

Extraction of Implicit Knowledge from WordNet

Wim Peters

NLP Group
Department of Computer Science
University of Sheffield
Regent Court
211 Portobello Street
Sheffield S1 4DP
U.K.
wim@dcs.shef.ac.uk

Abstract

Lexical knowledge databases such as WordNet contain much semantic information that is left implicit. In order to make maximal use of these resources it is important to make this implicit semantic information explicit. Metonymy and regular polysemy constitute a type of implicit ontological knowledge. This paper describes the semi-automatic extraction of systematically related word senses from WordNet by exploiting its hierarchical structure, and the identification of relations that link these on the basis of the glosses.

1. Introduction

WordNet (Fellbaum 1998) contains far more semantic information than its ontological organization shows. Word senses are related to senses of other words by means of a small number of basic semantic relations such as synonymy and hypernymy. Other types of encyclopaedic knowledge and semantic relations are implicitly present in the structure of WordNet in the form of taxonomic correspondences and glosses. This non-formalized semantic information in WordNet can be processed in order to distil more implicit knowledge (see e.g. Harabagiu 2000).

2. Relations between senses

Systematic relatedness between senses is one type of knowledge that is mostly left implicit in resources. This phenomenon is called metonymy, or, more specifically, regular polysemy (Apresjan 1973).

Viewed traditionally, metonymy is a non-literal figure of speech in which the name of one thing is substituted for that of another related to it. It has been described as a cognitive process in which one conceptual entity, the vehicle, provides mental access to another conceptual entity (Radden 1999). In its basic form, it establishes a semantic relation between two concepts that are associated with word forms. The semantic shift expressed by the relation may or may not be accompanied by a shift in form. The semantic relation that is captured by metonymy is one of semantic contiguity, in the sense that in many cases there are systematic relations between metonymically related concepts that can be regarded as slots in conceptual frames (cf. Fillmore 1977).

Regular polysemy is a more specific instantiation of metonymy that covers the systematicity of the semantic relations involved. It can be defined as a subset of metonymically related senses of the same word displaying a conventional as opposed to novel type of semantic contiguity relation. Any systematic semantic relations between concepts are lexicalized, i.e. they are explicitly listed in dictionaries and independent of a pragmatic situation. For example, *The White House* is on the one hand an institution and on the other a building. The semantic relation between the two senses is 'is housed in'. It is a conventional pattern, not a nonce formation (a pragmatically defined novel metonymy), because it holds for related senses of two or more words (Apresjan, 1973) in the lexicon. It is this subtype of metonymy that we concentrate on in this paper.

3 Extraction from WordNet

A technique was developed (Peters 2000) for identifying sense combinations in WordNet where the senses involved potentially display a regular polysemic relation, i.e. where the senses involved are candidates for systematic relatedness.

In order to obtain these candidate patterns WordNet (WN) has been automatically analysed by exploiting its hierarchical structure for nouns. Wherever there are two or more nouns with senses in one part of the hierarchy, which also have senses in another part of the hierarchy, then we have a candidate pattern of regular polysemy. The patterns are candidates because there seems to be an observed regularity for two or more words. An example can be found in Figure 1 below.

The synset 2 bag holds 518 words (*seafood, food, substance, food product, nutrient, object*). Only a subset of these words is related to *herring*, the rest are associated with the other words that are subsumed by the hypernymic pattern. The concept 'food fish' is the hypernym of sense 1: "any fish used for food by human beings". Of the words in this

gloss 'fish' is found in the synset1 bag and 'food' in the synset 2 bag. The intermediate text span is 'used for' which consists of a past participle and a preposition. The outcome is the relationship 'animal used for food'. This relation is found 37 times. The relation 'used for' is found 23 times.

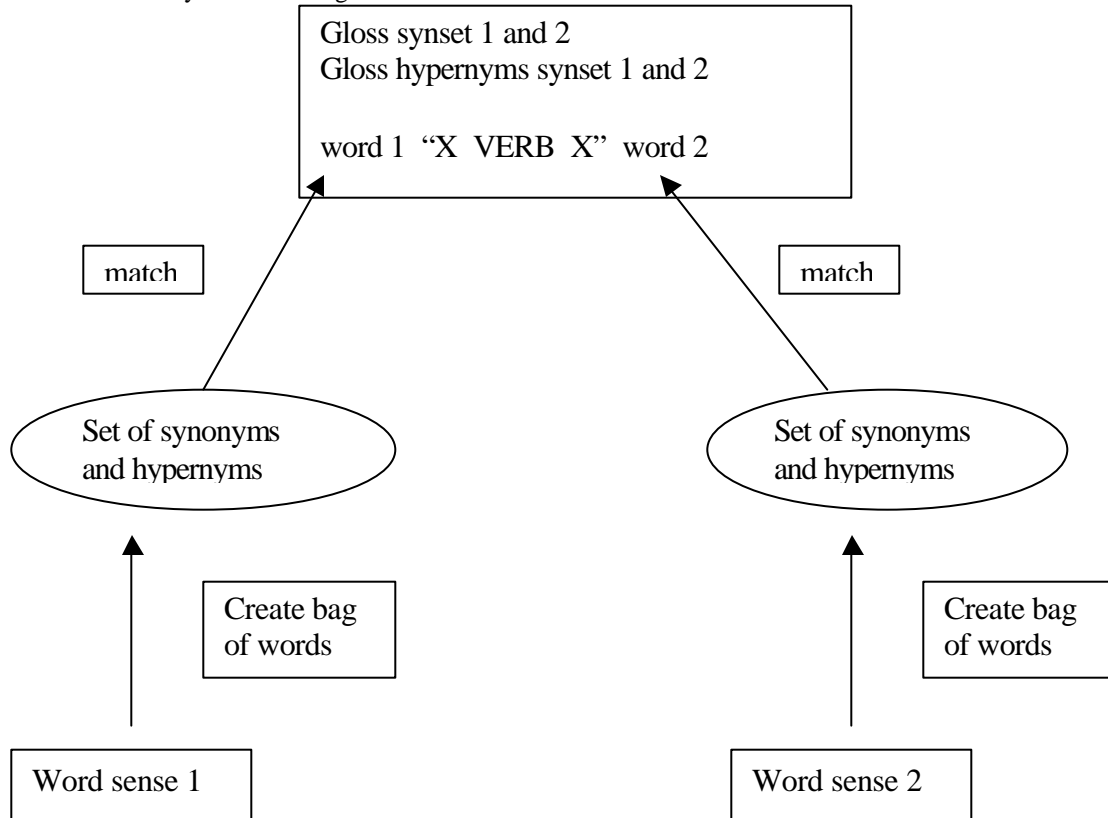


Figure 2: Mapping synset members onto glosses

The pattern **profession** and **discipline** (see figure 1) subsumes five words: *architecture, literature, politics, law* and *theology*.

Sense 6 of 'law' has the gloss 'the learned profession that is mastered by graduate study in a law school and that is responsible for the judicial system; "he studied law at Yale"

Bag synset 1 contains 'profession', bag synset2 'study'. In between is the verb 'is mastered by' which yields the relation 'profession is mastered by discipline' for this regular polysemic pattern. This relation is found 2 times. One other relation was found: 'concerned with', which occurs only once.

Other relations are:

writing (reading matter; anything expressed in letters of the alphabet (especially when considered from the point of view of style and effect); "the writing in her novels is excellent")

message (what a communication that is about something is about)

This pattern covers 36 words. Examples are *account, conclusion, declaration, epitaph*. The relation 'express' occurs once, 'state' occurs 24 times.

fabric (something made by weaving or felting or knitting or crocheting natural or synthetic fibers)

covering (a natural object that covers or envelops)

This hypernymic combination subsumes 5 words: *fleece, hair, tapa, tappa, wool*.

'made from' occurs once.

person (a human being; "there was too much for one person to do")

language (a systematic means of communicating by the use of sounds or conventional symbols; "he taught foreign languages"; "the language introduced is standard throughout the text")

This pattern subsumes 257 words such as *Tatar, Assyrian, Hopi, Punjabi*.

The relation 'speak' occurs 132 times.

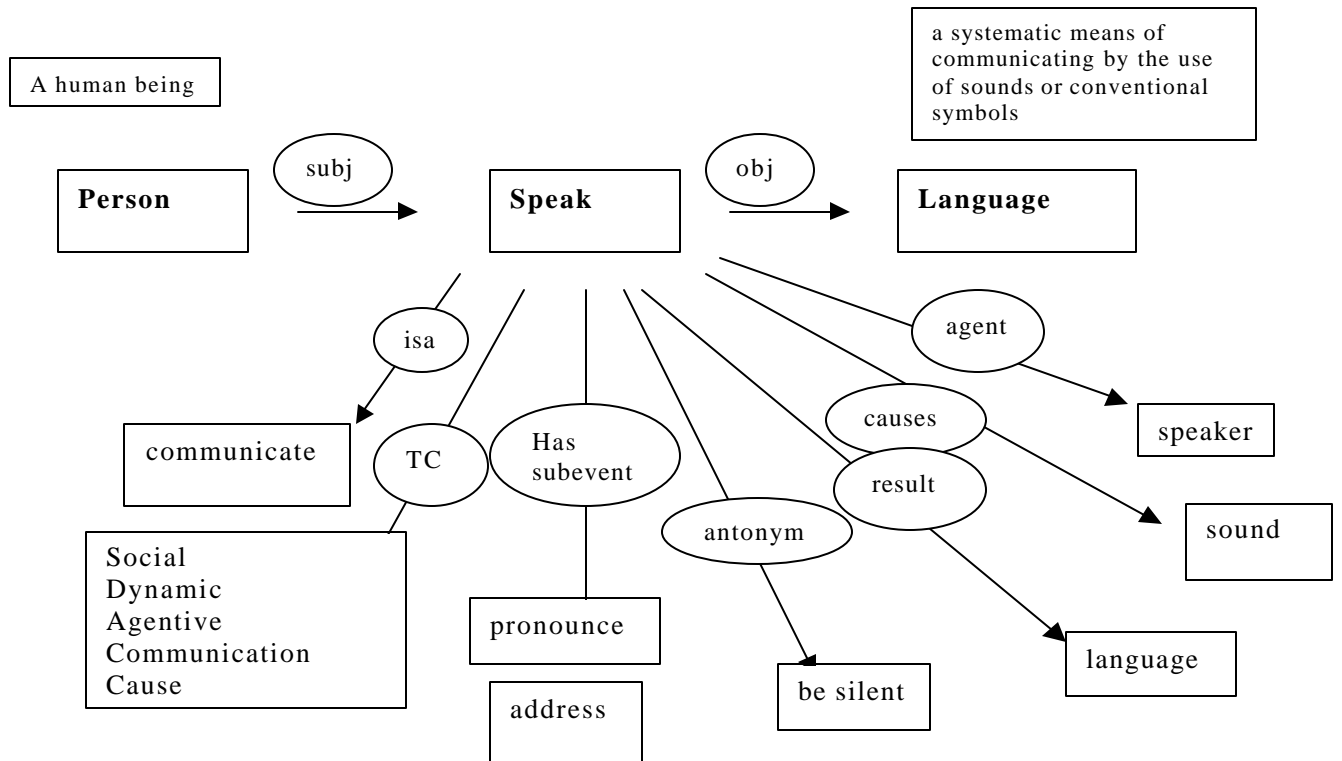


Figure 3: Expanded Ontological Fragment for the pattern person - speak - language

5 Expansion through EuroWordNet

Now we have obtained a number of patterns with specific relations it is possible to extend each ontological fragment consisting of concept triples (N-V-N) with explicit relations from EuroWordNet (Vossen, 1998). We have

chosen this database over Wordnet because it contains more kinds of semantic relations than WordNet, such as thematic relations and links that hold between concepts lexicalized by different parts of speech.

First, the applicable verb senses was chosen manually. After that, relational chains in the database were extracted. Figure 3 and 4 exemplify this process for the verbs 'speak' associated with the pattern person - language and 'master' linking profession and discipline. The 'TC' relation indicates the EuroWordNet top concepts that are described in great detail in (Rodriguez et al., 1998). The relations can all be considered additional slots in the

partial knowledge frame that started as a regular polysemic pattern. For instance, the additional knowledge fragments provided by EuroWordNet connect 'master' to 'knowledge', 'practice', 'learning' and 'teaching'. These can be used for inferencing purposes or knowledge extraction from texts.

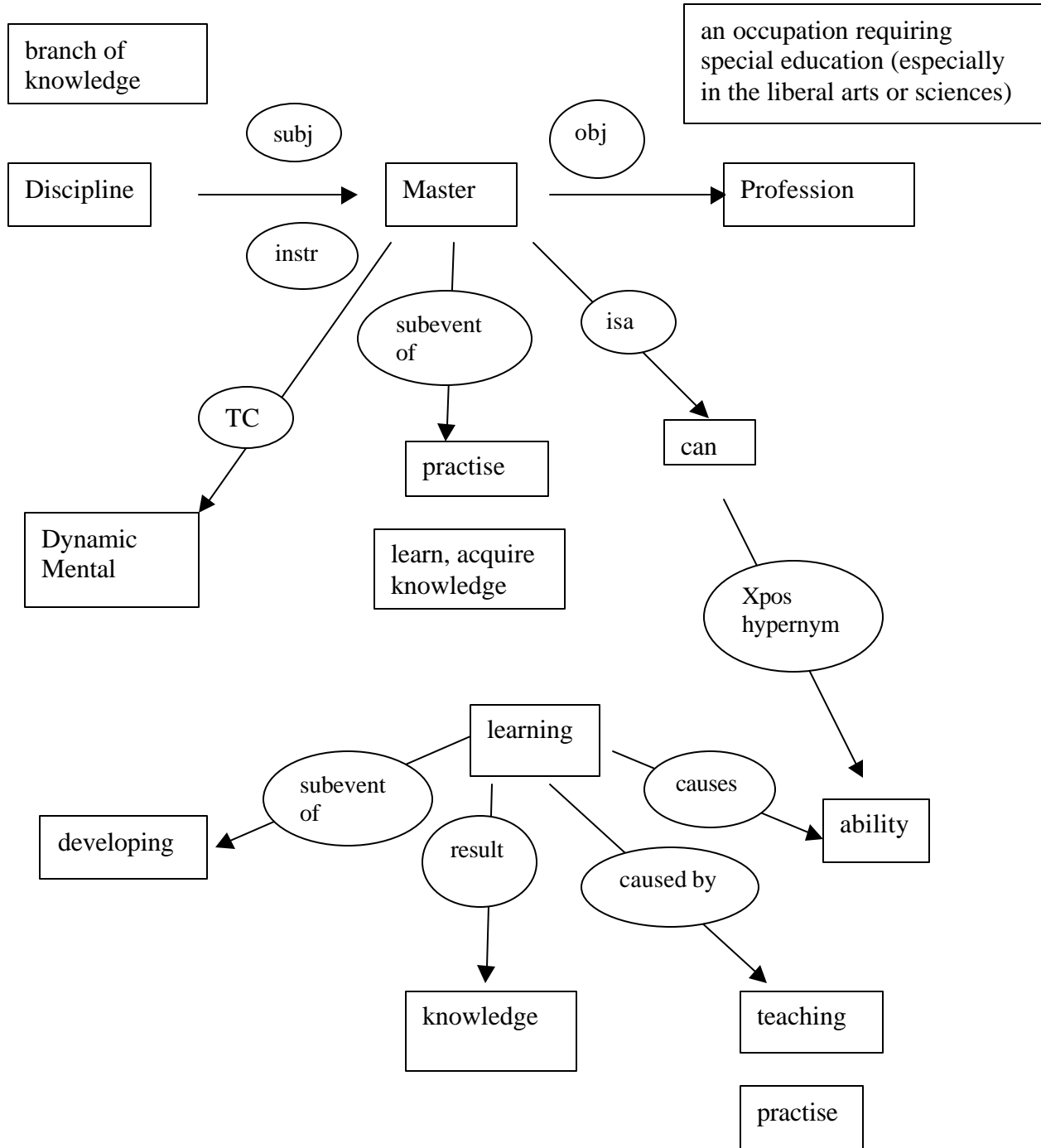


Figure 4: Expanded Ontological Fragment for the pattern **discipline - master - profession**

6 Discussion and conclusion

We have shown that the semi-automatic technique described above for extracting semantic relations between systematically related senses from WordNet glosses is

feasible. There are cases, however, where no relations can be extracted, and where the extracted relations are wrong. Further experimentation with the syntactic properties of the glosses might improve results.

References

Apresjan, J. (1973), *Regular Polysemy*
In: *Linguistics* 142: 5-32

Fellbaum, Christiane (ed.) (1998)
WordNet: An Electronic Lexical Database.
Cambridge, Mass.: MIT Press.

Fillmore, C (1977)
Scenes and frames semantics.
In: Zampolli, A (ed.) *Linguistic structures processing*.
Amsterdam: Benjamins, 55-81.

Harabagiu, S. and Maiorano, S. (2000), *Acquisition of Linguistic Patterns for Knowledge-based Information Extraction*
LREC 2000, Athens

Peters, W. and Peters, I. (2000), *Lexicalised Systematic Polysemy in WordNet*
In *Proc. Second Intl Conf on Language Resources and Evaluation*
Athens, Greece

Peters, W. and Wilks, Y. (2001), *Distribution-oriented Extension of WordNet's Ontological Framework*,
Proceedings RANLP2001, Tzgov Chark, Bulgaria

Peters, W., Guthrie, L. and Wilks, Y. (2002), *Cross-linguistic Discovery of Semantic Regularity*,
Proceedings Global WordNet Association, Mysore, India

Radden, G. and Kövecses (1999), *Towards a Theory of metonymy*
In: Panther, K.U. and Radden, G. (eds.) *Metonymy in language and Thought*.
John Benjamins, Amsterdam

Rodriquez, H., Climent, S., Vossen, P. Bloksma, L., Roventini, A.,
Bertagna, F., Alonge, A., Peters, W. (1998), *The Top-Down Strategy for Building EuroWordNet: Vocabulary Coverage, Base Concepts and Top Ontology*.
In: Nancy Ide, Daniel Greenstein, Piek Vossen (eds), *Special Issue on EuroWordNet. Computers and the Humanities*, Volume 32, Nos. 2-3 1998. 117-152.

Vossen, P. (1998), *Introduction to EuroWordNet*.
In: Nancy Ide, N., Greenstein, D. and Vossen, P. (eds), *Special Issue on EuroWordNet. Computers and the Humanities*, Volume 32, Nos. 2-3 1998. 73-89.