet to include categorical relations that appear to have some relation with basic-level categories and idealized cognitive models we could make it more suitable for disambiguation purposes. However, there are many questions to be solved about the basic-level categories, ICMs and their relations to context before a more comprehensive system can be developed. For example, something perceived as basic level varies amongst individuals: for an expert *eucalypt* may appear as a basic-level object, whereas for many ordinary city-dwellers it is *tree* that is seen as the basic-level object. The salience of the word within the context, i.e., whether it is in the background or in the foreground, affects the gestalt experienced also. There may be hundreds of different types of ICMs judging by the variety of examples given by Lakoff and others. Some of this information is already coded in different ontologies, albeit referred to by different terms, such as thematic relations, partonymy etc. It is likely, as Rosch and Lakoff have pointed out that basic-level structures are a mere surface phenomena, and one needs to dig deeper to get to the gist of what happens in the cognition when dealing with categories, in order to allow us to build structures that can be used in disambiguation

Complex as it might seem considering the reservations above, the research is justified on the grounds that a human being can disambiguate linguistic context better than a machine, and unless we are able to come up with a superior algorithm or mimic this disambiguating behavior, we can never be sure whether the results that our machine translation and other applications come up with are correct. We need to communicate globally and rapidly and need to be able to do it without the fear of being misunderstood.

References

1. Archambault, A., Gosselin, F., Schyns, P. (2000). A natural bias for the basic level? Proceedings of the 22nd Annual Conference of the CognitiveScience Society, NJ: LEA, 585-590.
2. Brewster, C. (2002). Techniques for Automated Taxonomy Building: Towards Ontologies for Knowledge Management. In Proc. CLUK Research Colloquium, Leeds, UK.
3. Brown, R. (1958). How Shall a Thing be Called? Psychological Review 65:14-21.
4. Cañas, A.J., Valerio, A. Lalinde-Pulido, J., Carvalho, M., Arguedas, M. (2003). Using WordNet for Word Sense Disambiguation to Support Concept Map Construction Proceedings of SPIRE 2003 – 10th International Symposium on String Processing and Information Retrieval, October 2003, Manaus, Brazil.
5. Digester can be downloaded from: http://jakarta.apache.org/commons/digester/.
6. Fauconnier, G. (1985). Mental Spaces. Cambridge, Mass. MIT Press.
7. Fillmore, Charles J. (1982). Frame semantics. In Linguistics in the Morning Calm, Seoul, Hanshin Publishing Co., 111-137.
8. Fellbaum, C. ed. (1998). WordNet: An Electronic lexical database, MIT Press.
9. Geertz, C. (1973). The Interpretation of Cultures. New York. Basic Books.
10. Gospodnetic, O., Hatcher, E., Lucene in Action, (2004), Manning Publications, Geenwich, CT.
11. InfoMap information mapping project web interface at http:// infomap.stanford.edu/.
12. Koffka, K. (1922). Perception: An introduction to the Gestalt-theorie. Psychological Bulleting, 19. pp. 531-585.
13. Lakoff, G. (1987). Women, Fire, and Dangerous Things. The University of Chicago Press. Chicago and London.
14. Langacker, R. (1986). Foundations of Cognitive Grammar, vol 1. Stanford Univ. Press.
15. Legrand, S., Pulido JGR. (2004). A Hybrid Approach to Word Sense Disambiguation: Neural Clustering with Class Labeling. Knowledge Discovery and Ontologies (KDO-2004) workshop in15th European Conference on Machine Learning (ECML) and 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD) Pisa, Italy, September 24, 2004
16. Li X., Szpakowics, S., Matwin, S. (1995). A WordNet-based Algorithm for Word Sense Disambiguation. Proceedings of IJCAI-95. Montréal, Canada, 1995.
17. Lucene Java Libraries can be downloaded from: lucene.apache.org/java/docs/ index.html.
18. Miller, G.A., Beckwith, R., Fellbaum, C.D, Gross, D., Miller, K. (1993). Five Papers on WordNet. Technical report, Princeton University, Princeton, N.J.
19. Mihalcea, R., Moldovan, D. (2000). An Iterative Approach to Word Sense Disambiguation, Proc. of Flairs 2000, pp. 219-223, Orlando, FL.
20. Nastase, V., Szpakowics, S. (2001). Word Sense Disambiguation in Roget's Thesaurus Using WordNet, Proc. of the NAACL WordNet and Other Lexical Resources Workshop. Pittsburgh, June 2001.
21. Rosch, E., Mervis, C.B..Gray, W.D., Johnson, D.M., and Boyes-Braem, P. (1976). Basic objects in natural categories. Cognitive Psychology, 23, 457-482.
22. Rosch, E., Principles of Categorization. (1988). In Readings in Cognitive Science, a Perspective from Psychology and Artificial Intelligence. Allan Collins & Edward E. Smith. Morgan Kaufmann Publishers. San Mateo, California. pp. 312-322.
23. Schütze, H. 1997. Ambiguity Resolution in Language Learning: Computational and Cognitive Models. CSLI Lecture Notes 71, CSLI Publications, based on revised PhD Thesis, Stanford University, Dept. of Linguistics, July 1995.
24. Stevenson, M. (2002). Combining Disambiguation Techniques to Enrich an Ontology, Proceedings of the Fifteenth European Conference on Artificial Intelligence (ECAI-02)

workshop on "Machine Learning and Natural Language Processing for Ontology Engineering", Lyon, France.
25. Takayma, Y., Flournoy, R.S., Kaufmann, S. (1998). Information Mapping. Centre for the Study of Language and Information. Stanford University. Available at: www-csli.stanford.edu/semlab/infomap/Papers/takayama-infomap.ps
26. Tversky, B. (1989). Parts, Partonomies, and Taxonomies. Developmental Psychology, Vol. 25. No.6, pp 983–995.
27. Tversky, B, Hemenway, K. (1984). Objects, parts, and categories. Journal of Experimental Psychology: General, 113. pp. 169–193.
28. TWA sense tagged context data in http://www.cs.unt.edu/~rada/downloads.html
29. Widdows, D. (2004). Geometry and Meaning. CSLI Lecture Notes 172, CSLI Publications, Stanford University.
30. Wittengstein, L. (1953). Philosophical Investigations. MacMillan, New York.

Article V

STRUCTURING METAPHORS WITH BASIC-LEVEL CONCEPTS FOR WORD SENSE DISAMBIGUATION

by

S. Legrand 2007

In J. Šķilters (Ed.): *Baltic International Yearbook of Cognition, Logic, and Communication.* Vol.2: Complex Cognition and Qualitative Science, 2007, pp. 171-188

STRUCTURING METAPHORS WITH BASIC-LEVEL CONCEPTS FOR WORD SENSE DISAMBIGUATION

Steve Legrand
University of Jyväskylä
Finland
stelegra@cc.jyu.fi

ABSTRACT
Metaphorical mappings between source and target domains seem to correspond to abstract concept formation in humans. Findings indicate that source categories in metaphors may be based at least partially on ontological relations. This paper presents a model of how abstract linguistic concepts might be formed with the help of metaphors. The model is based on basic-level categories by Rosch (1988), Contemporary Theory of Metaphor by Lakoff and Johnson (1980), and Conceptual Mapping Model by Ahrens (2002). Recent experimental research (Legrand, 2006) indicates that basic level categories can be helpful in word sense disambiguation. Metaphorical expressions are especially hard to disambiguate by computational means: thus structuring them with the help of basic level categories will make it possible to use various word disambiguation methods on them as well.

Keywords: metaphor, disambiguation, basic-level categories, ontology

1. Introduction

If someone claims: "*My car drinks petrol*", we are able, quite easily, to interpret the sentence. It would probably be harder to understand a sentence like: "*This car drinks gas left and right like it was nobody's business...*", although most of us would still be able to understand it. However, a software application trying to disambiguate these words and sentences in order to translate them into another language might not be able to accomplish its aim. Suppose that the application used selectional restrictions (Resnik, 1993) to help extract the meaning from the context. The verb *drink* commonly selects for a subject that belongs to the "animate" category and for an object that belongs to the "liquid" category. The second sentence, in which a machine is made to drink gas, might prove an unsolvable puzzle to the translation application and the translation to another language might turn out to be something like: "*This car imbibes flatulence left and right like a company that belongs to nobody...*" Such a translation might result, because the application cannot determine, based on the sentence alone, whether it is the object that selects the verb or the verb that selects the object.

The metaphorical linkage that we can easily understand and which modifies the selectional restrictions can be labeled as MACHINES ARE CREATURES. In a language other than English, this metaphorical correspondence might not even exist, and the sentence "*My car drinks petrol*", might seem extremely funny or nonsensical.

An individual attempting to describe artifacts or abstract things often finds it very hard, perhaps impossible, to talk or write about them without using metaphors. We do not experience abstract concepts the same way that we experience concrete entities. When someone says: "*There are holes in your argument*" it seems that he/she is constructing the abstract concept by using more familiar conceptual building blocks from the every-day experiential, human-sized world. Here, the *argument* concept is treated as a container that can have holes and become empty if there is a leakage. The source concept or domain in this case is *container* and the target concept or domain is *argument*.

The question naturally arises how to make a translation application (or any other artificial system dealing with natural language) aware of this human trait in such a way that the application could be made to take advantage of it. Whether the system mimicked the way that humans form abstract concepts or performed a functionally equivalent operation in its own way would be beside the point here, as long as it could disambiguate and make sense of the metaphorical expressions.

In this article, we propose a way in which the human cognitive system could be understood to build abstract concepts with the help of metaphors and basic-level schema. The model presented here does not pretend to be a faithful model of the entire human abstraction mechanism, though it does rely on some widely respected empirical results in its construction. The novelty of the approach lies in structuring metaphors with the help of basic level categories. It is hoped that the model will be able to help scientists dealing with natural language disambiguation, at least in cases where metaphors are used to construct abstract concepts. To enable a system to map source domain information to target domain and vice versa we propose including these relations directly in relevant domain ontologies extended with links to structured metaphors.

Before presenting the model, the paper briefly reviews the concepts of basic-level categories (Section 2), metaphorical expressions and embodiment (Section 3); it then presents some examples of work done for the discovery of metaphorical information from ontologies (Section 4.1) as well as ways in which such information can itself be included in ontologies (Section 4.2). After this, the details of our metaphor ontology are presented (Section 5), the applicability of the model for word sense disambiguation is considered (Section 5), and critical viewpoints related to the theory on which our proposal is based are discussed (Section 6).

2. Basic-level categories

Roger Brown (1958) explained his notion of a "first level" or "basic-level" category as a kind of category which allows children to learn object categories and name them. On this account "first level" categories fall somewhere between the most general and the most specific level. Later Rosch et al. (1976) and Rosch (1988) designed a series of experiments in which they demonstrated that these basic-level categories, as they were named, were somewhat inconsistent with the classical theory of categories Among other things, Rosch and her colleagues showed that the members of categories exhibit graded typicality effects and a continuum of representativeness. This runs counter to one of the main assumptions of the classical theory of categories, which states that no member can be more typical or more representative of a category than some other member and that categories are fixed and rigid.

From the point of view of human cognition, the categories seem to be divided roughly into three kinds: superordinate (*animal*), basic-level (*dog, fish, bird*), and subordinate (*alsatian, spaniel / mullet, perch / hawk, canary*). Categories at the basic-level are verified fastest, basic-level names are learned before subordinate names, objects falling within these categories can be named more quickly, and the names themselves tend to be shorter. The basic-level categories are maximally informative by sharing most of the features that are common to all subordinate members of the category and the least number of features with other contrasting categories. The features a basic-level object shares with the members of its own category are known as distinguishing features. There are at least three kinds of distinguishing features:

1. Parts: usually nouns. There are at least 6 different kinds of these *meronymic* relationships.
2. Functions: usually verbs that describe interactions of parts with their internal and external environment.
3. Attributes: usually adjectives that modify nouns.

For example, the basic-level features of a bird would include parts such as *beak* and *wings*, functions such as *fly* and *preen*, and attributes such as *light* and *fluffy*.

Tversky and Hemenway (1984) and Tversky (1989) propose that parts may play a major role in the recognition of basic-level objects, which may have something to do with the so-called Gestalt peception (Koffka, 1922) related to part-whole configuration. In their proposal there is a strong suggestion that our basic-level object perception may well be based around this part-whole division. Parts, in turn, are related to functions, which are based on the shape and interactions of the parts of these basic-level objects.

3. Metaphors and embodiment

To make other people understand us (aside, of course, from speaking the same language) we need to be able to describe our thoughts in terms that can be understood. In metaphorical expressions things in the target domain are described in terms of the source domain. This facilitates understanding because the source domain is usually well understood, whereas the target domain is normally less so (indeed, the target domain entity is often abstract).

According to Lakoff and Johnson (1980) we organize our experience into structured wholes with the help of experiential gestalts. A gestalt experience is a kind of multidimensional schema forming a coherent structural and/or functional whole. With the help of such schemata from various source domains we can explain and understand concepts in the target domain. Thus, in the sentence

> *At this point, we'll have constructed the core of our argument and can easily defend it.*

the *argument* concept (target) is structured with the help of four different metaphors:

AN ARGUMENT IS A JOURNEY	(*at this point*)
AN ARGUMENT IS A BUILDING	(*constructed*)
AN ARGUMENT IS A CONTAINER	(*core*)
AN ARGUMENT IS WAR	(*defend*)

Lakoff and Johnson (1999) maintain that most of our experience is based on bodily experience, that is, our understanding of the world is determined by our senses, by our ability to move about and to manipulate objects, by our culture and by any other things that have a bearing on our interactions with the environment. Not surprisingly then, the metaphors we use are also related to our bodily experience, and the source domain is very often related to our gestalt experience.

As noted in Section 2 on basic-level categories, parts and functions are closely related and this applies to parts and functions of the body as well. Varela et al (1991) give some good examples of how intertwined our bodies and minds are at the basic, interactionary level. Their "enaction" or embodied cognition concept suggests that the separation of the world from the individual is artificial. For example, instead of maintaining that our color perception reacts to outside visual stimuli, we should admit that the colors we see are not input from the outside world, but rather an enactment from our part: the color receptors in our visual system actively select certain wavelengths from the radiomagnetic spectrum, and the neural circuits in our brain interact with this selection. Not all the animals select the same wavelengths and the reason we do select these particular wavelengths rather than others is that it gives us an evolutionary advantage. Similarly, we do not simply pick pre-given categories from the world, rather we enact them and to some extent impose them on the world through the functioning of our sensorimotor system. The categories we enact most readily are the basic-level categories, which are fundamental to our interaction with the environment.

Some of these points can be illustrated further by the case of Wierzbicka (1980), who found herself in a dilemma when trying to give definitions to external body parts using only the semantic primitives in her system of 'lingua mentalis'. She found it quite impossible to define them without reference to their function, which led to circularity in definitions. At the end she had to resort to locational definitions, which naturally do not (in most cases) comprise the true essence of the things defined. Reading about biological organs in any text book should be sufficient to convince most of us that structure and function are in this context quite inseparable.

4. Metaphorical knowledge and ontologies

In this section we draw a distinction between research that attempts to discover metaphorical knowledge from ontologies without modifying them on the one hand, and research which aims to modify ontologies with metaphorical knowledge on the other. There is a grey area in between, and the separation is made mainly to emphasize the processing complexities that are faced if ontologies are left unmodified. There seems to be a rough give and take between changing the processing environment (changing the ontologies) or adapting to it in terms of increased processing effort. As an example of processing intensive applications which do not modify the underlying ontology we present CorMet (4.1), and as an example of research directed at modifying the underlying ontology we present some examples that use WordNet ontology (4.2).

4.1 Discovery of metaphorical knowledge from ontologies

Mason's (2004) CorMet system, which was built to discover metaphorical mappings between concepts, goes about it by clustering, with the help of various statistical measures, the context terms containing metaphorical expressions and then mapping the results obtained onto a knowledge base, Wordnet (Section 4.2), which is a kind of ontology. Domain-specific documents are obtained by searching the net with suitable keywords. Verb-stems that are the most characteristic for each domain are discovered by analyzing the documents on the basis of which case-frames for sentences containing these verbs with their accompanying nouns are extracted. These case frames are subjected to learning algorithms in order to establish their domain-specific selection preferences, and the selectional preference similarities between any two domains are then found by comparing the results of a clustering algorithm processing the WordNet ontology. Unlikely candidates are winnowed out with the help of so-called polarity. Finally the results are verified against the Master Metaphor List (Lakoff et al., 1991).

What the results show is only that it is possible to discover some metaphorical mappings; however, the CorMet system makes no attempt to interpret any of them. It would thus be hard to use the system for, for example, word sense disambiguation in its present stage. Overall, the process seems quite complicated given the results that it achieves, and the question arises of whether it would make more sense to code the metaphorical mappings into an ontology itself (4.2). It is fair to point out that metaphorical mappings between the word sense relations in WordNet are rather incidental, and in view of this the results obtained by CorMet were surprisingly good.

4.2 Addition of metaphorical knowledge into ontologies

None of the general ontologies such as CYC (Lenat, 1995) with its vast common sense database, or Mikrokosmos have made any special effort to link metaphorical knowledge to their ontologies, although the Mikrokosmos authors maintain that the structure of their ontology is conducive to such a linkage (Nirenburg and Raskin, 2004). Both CYC (Reed and Lenat, 2002) and another general ontology, SUMO (Niles and Pease, 2003), have recently been linked to WordNet (Miller et al., 1993), and it is WordNet and its multilingual extension EuroWordNet (Vossen, 1999) that are now the focus of interest for ontology modificaton with metaphorical knowledge.

WordNet, according to its website in (http://wordnet.princeton.edu/) is a lexical database or a lexical online reference system inspired by psycholinguistic theories of human lexical memory. For example, there is some evidence from studies conducted on aphasia patients that nouns form a separate lexical subsystem from adjectives or verbs, and this is reflected in the corresponding divisions of Wordnet. Some other evidence indicates that nouns are organized hierarchically. Many other psycholinguistic studies have been taken into consideration. However, any metaphorical knowledge that has found its way into WordNet has no obvious way of being used for systematic extraction and/or exploitation for NLP purposes. Also, the basic-level relations referred to in Section 2 have not been coded in the database, although this was originally advocated by Miller et al. (1993). Some of these relations can be derived from the glosses, though. WordNet is structured with the help of synsets (sets of synonymous words) with basic semantic relations between them.

EuroWordNet is a multilingual database which connects several European languages through its sharable top ontology, providing, with its 63 semantic distinctions, a common semantic framework for all the languages connected through it. The structure of each individual language-specific WordNet is based on the original Princeton WordNet described above. EuroWordNet links language-specific WordNets for various languages to a top ontology via a so-called InterLingual Index (ILI) and each WordNet, in turn, is internally linked via synsets.

Alonge and Castelli (2002) suggested a way of encoding metaphorical information in EuroWordNet to make the discovery of mappings between source and target domains easier for purposes such as word sense disambiguation. A composite ILI would mediate between the metaphorically related synsets and the top ontology domains (source and target) linking them. Apart from indicating the existing metaphorical mappings, the system could be used to infer potential metaphorical mappings. In this case the top-level ontology is internal to the EuroWordNet.

A system inspired by this suggestion, the Hamburg Metaphor Database (Eilts and Lönneker, 2002), has been partially implemented since. There is a web interface (http://www.rrz.uni-hamburg.de/metaphern/datenbank_en.html) which allows the user to explore metaphorical information in German and French based on the lexical information in EuroWordNet and metaphorical information in Master Metaphor List.

Ahrens et al. (2004) go a step further by using SUMO (Pease and Niles, 2003), a top-level ontology not specifically designed for WordNet, although linked to it (Niles and Pease, 2003) in their quest to verify mapping principles (Ahrens et al., 2003) between the source and target domains of conceptual metaphors. SUMO has an advantage over EuroWordNet top ontology in that it can be used to infer information through automatic reasoning. In this case, however, no automatic reasoning was implemented. A Mandarin corpus used to gather the metaphorical examples was linked to WordNet and the overlapping items in WordNet definitions and SUMO categories were used to verify the mapping to source and target domains. An interesting discovery that is relevant to this paper was that a source domain sometimes used inherited features from the domain subsuming it.

5. Metaphor ontology

This article proposes that rather than thinking in terms of a mere unspecified connection between the source and the target domains for metaphors, we should be more explicit about the types of relations that obtain between the domains and about reasons for their obtaining in order to make them more useful for word sense disambiguation and other purposes.

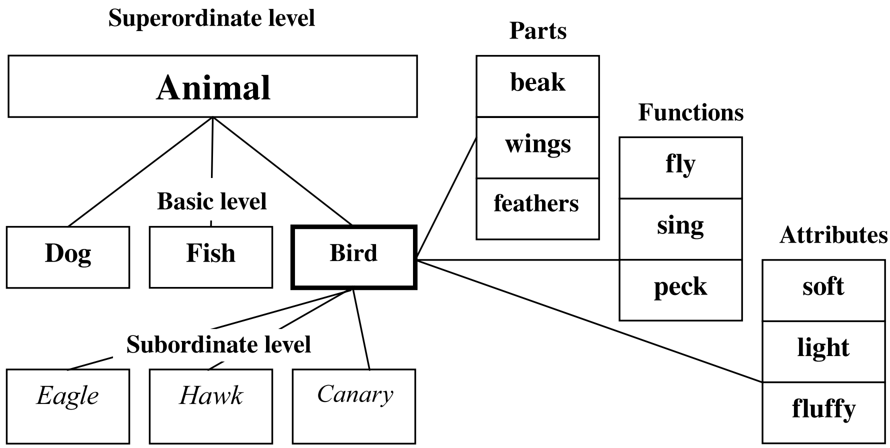


Figure 1: Some cognitive ontological relations of *bird*.

Figure 1 shows a conventional hierarchical ontology with its superordinate level, basic level and subordinate level. As discussed in the previous section, we usually find it easy to communicate about things on the basic level, often by using words referring to parts and functions. One should also remember that the basic level can vary according to the age and expertise of the observer. For example, the basic level depicted in Figure 1 might well be good enough for young children and city dwelling adults, whereas people who can tell birds apart might see concepts like *canary* as basic-level items.

It is more difficult to talk about abstract entities. These do not seem to have any definite basic level nor do they seem to have many representative instances, as is the case with concrete entities such as those described by the animal schema. One could think of emotions such as anger, love etc. as being basic level entities, but because they do not seem to have any particularly distinguishable parts or functions we need the help of metaphors when communicating about them.

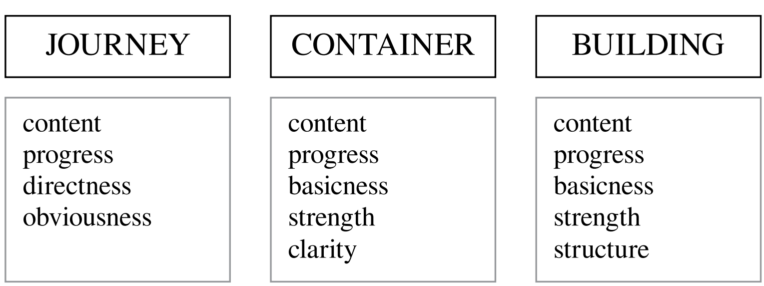


Figure 2: Aspects covered by some *argument* metaphors
(after Lakoff and Johnson, 1980)

For the argument concept Lakoff and Johnson (1980) listed some aspects which could be thought of as parts, but these are also abstract concepts and need some flesh around them in order to be understood. Figure 2 shows some of these aspects used in their metaphors. In the light of these metaphors it becomes possible to dispense with things such as obviousness, clarity, basicness etc., since we now have some less abstract or at least better-structured entities such as *building* or *container* to work with. In fact, we can now talk about arguments as we would about any other basic-level object.

Figure 3 shows how our model structures an abstract concept with a schema that is used for basic-level objects. Basic level objects, as we saw earlier, are objects that are the most informative and easiest to talk about, having parts and functions that we can clearly identify. Therefore, to be able to talk about an abstract object we need to add some 'handles', i.e., parts and functions. Comparing Figure 2 with Figure 3 we can see that their hierarchical levels coincide (Animal – Bird – Hawk v Conflict – Argument – Polemic), with the *argument* concept occupying the basic-level position, even though it might be somewhat misleading to think of abstract objects as basic-level objects. (Figures 2 and 3 have the same hierarchical structure, although the configurations look different due to issues of space).

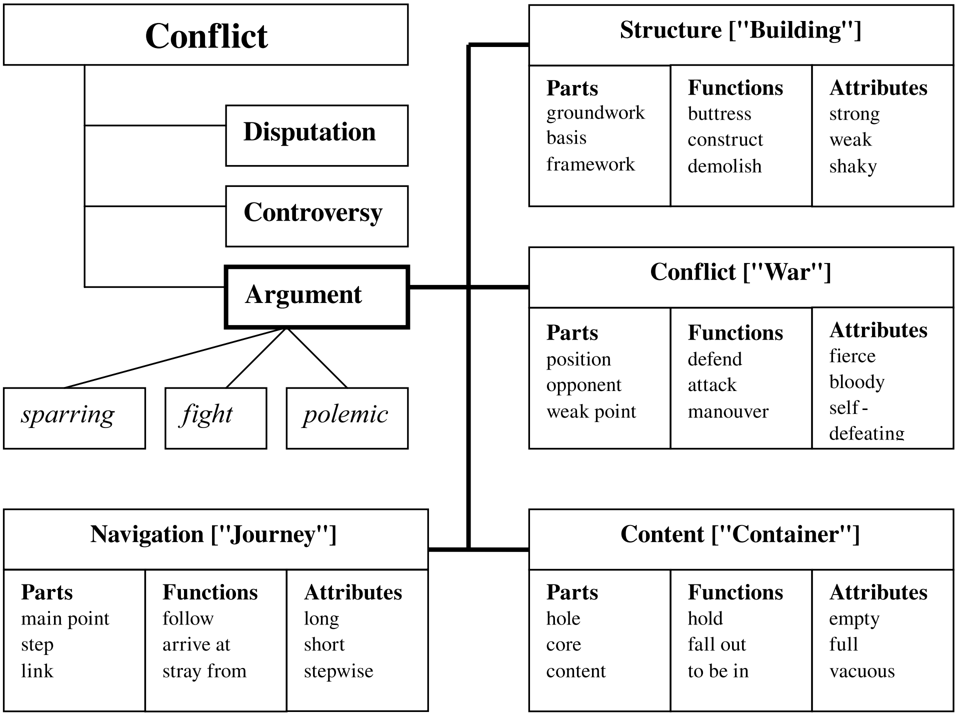


Figure 3: Schematic view of the *argument* concept

The schema used corresponds to both cases, but in order to be able to talk about *argument* we need some further help from other domains. While in Figure 2 *bird* simply has parts and functions, in Figure 3 parts and functions are subdivided into four separate categories according to the features characterizing the *argument* concept: *structure*, *content*, *conflict* and *navigation*.

This division structures the concept *argument* along the lines suggested by Lakoff, but rather than using concrete concepts to structure the main concept directly, it may be better to think of concrete concepts highlighting some aspects of abstract concepts that structure the concept *argument* itself. The Lakovian metaphorical statements such X IS Y are to be understood as labeling conventions only and should not be taken as adequate general descriptions of the underlying metaphors. Here we talk about schema correspondence where the basic level schema details of Figure 2 correspond with the metaphor ontology schema of Figure 3, but we do not claim that either one has primacy. Four (out of a possibly much larger number of) structuring sources are shown in Figure 3 and discussed in more detail below: these are structure, conflict, content and navigation.

1. Structure
   "With this framework, you can construct a strong argument."
   (ARGUMENT IS A BUILDING metaphor).
   While building has a physical structure, argument has an abstract structure (Ahrens et al., 2003). There are many different kinds of structures: a structure may be based on sequence, it may be about material composition etc. In our case, structure seems to concern the coherence and balance of an argument. One of the early experiences of a human being is interaction with gravity; trying to balance and keep oneself upright, trying to avoid falling and getting hurt. There is a clear link to the principle of conceptual embodiment here.
   Structure is used as a building block in other metaphors such as THEORIES ARE BUILDINGS, IDEA IS A BUILDING, and SOCIETY IS A BUILDING.

2. Conflict
   "We attack his positions while defending our own."
   (ARGUMENT IS WAR metaphor).
   Like Ritchie (2003), many others have pointed out that *conflict* might be better than *war* as a term to describe this important aspect of argument. This would also be more in line with Lakoff's own theories about embodiment as a basis of metaphors. While very few of us have direct experiences with war, practically all of us have, as children, engaged in war-like situations with other children, thus gaining first-hand experience with strategies,

opponents, defense etc. Later on this mindset probably gets transferred to play situations and games of different sorts and enforced by virtual exposure to war situations through media, especially TV in our times. Human beings are not the only ones whose young engage in mock fights or war games: all animal species exhibit similar tendencies, as the mastery of the concepts involved has survival value for the species. Rational argument might have developed later on, aiming at accommodation rather than confrontation, but there is a fictitious opponent there too. LOVE IS WAR, BUSINESS IS WAR and ECONOMY IS WAR are other well-known metaphors using war as a constituent.

3. Content
   "Your argument has very little substance."
   (ARGUMENT IS A CONTAINER).
   The participants in an argument need something to argue about, the subject matter, i.e., its content. Both content and container are equally relevant here. We can experience being inside or outside, perceive something as being inside or outside of our body, and we can see things such as cups and houses having both an inside and an outside. Out of this dichotomy there arises the boundary concept. The container schema has a close relation to centre-periphery schema and other schemas which can also be linked to embodiment (Shoul, 2003).
   IDEA IS A CONTAINER, TIME IS A CONTAINER and LIFE IS A CONTAINER are just few of the container metaphors used.

4. Navigation
   "We are well on our way to solving this problem."
   (ARGUMENT IS A JOURNEY metaphor).
   Old and Priss (2001) give some good reasons for using the navigation metaphor, although they do not explicitly recommend it for replacing journey in the argument metaphor. The reasons, incidentally, are related to Lakoff's embodiment principle. The authors point out that navigation is essential to animals and humans: we navigate toward a target in the distance taking the most direct and economic route and avoiding obstacles. We learn to do this early on during our childhood development. As in the case of conflict, it seems more likely that the concept of navigation or orientation in space and/or progress or some such concept is formed in our mind even before we learn to speak, and would, thus, be more fundamental than journey as a building block to other abstract concepts.
   The navigation metaphor would then include LIFE IS A JOURNEY, TIME IS A JOURNEY and LOVE IS A JOURNEY among others.

This ontological model resembles Ahrens' (2002) Conceptual Mapping Model and has been influenced by it. Our model makes the connection between the basic-level categories and the formation of abstract concepts more explicit. Mappable correspondences that are used in the CM Model consist of a) entities b) qualities and c) subject/object functions in the source domain. What Ahrens refers to as "entities" are treated as parts in Figure 3 in our model, and both the subject and object functions are referred to as functions. Her qualities correspond to our attributes. We do not give any ontological precedence to the basic-level schema over the schema used for the model. What we are trying to show is that there are notable correspondences between the basic-level schema and the schema used here. It should be added that it is likely that apart from the four structuring sources analyzed here there could be and probably are many more.

6. Metaphor ontologies in word sense disambiguation

Rosch et al (1988) and Lakoff (1987) have pointed out that basic-level structures are a mere surface phenomena. One needs to dig deeper to get to the gist of what happens in the cognition when dealing with categories, in order to allow us to build structures that can be used in word sense disambiguation. Metaphorical expressions are very hard to disambiguate by simply matching context words with conventional domain ontologies; it is more straightforward when the expression is not metaphoric. A simple example will illuminate this:

Non-metaphorical expression:

*A. The **building** was **constructed** by **carpenters**, **bricklayers** and **labourers**. The **materials** used consisted of **concrete**, **timber** and some **plastic tubing**.*

Metaphorical expression

*B. The **argument** was **constructed weakly**. It was **cemented** together in a **shoddy** way, and its **base** was not **strong** enough to **support** the rest of the argument's **framework**.*

In both cases, most of the context words seem to be connected to building construction domain ontology. We can easily see that the B paragraph is metaphorical and thus would, instead, opt for an ontology dealing with arguments or the like. A computer program, however, using simple word-counts, would most probably select the building construction ontology to match against, which in this case would be wrong. To get over this problem, it is necessary to connect the building construction ontology, through basic level link words, to the argument ontology (Figure 3 – note that only a very small portion of these link words are shown in the figure).

There are clear indications that basic level categories can be very helpful in word sense disambiguation (Legrand, 2006). The research method used by Legrand could be modified to apply it to metaphorical expressions also, which is one of the most difficult areas for disambiguation. The ontology to be used, be it WordNet or any other general or domain-specific ontology, would need to be extended with basic level links based on structuring of metaphorical expressions. For example, words such as "*construct*", "*structure*" and "*support*" that are commonly used in metaphoric expressions treating *argument* as a construction could act as linking concepts to the target domain. Extended metaphors might then be given a license to make use of the other target domain concepts as well.

7. Discussion

Our model suggests how abstract human concept formation could be rooted to embodiment as proposed by Lakoff and Johnson (1999) via a schema resembling the basic-level schema. Instead of using words like *building* and *war* to label the metaphoric correspondences, we use words such as *structure* and *conflict* which, as shown, are more fundamental to human cognition. The additional example metaphors at the end of each structural part description indicate that these basic building blocks can be applied to many other abstract concepts in metaphorical expressions.

Lakoff and Johnson's (1980) Contemporary Theory of Metaphor has its critics, and attempts to modify or refute it have been made. Our model uses its insights, but is not meant to explain or support any theories about how abstract concepts are formed in reality, rather, we find the correspondences in the mapping theory and basic-level category schema potentially useful for our purpose, which is to facilitate word sense disambiguation in different kinds of software applications. If the model happens to have any bearing on abstract concept formation in metaphorical expressions, that's all for the good, but to assess the model's usefulness in this respect more empirical research is needed.

Some experimental evidence related to metaphorical concept formation that is not entirely favorable to Lakoff's theories is gradually appearing. Keysar et al. (2000), for example, have run a series of experiments in which they claim to show that there is no significant difference in the subjects' mean reading times of paragraphs of conventional metaphoric expressions when compared with paragraphs with no mapping, implicit or explicit, to any source domain that could give rise to a metaphoric expression. What is interesting, and might seem even counterintuitive, is that the reading response times to novel (as opposed to conventional) metaphoric expressions were much shorter. Quite recently, Lakoff has embarked on an attempt to at least partially explain this and other anomalies with the help of neural theory (Lakoff, 2006).

Also, when looking at certain examples of data presented in the article, it seems that some explanations for the reading times for conventional expressions might have been overlooked. For example, in one explicit mapping scenario the paragraph

> *As a scientist, Tina thinks of her theories as her children. She is a prolific researcher, conceiving an enormous number of new findings each year. Tina is currently weaning her latest child.*

may strike one as ambiguous, explaining the relatively lengthy processing time when compared with an entirely novel metaphor. Although the last sentence in the paragraph is meant to belong to the metaphor, it could be interpreted as literal as well. In the novel metaphorical expression that the authors use in their experiment the expression is much more forceful and striking and, it seems, the whole paragraph is therefore in these cases less ambiguous. To be fair, this is just an isolated example, and does not seem to indicate anything systematic, especially on the basis of only the few examples available.

Supposing conventional metaphoric expressions had the same status in mental processing as other conventional expressions, how could that be related to the model we have proposed here? Wouldn't it be likely that the underlying processing structures would resemble each other, as they do in our model? Keysar et al. draw the conclusion that conventional metaphoric expressions do not instantiate metaphoric mappings and that only the novel ones do, but our model suggests that one should also be prepared to entertain the possibility that all conventional expressions including metaphorical conventional expressions create mappings to a basic-level schema type ontology, partly explaining the extra processing time when compared to novel metaphorical expressions. This, however, needs more empirical research.

It must be pointed out that the simple three-way categorization of parts-functions-attributes of our model, when related to real world knowledge, is likely to be too rough, and probably warrants further subdivisions. Only a fraction of each source domain's structuring resources are allowed to be mapped to the target domain, and in this respect the mapping is quite selective. However, this leaves room for the possibility of novel metaphors (once they have become conventionalized), which percolate across language boundaries from time to time. One such novel metaphorical expression is the infamous "mother-of-all-wars" and its entailments suggesting that wars are given a birth etc. Even though the original metaphor in this case may have been due to mistranslation (http://lists.village.virginia.edu/lists_archive/Humanist/v04/1131.html), it is now widely accepted and used in the English-speaking world, and can even be extended to cover other abstract concepts through expressions such as “mother-of-all-arguments”. This vividly demonstrates the adaptation and change that source categories for metaphors can undergo.

Also, it is not always obvious how certain concepts should be categorized: we might think of concepts such as heavy, blue etc. as qualities, but on the other hand we could easily decide to lump them with function words and call them, perhaps, interactionary features. Is “empty” a quality or does the word merely provide information about the shape of the object it is used to describe, or perhaps about the parts of the object? These are issues that need some further study.

Overall then, our future investigation should concentrate on details of the building blocks of the model presented here. Can they be further decomposed, and with a more extensive decomposition will we end up with a circular model? Lakoff (in a personal communication cited by Murphy (1996)) and other researchers have suggested that the abstract concept might have a slotted structure accepting fillers of a certain kind for concept building from the source domains. When combined with the assumption that these fillers from the source domains might also be abstract and their structure might be metaphorically structured, this would lead to circularity, and therefore needs careful investigation. It seems that the relation between the source domain and the target domain is not one-to-one: usually, certain parts of the target domain are structured by only some part of the source domain or by parts of multiple source domains.

A hypothesis that could be entertained on the basis of the model here is that the metaphorical terms used to structure a source domain could be transformed to a more general level of ‘embodiment primitives’ (such as conflict, content etc. and their bodily instantiations) and would ultimately define all the metaphoric expressions. This would also keep the mental processing times reasonable. To test our hypothesis we would need better defined and/or more fine-graded ontological distinctions in our basic-level schema-based ontological model. Also, more empirical work in this area is needed.

Another difficulty, which the model here does address is that pointed out by Ortony (1988): it is more likely that a child has an idea about an emotional concept such as anger or conflict before the notion of war emerges as the child learns about the world. The model shows how the ARGUMENT IS WAR metaphor is effectively generalized as *conflict*. Ahrens et al. (2003) also prefer to generalize it as *contest*. The important thing to note here is that this would basically mean that instead of talking only about conceptual structuring by metaphors, we would come up with two levels of structuring: one would be the level in which concepts would be conceptually structured by our sensorimotor system, and the other where the linguistic structuring with the help of metaphors would take place to better enable us to describe concepts using our everyday language.

References

Ahrens, K., 2002. When Love is not Digested: Underlying Reasons for Source to Target Domain Pairing in the Contemporary Theory of Metaphor. In YuChau E. Hsiao (ed.) *Proceedings of the First Cognitive Linguistics Conference*, Taipei: Cheng-Chi University, pp. 273-302.

Ahrens, K., Chung S-F., Huang C-R., 2003. Conceptual Metaphors: Ontology-based Representation and Corpora Driven Mapping Principles. *Proceedings of the ACL Workshop on the Lexicon and Figurative Language*. Kyoto. pp. 35-41.

Ahrens, K., Chung S-F., Huang C-R., 2004. From Lexical Semantics to Conceptual Metaphors: Mapping Principle Verification with WordNet and SUMO. In: Ji, Lua and Wang (eds). *Recent Advancement in Chinese Lexical Semantics: Proceedings of 5th Chinese Lexical Semantics Workshop (CLSW-5).* Singapore: COLIPS. pp. 99-106.

Alonge, A., Castelli, M., 2002. Which way should we go? Metaphoric expressions in lexical resources." In: *Proceedings of the third Language Resources and Evaluation Conference. Las Palmas*, Gran Canaria. Paris: European Language Resources Association. VI: 1948-1952.

Brown, R., 1958. How Shall a Thing be Called? *Psychological Review 65*:14-21

Eilts, C., Lönneker, B., 2002. The Hamburg Metaphor Database, *metaphoric.de, 03/2002.,* Online journal. Available in: http://www.metaphorik.de/Journal/ [5th March 2007].

Keysar, B., Shen, Y., Glucksberg, S., Horton, W.S., 2000. Conventional Language: How Metaphorical Is It? *Journal of Memory and Language 43:* pp. 576-593.

Koffka, K., 1922. Perception: An introduction to the Gestalt-theorie. *Psychological Bulletin, 19.* pp. 531-585.

Lakoff, G., 1987. Women, Fire, and Dangerous Things. The University of Chicago Press. Chicago and London.

Lakoff, G., 2002. Moral Politics: How Liberals and Conservatives Think. 2nd edition. Chicago and London: The University of Chicago Press.

Lakoff, G., 2006. The Neural Theory of Metaphor. 2006. Available in http://hci.ucsd.edu/coulson/cogling/lakoff10.pdf [5th March 2007]

Lakoff, G., Espenson, J., Schwartz, A.1991. Master metaphor list. Second draft copy. Cognitive Linguistics Group. University of California Berkeley. Available in http://araw.mede.uic.edu/~alansz/metaphor/METAPHORLIST.pdf [5th March 2007]

Lakoff, G., Johnson, M., 1980. Metaphors We Live By. Chicago and London: The University of Chicago Press.

Lakoff, G., Johnson, M., 1999. Philosophy in the Flesh. New York. Basic Books.

Legrand, S., 2006. Word Sense Disambiguation with Basic-Level Categories. In Advances in Natural Language Processing. Ed. Alexander Gelbukh, Research in Computing Science. Vol.18, IPN, Mexico, pp. 71-82

Lenat, D. B., 1995. Cyc: A Large-Scale Investment in Knowledge Infrastructure. The Communications of the ACM 38(11):33-38

Mason, Z.J., 2004. CorMet: a computational, corpus-based conventional metaphor extraction system, *Computational Linguistics, v.30* n.1, March 2004, pp.23-44

Miller, G.A, Beckwith, R., Fellbaum, C.D., Gross, D., Miller, K., 1993. Five Papers on WordNet. Technical report, Princeton University, Princeton, N.J.

Murphy, G.L., 1996. On Metaphoric Representation. *Cognition, 60,* pp.173-204.

Niles, I., Pease, A. 2003. Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In Proceedings of the 2003 International Conference on Information and Knowledge Engineering (IKE '03), Las Vegas, Nevada, June 23-26, 2003.

Nirenburg, S., Raskin,V., 2004. Ontological Semantics. MIT Press. Cambridge, Mass; London.

Old, L.J., Priss, U., 2001. Metaphor and Information Flow. In: *Proceedings of the 12th Midwest Artificial Intelligence and Cognitive Science Conference*, pp. 99-104

Ortony, A., 1988. Are Emotion Metaphors Conceptual or Lexical? *Cognition and Emotion, 2,* pp. 95-103.

Pease, A., Niles, I., Li, J., 2002. The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. In: *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web*, Edmonton, Canada, July 28-August 1, 2002.

Reed, S., Lenat, D.B., 2002. Mapping Ontologies into Cyc. In AAAI 2002 Conference Workshop on Ontologies For The Semantic Web, Edmonton, Canada, July 2002

Resnik, P.S., 1993. Selection and Information: A Class-Based Approach to lexical relationships. Ph.D. thesis, Computer and Information Science Department, University of Pennsylvania.

Ritchie, D., 2003, ARGUMENT IS WAR – Or is it a Game of Chess? Multiple Meanings in the Analysis of Implicit Metaphors. *Metaphor and Symbol, 18(2),* pp. 125-146.

Rosch, E., Principles of Categorization. 1988. In: Readings in Cognitive Science, a Perspective from Psychology and Artificial Intelligence. Allan Collins & Edward E. Smith. Morgan Kaufmann Publishers. San Mateo, California. pp. 312-322.

Rosch, E., Mervis, C.B..Gray, W.D., Johnson, D.M., Boyes-Braem, P., 1976. Basic objects in natural categories. *Cognitive Psychology, 23*, 457-482.

Shoul, M., 2003. Empathy and the innate response to architectural forms and spatial arrangements. Michael Shoul Memorial Paper. *4th International Space Syntax Symposium,* London, 7-19 June, 2003.

Tversky, B., 1989. Parts, Partonomies, and Taxonomies. *Developmental Psychology, Vol. 25*. No.6, pp 983-995.

Tversky, B, Hemenway, K., 1984. Objects, parts, and categories. *Journal of Experimental Psychology: General, 113*. pp. 169-193.

Varela, F.J., Thompson, E., Rosch, E., 1991. The Embodied Mind. MIT Press. Cambridge, Mass; London.

Vossen, P. (eds), 1999. EuroWordNet: A Multilingual Database with Lexical Semantic Networks, Kluwer Academic Publishers, Dordrecht

Wierzbicka, A.,1980. Lingua Mentalis. The Semantics of Natural Language. Academic Press. Australia.

Article VI

IDENTIFYING ONTOLOGY COMPONENTS FROM DIGITAL ARCHIVES FOR THE SEMANTIC WEB

by

J.R.G. Pulido, R. Herrera, M. Aréchiga, A. Block, R. Acosta, S. Legrand 2006

In S. Sahni (Ed.): *The IASTED Conference on Advances in Computer Science and Technology. Puerto Vallarta, Mexico, January 23-25, 2006,* pp. 7-12

Proceedings of the IASTED International Conference
ADVANCES IN COMPUTER SCIENCE AND TECHNOLOGY
January 23-25, 2006, Puerto Vallarta, Mexico

IDENTIFYING ONTOLOGY COMPONENTS FROM DIGITAL ARCHIVES FOR THE SEMANTIC WEB

JRG Pulido* R Herrera† M Aréchiga‡ A Block§ R Acosta¶ S Legrand‖

ABSTRACT

This paper describes an approach for identifying Ontology components by using Self-Organizing Maps (SOM). Our system represents the knowledge contained in a particular domain, any kind of digital archive, by assembling and displaying its ontology components. This novel approach provides a solution to the problem of semi-automatic ontology construction, supports mechanisms that explore domains, and allows knowledge components to be displayed in a browsable manner. Further processing may be carried out on the extracted knowledge to be embedded on the semantic web for software agents to use.

KEY WORDS

Semantic web, ontology learning, self-organizing maps.

1 Introduction

It is known that the web contains several billion of static pages connected by hyperlinks [26, 29]. Reaching them is a gigantic challenge having into account that current search engines only contain a small percentage of the total of documents in the web. Furthermore, this small amount of reachable documents is in an unstructured way, meaning that software agents understand actually nothing about the actual content of them. In other words, these documents can be read but not undestood [3]. It would be useful to develop representations of the information contained in digital archives and create intelligent systems supporting interactive searching. In this paper we describe an approach for helping in the semi-automatic construction of ontologies for such web sites. The remainder of this paper is organized as follows. In section 2 some related work is introduced. Our approach is outlined in section 3. Results are presented in section 4, and conclusions and further work in section 5.

*Faculty of Telematics, University of Colima, México, jrgp@ucol.mx
†SIABUC Dept, University of Colima, México, rherrera@ucol.mx
‡Faculty of Telematics, University of Colima, México, mandrad@ucol.mx
§Faculty of Telematics, University of Colima, México, arted@ucol.mx
¶Faculty of Telematics, University of Colima, México, acosta@ucol.mx
‖Computer Science School, University of Jyväskylä, Finland, steveleg@gmail.com

2 Related Work

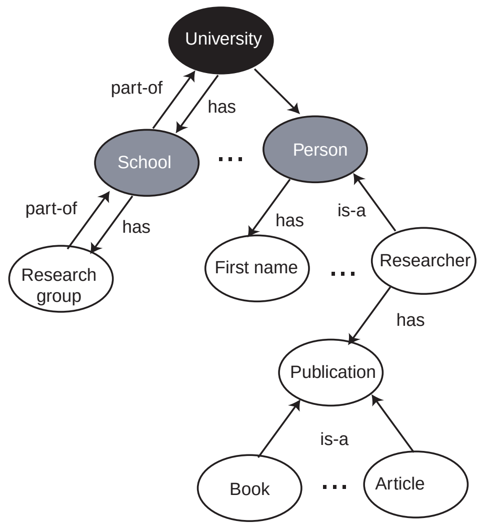
One of the most important challenges that the *semantic web* poses in dealing with large amounts of on-line knowledge is the mapping of unstructured information, suitable for humans, to formal representation of knowledge [5]. In the next subsections we have a brief look at some work done on Ontologies as well as Semantic Maps.

2.1 Constructing Ontologies

A representation that brings order and structure to a web site can be referred to as an *Ontology*. Representing knowledge about a domain as an ontology is a challenging process which is difficult to achieve in a consistent and rigorous way. It is easy to lose consistency and to introduce ambiguity and confusion [4]. An important observation in this context is that there is a significant manual effort involved in translating ontologies [27]. Nevertheless, ontologies are a useful form of knowledge representation which may be used to support the design and development of intelligent software applications and expert systems. Web ontologies can take rather different forms to traditional ones. New approaches, including advanced ontology languages have been proposed, such as OIL, DAML, OWL [2, 15, 10, 14, 8]. In [13] the use of the so-called *Simple HTML Ontology Extension* (SHOE) in a real world internet application is described. This approach allows authors to add semantic content to web pages, relating the context to common ontologies that provide contextual information about the domain. A similar approach is presented in [1]. Most *tag-annotated* web pages tend to categorize concepts, therefore there is no need for complex inference rules to perform automatic classification. One of the most exciting uses of an ontology, in the context of the semantic web, is to support the development of agent-based systems for web searching [9, 21].

2.2 Semantic Map Systems

An interesting project is presented in [18], where the results of applying the *WEBSOM2*, a document organization, searching and browsing system, to a set of about 7 million electronic patent abstracts is described. In this case, a document map is presented as




```
<html><head><a ONTO=''page:Researcher''/>
<title>Richard Benjamins</title>
</head><body><h1>
<a ONTO=''page[firstName=body]''>Richard</a>
<a ONTO=''page[lastName=body]''>Benjamins</a>
<a ONTO=''page[affiliation=body]''>
Artificial Intelligence Research Institute (IIIA)</a>
<a href=''http://www.iiia.csic.es''>CSIC</a>,
Barcelona, Spain.
</h1>
.
.
.
</body></html>
```

Figure 1. **Left** Basic taxonomy for an academic domain. **Right** Embedded ontology into a web page.

a series of HTML pages facilitating exploration. In [25] a distributed architecture for the extraction of meta-data from WWW documents is proposed and is particularly suited for repositories of historical publications. This information extraction system is based on semi-structured data analysis. The system output is a *meta-data object* containing a concise representation of the corresponding publication and its components. In that research gatherers have been designed as a combination of a parser, based on a context-free grammar, and a web robot, which navigates the links contained in the basic document type to infer the document structure of the entire site. These meta-data objects can be interchanged with other web agents, then classified and organized.

3 Methods

Our software is written in Java, which offers robust, multiplatform, and easy networking functionalities. Being a object-oriented programming language, it also facilitates reuse as well. Speed is not an issue anymore as computer processors are faster and faster. Java and its various APIs are powerful enough for constructing ontology software systems. The idea of combining ontologies and semantic maps has motivated our work. For the semantic web to become a reality, we need to transform the current web into a web where software agents are able to negotiate and carry out trivial tasks for us. Doing this manually, would mean a bottleneck for the semantic web. We need software tools that help us accomplish this enterprise.

Our system consists of two applications: Spade and Grubber [7, 6]. The former pre-processes html pages and creates a document space. The latter is fed with the document space and produces knowledge maps that allow us visualize ontology components contained from a digital archive. They may later be organized as a set of *Entities*[1], *Relations*[2], and *Functions*[3]. Problem solvers use this triad for inferring new data from knowledge bases [11, 12, 28, 22].

3.1 The Algorithm

SOM can be viewed as a model of unsupervised learning and an adaptive knowledge representation scheme. Adaptive means that at each iteration a unique sample is taken into account to update the weight vector of a neighbourhood of neurons [17]. Adaptation of the model vectors take place according to the following equation:

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)] \qquad (1)$$

where t $\in \mathcal{N}$ is the discrete time coordinate, $m_i \in \Re^n$ is a node, and $h_{ci}(t)$ is a neighbourhood function. The latter has a central role as it acts as a smoothing kernel defined over the lattice points and defines the stiffness of the surface to be fitted to the data points. This function may be constant for all the cells in the neighbourhood and zero elsewhere. A common neighbourhood kernel that describes a natural mapping and that

[1] Anything about which something can be said.

[2] Interconnections between entities in a universe of discourse (eg part-of).

[3] A special type of interrelation (eg is-a).

is used for this purpose can be written in terms of the Gaussian function:

$$h_{ci}(t) = \alpha(t)\exp(-\frac{||r_c - r_i||^2}{2\sigma^2(t)}) \qquad (2)$$

where $r_c, r_i \in \Re^2$ are the locations of the winner and a neighbouring node on the grid, $\alpha(t)$ is the learning rate ($0 \leq \alpha(t) \leq 1$), and $\sigma(t)$ is the width of the kernel. Both $\alpha(t)$ and $\sigma(t)$ decrease monotonically.

The major steps of our approach are as follows:

a) Produce a *document space* A document space is created with the individual vector spaces.

b) Construct the SOM By using a suitable number of cells and iterations the map [24] is trained with the *docuspace*.

Once the SOM is done, ontology components can be visualized clustered together. One important difference between our approach and Kohonen's is that we do not use *average context* [24, 16] to create the *docuspace*. This helps us reduce the dimensionality of the dataset. Contextual information is clustered together anyway. Preliminary results were surprisingly close to our intuitive expectations. After this, some other ontology tools such as editors can be used to organize this knowledge. Finally, it can be embedded into the digital archive (Fig.1) where it was extracted from by means of any of the ontology languages that exist.

4 A domain two datasets: The animal kingdom case

This section presents two experiments that we have have carried out. First we present, though in a different and enhanced way, some results that have been published before by other authors. Then in subsection 4.2, the results from applying our system to a bigger domain, from the same kingdom, are shown. Both subsections describe two maps of the domain in two ways, the front view showing *Attributes*, and the transposed view showing *Entities*. Both views display knowledge componentes of the domain clustered together.

4.1 The animal dataset

In [24, 23] the *animal dataset* is presented by means of a html page. Our approach uses a 4x4 SOM and presents the same data by using colored areas. In our experiment we found that one dominant characteristic amongst the animals is their size, e.g. birds are small, mammals come in two sizes. On the other hand, birds of prey and hunting mammals, small animals with feathers, big animals with hooves, and the ones with four legs and hair are also clustered together.

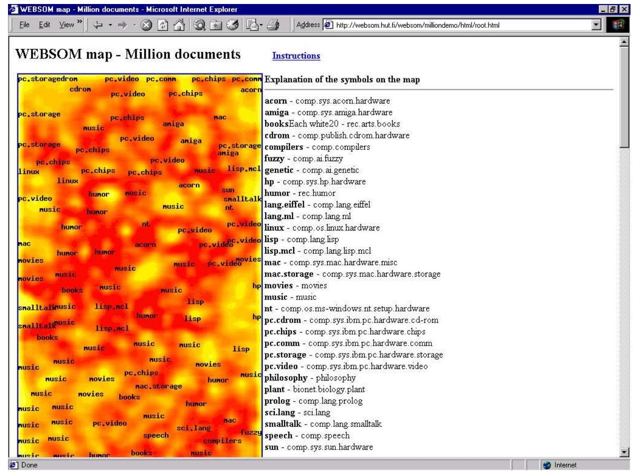


Figure 2. Websom: an online semantic map.

This is consistent with earlier tests carried out on the dataset. Both SOMs are shown in figure 3. It must be noticed that the vector spaces for *zebra and horse, and owl and hawk* are equal. The ones for *hen and duck* are approximately equal. Similarly, the vector spaces for the Attributes *feather and two legs, and hair and four legs* are equal. That is why some areas overlap and produce a combination of colorings.

4.2 The zoo dataset

For our second analysis another animal dataset has been used. This is commonly known as the *zoo dataset*. It contains 101 instances belonging to a seven already identified classes and 16 attributes. It combines 15 boolean attributes and a numerical one. The original dataset also includes *name* and *class*. We have used the former as one of the header files and the latter has been omitted here as our approach is going to find the classes. Two 10x10 SOMs have been used here for the analysis. It took just about 30 seconds training each map, the front and transposed view, on a IBM netvista computer (1.7GHz, 256Mb RAM). Browsing the SOMs gives us a clear idea and helps us understand what the domain is all about.

For instance we can readily identify *birds*[4], *fish*[5], *insects*[6], and *mammals*[7] within the domain. *Amphibians* and *reptiles* have not been easy to find as they overlap other classes. Attributes like *toothed, backbone, tail* are shared by *haddock* and *pitviper* for instance. Other attributes that can be seen clustered together by our software tool and that, of course, are not shared by the mentioned instances are *eggs, preda-*

[4] eg lark, parakeet, pheasant.
[5] eg sole, chub, carp.
[6] eg flea, gnat, housefly.
[7] eg elephant, antelope, hare.

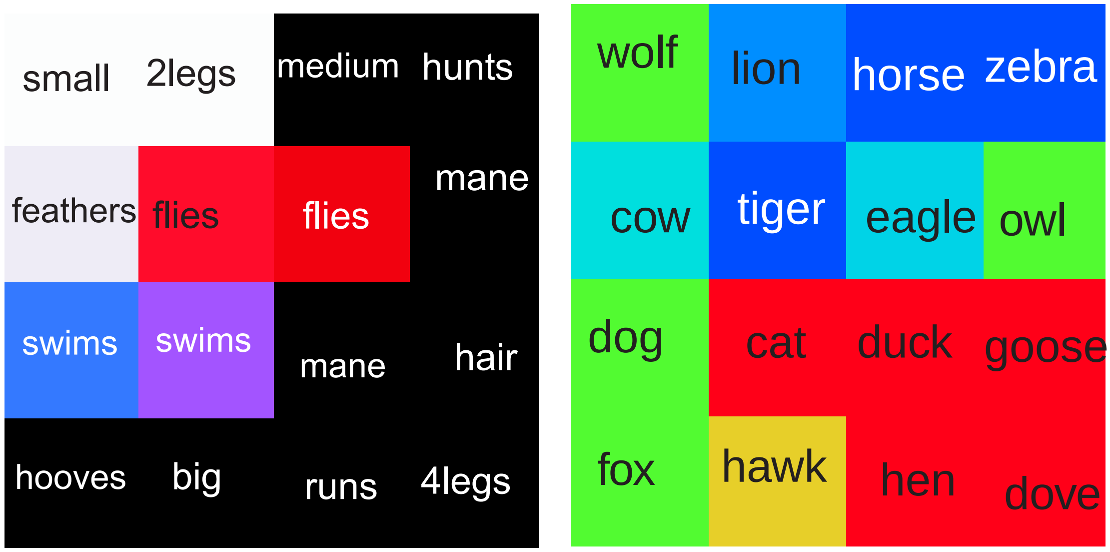


Figure 3. The animal dataset, each cell labelled. **Left**. *Attributes* describing groups of animals. **Right**. *Entities* (animals) sharing attributes.

tor, venomous. Sharing the terms *domestic* we find animals such as *chicken, calf, pussycat*, and so forth. We may consider *backbone* as an important attribute for it determines whether these animals are vertebrate or not (Fig.5).

The experiments we have presented in this section show that Self-Organizing Maps are an efficient software tool to analyse domains. We have reported the use of SOM for other domains [7]. The next step would be to use other ontology tools to organize and embed this knowledge into web pages.

5 Conclusion

An ontology can be used to give a sense of order to unstructured digital sources. It also provides a common vocabulary of *concepts* and *relationships* which may be used to inform a viewer, a search engine, or other software entities such as agents. A common ontology would enable collaborators to work together with a minimal risk of misunderstanding. Principled techniques that allow the ontological engineer to deal with the problems caused by such complexity need to be developed, and the ideas in this paper have shown promise as avenues of investigation. The novelty of our approach is that SOM offer clustering and visualization features not present in other techniques, and as it has been presented helps in the semi-automatic construction of ontologies by identifying components from digital archives. Further research avenues we are working on involve the use of hybrid systems in such a way that by combining clustering techniques with the already trained feature vectors we may refine the classification of the knowledge componentes from the domain [19, 20]. Must be said that a domain expert is always required in order to obtain a desirable level of accuracy in the ontology. Should that be done manually, then the semantic web will not become a really in the next couple of decades due to this bottleneck. Ontology learning tools are essential for the realization of the semantic web for the job to be done is quite complex.

References

[1] R Benjamins et al. $(KA)^2$: Building ontologies for the internet: A midterm report. *Int.J.Human-Computer Studies*, 51(3):687–712, 1999.

[2] B Berent et al. A roadmap for web mining: from web to semantic web. In B Berent et al., editors, *First European Web Mining Forum (EWMF)*, volume 3209 of *LNCS*, pages 1–22. Springer, 2004.

[3] T Berners-Lee et al. The Semantic Web. *Scientific American*, 284(5):34–43, May 2001.

[4] R Brachman. What is-a and isn't: An analysis of taxonomic links in semantic networks. *IEEE Computer*, 16(10):10–36, 1983.

[5] L Crow and N Shadbolt. Extracting focused knowledge from the Semantic Web. *Int.J.Human-Computer Studies*, 54:155–184, 2001.

[6] D Elliman and *JRG Pulido*. Visualizing ontology components through self-organizing maps. In D Williams, editor, *6th International Conference*

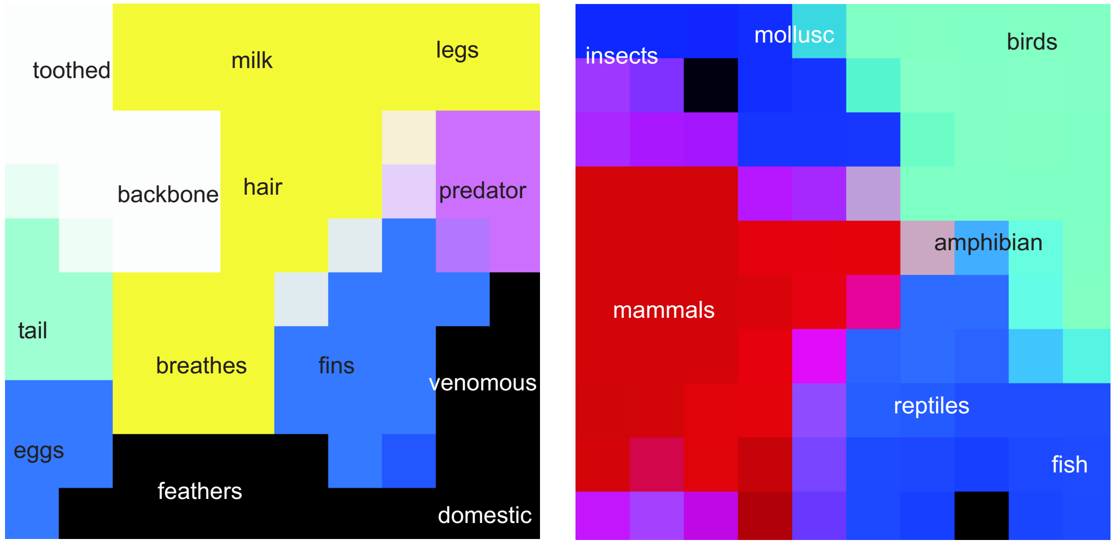


Figure 4. The zoo dataset, groups labelled. **Left**. *Attributes* describing groups of animals. **Right**. *Entities* (animals) sharing attributes.

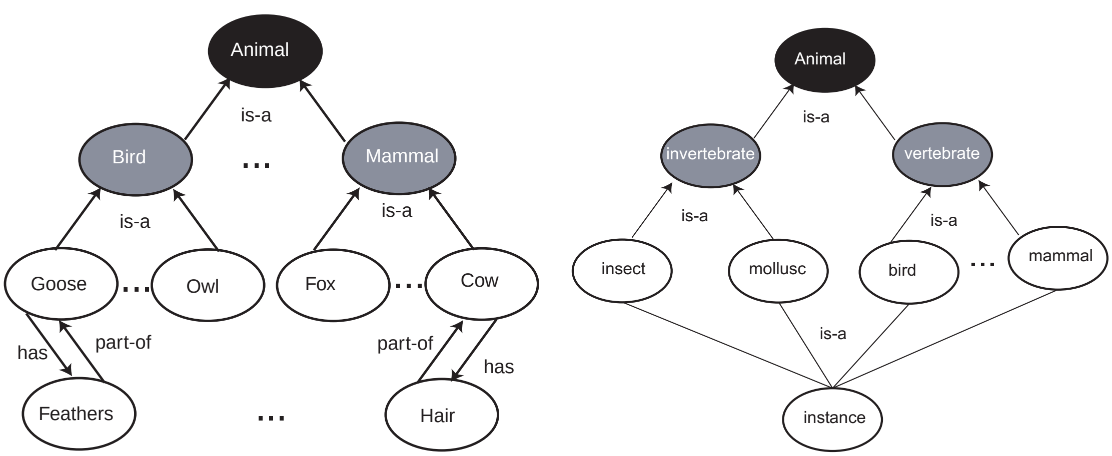


Figure 5. Two basic taxonomies from the animal kingdom. **Left**. Animal dataset (sec.4.1), attributes at the bottom are *part-of* instances. **Right**. Zoo dataset (sec.4.2), attributes not shown. Note that the former is actually a subset of the latter (vertebrate).

on Information Visualization (IV02), London, UK, pages 434–438. IEEE Computer Soc.Press, Los Alamitos, 2002.

[7] D Elliman and *JRG Pulido*. Self-organizing maps for detecting ontology components. In H Arabnia et al., editors, *The 2003 Int.Conf.on Artificial Intelligence (IC-AI), Las Vegas, USA*, pages 650–653. CSREA Press, 2003.

[8] D Fensel et al. OIL in a nutshell. In R Ding et al., editors, *Proc.European Knowledge Acquisition Conf.*, LNAI. Springer-Verlag, Berlin, 2000.

[9] S Geffner et al. Browsing large digital library collections using classification hierarchies. In S Gauch, editor, *8th Int. Conf. on Information Knowledge Management, CIKM'99*, pages 195–201. ACM, New York, 1999.

[10] A Gómez and Oscar Corcho. Ontology languages for the Semantic Web. *IEEE Intelligent Systems*, 2002.

[11] A Gómez et al. Knowledge maps: An essential technique for conceptualisation. *Data & Knowledge Engineering*, 33:169–190, 2000.

[12] J Gordon. Creating knowledge maps by exploiting dependent relationships. *Knowledge-Based Systems*, pages 71–79, 2000.

[13] J Heflin and J Hendler. Dynamic ontologies on the web. In *American Association For Artificial Intelligence Conf.*, pages 251–254. AAAI Press, California, 2000.

[14] J Hendler and E Feigenbaum. Knowledge is power: The Semantic Web vision. In N Zhonh et al., editors, *Web intelligence: Research and development*, volume 2198 of *LNAI*, pages 18–29. Springer-Verlag, Berlin, 2001.

[15] I Horrocks et al. From SHIQ and RDF to OWL: The making of a web ontology language. *Journal of web semantics*, 2003.

[16] S Kaski et al. WEBSOM - Self-organizing maps of document collections. *Neurocomputing*, 6:101–117, 1998.

[17] T Kohonen. *Self-Organizing Maps*. Information Sciences Series. Springer-Verlag, Berlin, 3rd edition, 2001.

[18] T Kohonen et al. Self organization of a massive text document collection. In E Oja and S Kaski, editors, *Kohonen Maps*, pages 171–182. Elsevier Sci, Amsterdam, 1999.

[19] S Legrand and *JRG Pulido*. A hybrid approach to word sense disambiguation: Neural clustering with class labeling. In P Buitelaar et al., editors, *Workshop on knowledge discovery and ontologies, 15th European Conference on Machine Learning (ECML), Pisa, Italy*, pages 127–132, September 2004.

[20] A Maedche and V Zacharias. Clustering ontology-based metadata in the semantic web. In T Elomaa et al., editors, *Principles of Data Mining and Knowledge Discovery: 6th European Conference, PKDD 2002*, volume 2431 of *LNCS*, pages 348–360. Springer, 2002.

[21] J McCormack and B Wohlschlaeger. Harnessing agent technologies for data mining and knowledge discovery. In *Data Mining and Knowledge Discovery: Theory, Tools and Technology II*, volume 4057, pages 393–400, 2000.

[22] E Motta et al. Ontology-driven document enrichment: principles, tools and applications. *Int.J.Human-Computer Studies*, 52:1071–1109, 2000.

[23] A Rauber and D Merkl. Automatic labeling of self-organizing maps: Making a treasure-map reveal its secrets. In *Pacific-Asia Conf. on Knowledge Discovery and Data Mining*, pages 228–237, 1999.

[24] H Ritter and T Kohonen. Self-organizing semantic maps. *Biological Cybernetics*, 61:241–254, 1989.

[25] I Sanz et al. Gathering metadata from web-based repositories of historical publications. In A Tjoa and R Wagner, editors, *9th Int. Workshop on Database and Expert Systems Apps*, pages 473–478. IEEE Computer Soc.Press, Los Alamitos, 1998.

[26] P Schneider and D Fensel. Layering the Semantic Web: Problems and directions. In I Horrocks and J Hendler, editors, *1st Int.Semantic Web Conf.*, volume 2342 of *LNCS*, pages 16–29. Springer-Verlag, Berlin, 2002.

[27] M Uschold and M Gruninger. Ontologies: Principles, methods, and applications. *Knowledge Engineering Review*, 11(2):93–155, 1996.

[28] A Waterson and A Preece. Verifying ontological commitment in knowledge-based systems. *Knowledge-Based Systems*, 12:45–54, 1999.

[29] Y Yang et al. A study of approaches to hypertext categorization. *J.Intelligent Information Systems*, 18(2/3):219–241, 2002.

Article VII

BUILDING AN OPTIMAL WSD ENSEMBLE USING PER-WORD SELECTION OF BEST SYSTEM

by

H. Saarikoski, S. Legrand 2006

In J. F. Martínez-Trinidad, J. A. C. Ochoa, J. Kittler (Eds.): Progress in Pattern Recognition, Image Analysis and Applications : *11th Iberoamerican Congress in Pattern Recognition, CIARP 2006 Cancun, Mexico, November 14-17, 2006 Proceedings. Lecture Notes in Computer Science*. No. 4225. Springer Berlin / Heidelberg, 2006, pp. 864-872.

Building an Optimal WSD Ensemble Using Per-Word Selection of Best System

Harri M.T. Saarikoski[1] and Steve Legrand[2]

[1] KIT Language Technology Doctorate School, Helsinki University, Finland
Harri.Saarikoski@helsinki.fi
[2] Department of Computer Science, University of Jyväskylä, Finland
stelegra@cc.jyu.fi

Abstract. In Senseval workshops for evaluating WSD systems [1,4,9], no one system or system type (classifier algorithm, type of system ensemble, extracted feature set, lexical knowledge source etc.) has been discovered that resolves all ambiguous words into their senses in a superior way. This paper presents a novel method for selecting the best system for target word based on readily available word features (number of senses, average amount of training per sense, dominant sense ratio). Applied to Senseval-3 and Senseval-2 English lexical sample state-of-art systems, a net gain of approximately 2.5 - 5.0% (respectively) in average precision per word over the best base system is achieved. The method can be applied to any base system or target word in any language.

1 Introduction

Based on recent evaluation of WSD systems, progress in disambiguation methods have reached a standstill. The 15 best systems in Senseval-3 English sample task ended up within 2% of each other [10] while in Senseval-2 the number of systems within that range was only five [1]. Numerous methods of disambiguation have been tried out in Senseval evaluations. For instance, most classifiers found effective in data mining experiments have been tried out: in Senseval-3 for example there were experiments with support vector machines (IRST-kernel, nusels), Naive Bayes (CLaC1, all htsa systems), Neural Networks (MC-WSD, UJAEN) and Maximum Entropy algorithms (HKUST-me, CLaC2) [10]. Multi-classifier experiments have also been very popular [3,19,12,17]: in Senseval-3 evaluation, classifier ensembles were as popular as single-classifier systems (e.g. SyntaLex, NRC, HKUST-all and BCU systems, and Duluth-ELSS) [10].

The first conclusion from these experiments has been that different disambiguation methods result in different performance results. A second conclusion is that there is a 'word bias', i.e. each word poses a different set of learning problems. To solve these biases, all we need is an exact definition of the type of system that is best equipped to handle a particular target word. [18] showed that word grain, amount of training and most frequent (dominant) sense bias in training data are factors that have a profound influence on system performance. For instance, disambiguating a hard word (40

J.F. Martínez-Trinidad et al. (Eds.): CIARP 2006, LNCS 4225, pp. 864 – 872, 2006.

senses, average of 4 training examples per sense out of which dominant sense gets 25%) is a different type of learning task than disambiguating an easy word (2-sense word with 40 examples at 80% dominant share). Since classifiers have different solutions to deal with the different learning tasks, it is reasonable to assume that system strengths tend to follow changes (drops and rises) in these three word factors. We further propose that system strength is focused on a particular region of this 'word space' (see Figures 1 and 2), which allows effective predictors of best system per word to be built.

This paper presents a novel method using the three word features that fairly accurately predicts the strong regions of given base systems. To our knowledge, only one such per-word ensemble using word features as system selection criterion has been implemented [11] where they selected the system according to target word part-of speech. Despite the fact that the two-system ensemble ended up at the bottom of the Senseval-2 evaluation (20% off the state of the art), it still achieved three wordwins, which indicates the viability of the per-word selection method in general.

In section 2, we present the machine-learning tools we used for predictions. In section 3, we define the three word-based factors and the predictors built on them. In section 4, we present the disambiguation method based on those predictors, and in section 5, we test the method in practice on two different datasets. Sections 6 and 7 discuss and conclude on the findings.

2 MOA-SOM Toolkit

Study of disambiguation systems lacks a diagnostic tool that could be used to meta-learn the effects of these factors. As a result, the following types of questions are largely unanswered: Which are the words where a system is at its strongest? What type of ensembles of systems achieve optimum performance for give target word?

We are developing a meta-classifier (MOA-SOM, 'mother-of-all-self-organizing-maps') to handle such learning tasks. The tool clusters publicly available WSD system scores [10,1,4] stored in database [13] based on features defining the systems (e.g. classifier algorithm, feature sets) and target words (e.g. PoS, training, word grain) by calculating the amount of correlation between systems and words. The output from MOA-SOM is the optimal classifier, feature and configuration for that target word. The feature matrix can be fed to SOM using either system names as labels and words as data points or vice versa. The SOM used is based on hierarchically clustering DGSOT [7] which was found useful in earlier WSD experiments [6]. For these tests we additionally employed the machine-learning algorithms implemented in Weka toolkit [16] for predictors.

In the next section, we present the three factors in more detail and how we combined them to build machine-learning predictors of system differences.

3 Predictor Building

In this section we present the factors predicting system performance and the predictors using those factors for prediction of system differences.

3.1 Factors

We introduce here the three word-based factors in explaining variations in system performance (Train, Grain, and DomSub). *Train* is average number of training instances per sense, *Grain* is the number of senses (as recorded in WordNet / WordSmyth sense repositories used in Senseval evaluations). DomSub is aimed to differentiate between systems that differ in their inherent bias to deal with big vs small dominant (most frequent) sense shares. The formula we used for DomSub is $\texttt{DomSub} = \texttt{dom}^2 + \texttt{sub}^2$ where *dom* and *sub* are the shares of dominant and subdominant sense out of all training instances for the current target word[1]. For example, for a word with 80% / 20% shares for dominant / subdominant senses, DomSub is $0.8^2 + 0.2^2 = 0.68$.

Next we present the types of predictors we used in our experiments.

3.2 Predictors

A few factor formulas emerged as best predictors of system difference predictors. To train the predictors, we used both manual rules and machine-learning algorithms:

(1) Bisections (baseline). To achieve a bisection baseline, we first sort the data according to a selected factor (e.g. T, G, D, T+G+D), then split the data in two and calculate the net gain by each system for each half and average that by dividing it by two. The best weighting scheme we found was the square root of the unweighted sum of normalized values of the three factors: `sqrt (a*T + b*G + c*D)` where *G* stands for Grain, *T* for Train, *D* for DomSub values of target words and integers *a*, *b* and *c* normalize the weights of the three factors. Note that since this set of predictors is limited to one factor at a time, it cannot express decision rules containing multiple factors which tends to make them less reliable.

(2) Machine-learned models. To predict the best system for words, we trained some of the most efficient learning algorithms implemented in Weka toolkit [16] (Support Vector Machine, Maximum Entropy, Naive Bayes, Decision Trees, Random Forests as well as voting committee, training data bagging and algorithm boosting methods). For training we used the abovementioned word factors both individually and in various permutations (e.g. T-G).

Next we outline the method of using these predictors for system-word pairing.

4 Method

In this section, we outline a method for defining and selecting maximally complementary base systems integrated inside a disambiguation algorithm:

1. **Base system selection.** Run candidate base systems on training words. Investigate their performance at different types of words. Based on their performance at training words, select systems whose strong regions are as large and as distinct as possible using the following criteria:

---

[1] We consider the increment from the rest of the senses (typically < 0.05) as largely negligible.

- biggest gross gain (see Evaluation) from candidate base systems
- largest number of training words won by the system

2. **Training the predictor.** Using the training run data, train the predictors to recognize the best base system using readily available factors (e.g. word grain). Predictor can be constructed by setting decision rules manually, e.g. "use system#1 (Decision Tree -based) when number of senses (grain) < 5, system#2 (Naive Bayes -based) when grain is > 5 but not when 20 < train < 25". Alternatively, use a machine-learning algorithm to induce the rules from the training data.
3. **Testing.** Run selected base systems and the ensemble (according to the best predictor for that ensemble) on test words.
4. **Evaluation.** Evaluate the ensemble by comparing it to the better of the base systems. Also evaluate the predictor using *net gain* measure calculated from the following formula:

```
((PredictionAccuracy - (1.0 / NumberOfSystems))
                       *2)
                 * GrossGain
```

PredictionAccuracy is the number of correct system-for-word predictions out of all test words and *NumberOfSystems* is the number of classes/systems to predict. *GrossGain* is a measure of the potential of the base systems when they form an ensemble, resulting from a perfect system-for-word prediction for all test words. It is calculated from all-words average net gain by either base system (e.g. in a test set of two words, if system#1 wins over system#2 by 2% at word#1 and system#2 wins over system#1 by 4% at word#2, then gross gain for all test words is (2+4) / 2 = 3%). Net gain is then calculated as follows: in a two-system ensemble with 0.80 prediction accuracy and 8.0% gross gain, net gain is ((0.80-0.50)*2)) * 8.0% = 4.8%. It should be noted that in a two-system prediction task, prediction accuracy of 0.50 results in zero net gain, same as random selection of system.

Next we apply this method to two separate Senseval datasets (four prediction tasks each), using state of the art systems and predictors that proved best in our tests.

5 Testing

In this section, we apply the method to two Senseval evaluations.

5.1 Senseval-2 English Lexical Sample

System Selection. We trained the predictors with 39 words [2] and considered all supervised systems as candidates for base systems [3]. We selected the systems based on criteria in Step 1 of the method: looking at wins by best systems in training words, SMU [9] got 10 wins, JHU [17] nine, KUN(LP) [14] got four and CS(224N) [8] three.

[2] We discarded words where the wordwinner system's margin over next best system was < 2%.

[3] We ignored low-recall (<99%) and low-precision (> 4% behind best) systems.

Strong region of the latter was almost identical the same as that of JHU, yet smaller, and even though the abovementioned Alicante system [8] scored 3 wordwins, it cannot be used because of its poor overall performance (20% behind top) (criteria 3). In Figure 1, we see the Train-Grain regions (most important criteria 1) of the two top 'wordwinners' (SMU and JHU).

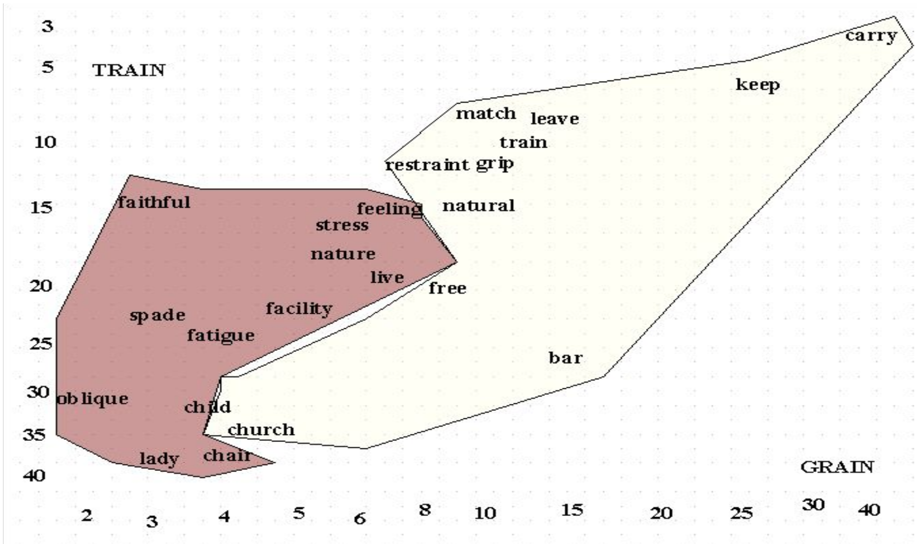


Fig. 1. Strong regions of two Senseval-2 systems in Train-Grain space (sample of training words shown). JHU region can be found on right (mid/high-grain), SMU region on left (low-grain, mid/high-train).

Table 1. Results from applying the method on selected base systems from Senseval-2

system pair (gross gain)	best predictor (factor/classifier)	prediction accuracy	net gain of ensemble
JHU+SMU (8.0%)	(1) (T-G) / (T+G)	0.63	2.6%
	(2) SVM *	0.80	4.8%
SMU+KUN (**8.4%**)	(1) T+G+D	0.70	3.4%
	(2) SVM	**0.82**	**5.0%**
JHU+KUN (5.5%)	(1) T+G+D	0.56	1.7%
	(2) SVM	0.75	2.8%
JHU+SMU+KUN (9.5%)	(3) SVM	0.55 [4]	4.2%

We see from Figure 1 that SMU and JHU populate distinct and large learning regions in Grain-Train space. KUN (not showing) also occupies a large region, focused on high-grain, low-train words such as *call* and *dress* and would be located approximately between JHU's strong words *keep* and *leave*.

[4] Note that having three systems to predict yields a naturally lower prediction accuracy.

Strength of KUN is, however, in the steady quality of its performance with all words, not manifesting any huge drops with any word.

Testing. We tested the model(s) on 19 words and three possible two-system combinations of the three top wordwinning systems (SMU, JHU and KUN) as well as an ensemble of all three systems.

SMU+KUN appears to have the highest gross gain, prediction accuracy and net gain, making it the maximally complementary system pair for this dataset. Furthermore, it seems that 3-system prediction (JHU+SMU+KUN) with more gross gain loses to 2-system predictions in prediction accuracy ending up with a slightly lower net gain.

5.2 Senseval-3 English Lexical Sample

System Selection. We trained the predictors again with 39 words and considered 15 top systems and selected the three top wordwinners for candidate base systems: IRST-kernel [15] with 8, htsa3 [2] 4 and nusels [5] with 3 training words won. Let us investigate the strong Train-Grain regions of the two top wordwinners.

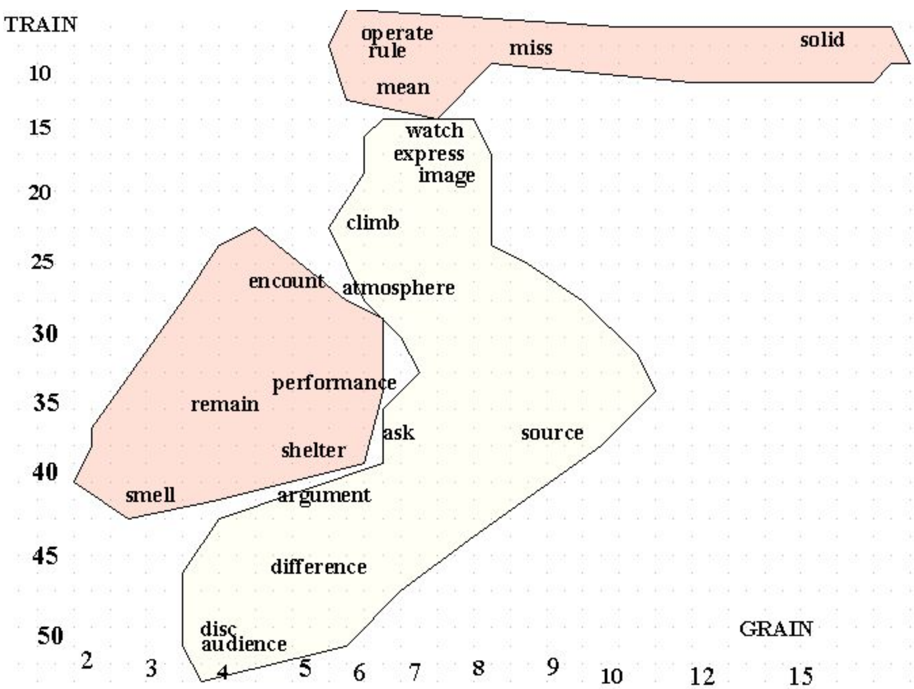


Fig. 2. Strong regions of two Senseval-3 systems in Train-Grain space (sample of training words shown). htsa3 is the lighter shade intact region in the middle, IRST-kernel holds the other two regions, one on left, one on top.

These two systems (htsa3 and IRST-kernel) complement each other very well. IRST-kernel occupies two regions but since training data contains no words from this area the regions cannot be merged.

Testing. We tested the model on 19 words and three two-system combinations of the three wordwinning systems (htsa3, IRST-kernel and nusels) as well as an ensemble of all three systems.

Table 2. Results from applying the method on selected base systems from Senseval-3

system pair (gross gain)	best predictor (factor/classifier)	prediction accuracy	net gain of ensemble
htsa3+IRST-kernel (4.1%)	(1) T+G+D	0.80	2.5%
	(2) NaiveBayes	**0.82**	**2.7%**
htsa3+nusels (3.6%)	(1) T+G+D	0.65	1.2%
	(2) DecisionTree	0.70	1.4%
nusels+ IRST-kernel (**4.4%**)	(1) (T-G) / (T+G)	0.80	2.6%
	(2) SVM	0.80	2.6%
htsa3+IRSTk+nusels (6.1%)	(2) MaxEnt	0.55	**2.7%**

Table 2 shows nusels+IK is the maximally complementary system pair in terms of net gain but another system pair (nusels+IRST-kernel) has the higher potential (gross gain). It should also be noted that the more challenging three-system prediction task (htsa3+IRSTk+nusels) produces equally high net gain as htsa3+IRST-kernel pair.

6 Discussion

Best predictors turned out to vary according to base system pair, both in terms of learning algorithm and input features. The most reliable learning algorithms turned out to be Support Vector Machines and slightly less consistently Maximum Entropy and Naive Bayes classifiers. Machine-learning models (2) tend to work better than the corresponding bisection baseline (1). The contribution of individual factors to system differentiation seems to depend heavily on the base system pair: combination of factors tended to work better than individual factors but there were (e.g. T+G+D for SMU/JHU pair) but sometimes one factor differentiated better (e.g. DomSub for IRST-kernel/htsa3). These findings lead us to conclude that this system prediction task - just like word sense disambiguation task itself - is in fact a set of tasks dependent on details and the difficulty of the task, and therfore, a customized predictor may need to be developed for given system pair.

7 Conclusion

We have presented a novel method for constructing effective WSD system ensembles. Predictors built on word-based factors (Grain, Train, DomSub) seem to yield very good predictions of optimal systems for words. The method was tested with two evaluations: in Senseval-2 the best net gain was 5.0% (out of maximal 8.4%) for

SMU/JHU pair, while in the more contested Senseval-3 it was 2.7% (out of 4.1%) for three-system ensemble (htsa3/IRST-kernel/nusels). The method is scalable to any ambiguous word and any assortment of base systems and the factors used to build predictors are readily available for all words.

Although most predictors exceed random selection baseline (zero net gain), further work is needed to make the prediction method more accurate and thereby maximize net gain. It should be kept in mind that base systems and their optimal predictors form a pair. Based on a more covering set of factors, we can then learn more reliable predictors, including for more than two or three systems. Particularly we need to account for other factors found influential to system performance: choice of feature sets [6,18] as well as choice of the classifier algorithm as well as the specifics of its sense decision procedure [2,18,19]. We also believe it is possible to fabricate two strong, 'opposite' systems that together optimally cover the word space, which probably makes predictions more reliable too. Furthermore, as with any supervised prediction system, providing more training words to a machine learner is likely to improve prediction accuracy (e.g. all the 165 English words and around 100 systems that have participated in the three English Senseval evaluations). Evaluation data from other languages can also be used since our preliminary comparisons with Spanish Senseval-2 data indicate that the same systems (JHU, CS, Duluths, UMCP) that excelled in a particular region in the English evaluation did so in Spanish as well. This phenomenon can be explained by the language-independence of both word factors and WSD systems and suggests that it is possible to build one 'optimal ensemble' that would be as effective for all languages.

Acknowledgments

Authors would like to express their gratitude to professor Alexander Gelbukh of Instituto Politecnico Nacional, Mexico, for his review of our paper. Work of first author is supported by Department of Translation Studies (Language Technology Doctorate School) of Helsinki University, Finland, and second author by COMAS Doctorate School of Jyväskylä University and Instituto Politecnico Nacional, Mexico.

References

1. Edmonds, P., and Kilgarriff, A. Introduction to the Special Issue on evaluating word sense disambiguation programs. Journal of Natural Language Engineering 8(4) (2002).
2. Grozea, C. Finding optimal parameter settings for high performance word sense disambiguation. In SENSEVAL-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, Barcelona, Spain (2004).
3. Hoste, V., Hendrickx, I., Daelemans, W. and A. van den Bosch. Parameter optimization for machine-learning of word sense disambiguation. Journal of Natural Language Engineering, 8(4) (2002) 311-327.
4. Kilgarriff, A. SENSEVAL: An Exercise in Evaluating Word Sense Disambiguation Programs. In Proceedings of LREC, Granada (1998) 581--588.
5. Lee, Y-K., Ng, H-T., and Chia, T-K. Supervised Word Sense Disambiguation with Support Vector Machines and Multiple Knowledge Sources. In Proceedings of SENSEVAL-3 workshop (2004).

6. Legrand, S., Pulido JGR. A Hybrid Approach to Word Sense Disambiguation: Neural Clustering with Class Labeling. Knowledge Discovery and Ontologies workshop at 15th European Conference on Machine Learning (ECML) (2004).
7. Luo, F., Khan, L., Bastani F., Yen I-L and Zhou, J. A dynamically growing self-organizing tree (DGSOT) for hierarchical clustering gene expression profiles, Bioinformatics 2004 20(16):2605-2617, Oxford University Press.
8. Manning, C., Tolga Ilhan, H., Kamvar, S., Klein, D. and Toutanova, K. Combining Heterogeneous Classifiers for Word-Sense Disambiguation. Proceedings of SENSEVAL-2, Second International Workshop on Evaluating WSD Systems (2001) 87-90.
9. Mihalcea, R. Word sense disambiguation with pattern learning and automatic feature selection. Journal of Natural Language Engineering, 8(4) (2002) 343-359.
10. Mihalcea, R., Kilgarriff, A. and Chklovski, T. The SENSEVAL-3 English lexical sample task. Proceedings of SENSEVAL-3 Workshop at ACL (2004).
11. Montoyo, A. and Suárez, A. The University of Alicante word sense disambiguation system. Proceedings of SENSEVAL-2 Workshop (2001) 131-134.
12. Mooney, R. Comparative experiments on disambiguating word senses: An illustration of the role of bias in machine learning. Proceedings of the Conference on Empirical Methods in Natural Language Processing (1996).
13. Saarikoski, H. mySENSEVAL: Explaining WSD System Performance Using Target Word Features. In NLDB-05, Natural Language for Databases: Proceedings of 10th International Conference on Applications of Natural Language Processing to Information Systems. Lecture Notes in Computer Science, N 3513, Springer (2005) 369–371.
14. Seo, H-C., Rim, H-C. and Kim, S-H. KUNLP system in Senseval-3. Proceedings of SENSEVAL-2 Workshop (2001) 222-225.
15. Strapparava, C., Gliozzo, A., and Giuliano, C. Pattern abstraction and term similarity for Word Sense Disambiguation: IRST at Senseval-3. In Proceedings of SENSEVAL-3 workshop (2004).
16. Witten, I., Frank, E. Data Mining: Practical Machine Learning Tools and Techniques (Second Edition). Morgan Kaufmann (2005).
17. Yarowsky, D., S. Cucerzan, R. Florian, C. Schafer and R. Wicentowski. The Johns Hopkins SENSEVAL2 System Descriptions. Proceedings of SENSEVAL-2 workshop (2002).
18. Yarowsky, D. and Florian, R. Evaluating sense disambiguation across diverse parameter spaces. Journal of Natural Language Engineering, 8(4) (2002) 293-311.
19. Zavrel, J., S. Degroeve, A. Kool, W. Daelemans, K. Jokinen. Diverse Classifiers for NLP Disambiguation Tasks. Comparisons, Optimization, Combination, and Evolution. TWLT 18. Learning to Behave. CEvoLE 2 (2000) 201–221

YHTEENVETO (FINNISH SUMMARY)

Teesini pyrkii osoittamaan kuinka sanojen merkitysten disambiguaatiota (WSD) voidaan parantaa ontologioihin koodatun taustatiedon ja erityisesti psykologisen taustatiedon avulla. Tämä taustatieto koostuu osittain ns. maailmantietoudesta. Syy ontologioiden käyttöön tässä tehtävässä on se, ettei konseptuaalista taustatietoa voida saada suoraan teksteistä WSD-järjestelmiin. Vaikka tekstin disambiguaatio on tietyssä määrin mahdollista ilman ontologioihin turvautumista, tällaisen taustatiedon käyttö WSD:ssä on erittäin hyödyllistä, erityisesti Semantic Webin kaltaisessa ympäristössä. Semantic Web (Semanttinen verkko) onkin ollut tämän teesin pääasiallinen motivaattori.

Suuri osa maailmantietoudesta, mikä on välttämätöntä ihmisen ymmärryksen kannalta, on vaikea koodata tavanomaisiin ontologioihin. Perustason kategoriat (basic-level categories) on eräs toiminnallisen tai kehollistettun (embodied) tiedon tyyppi ja liittyy myös maailmantietouteen sen psykologisena osana. Teesi kuvaa kuinka tavanomaisten ontologioiden ja itseorganisoituvien karttojen (SOM) avulla konseptit tekstissä voidaan automaattisesti ryhmitellä ja nimetä disambiguaatiota varten. Tätä ideaa laajennetaan ja sovelletaan sitten perustason kategorioihin perustuviin ontologisiin rakenteisiin. Teesi osoittaa, että perustason kategorioiden käyttö WSD:ssä parantaa huomattavasti disambiguaation tarkkuutta. Se myös selittää, kuinka linguistisia käsitteitä ,kuten metaforia, voidaan rakenteellisesti manipuloida siten että niistä tulee perustason komponenttaja, siis potentiaalisia tekijöitä WSD:n.

Teesin tarjoamaa lähestymistapaa voidaan hyödyntää sovelluksessa joka ei ainoastaan disambiguoi hybridien järjestelmien (sis. ontologisen maailmantietouden komponentin) avulla vaan myös valitsee parhaiten soveltuvan disambiguaatiojärjestelmän kullekin sanalle.

Asiasanoja: word sense disambiguation, WSD, basic-level categories, real-world knowledge, background knowledge, Semantic Web, ontology.