ISA-overload in NorNet

Julie Matilde Torjusen

Abstract

NorNet is a wordnet constructed on the basis of a traditional monolingual dictionary, still undergoing development. The wordnet does for instance still contain many cases of ISA-overload. In this paper I will show examples of ISA-overload in the semantic fields of persons and animals, and see if the relation of paranymy (Huang, Hsiao et al. 2008) or orthogonal hyponymy (Pedersen et al. 2009) are possible ways of solving the ISA-overload from these examples.

1 Introduction

In this paper I address the problem of ISA-overload in a wordnet constructed on the basis of a traditional monolingual dictionary. The wordnet is called NorNet (see Fjeld and Nygaard 2009), and is based upon the dictionary Bokmålsordboka (Wangensteen 2005). The hierarchical structure in NorNet is inherited from Bokmålsordboka. This leads to certain hierarchical problems, particularly ISA-overload. ISA is another word for hyponymy, 'the lexical relation corresponding to the inclusion of one class in another' (Cruse 1986:88). I return to the ISA-relation and ISA-overload in section 2. In section 3 I show some examples of ISA-overload from NorNet, and in section 4 I present the paranymy relation from Huang et al. (2008), as well as the orthogonal hyponymy solution used in DanNet (Pedersen et al. 2009). Both of these relations may be possible solutions for solving the problem of ISA-overload in NorNet.

2 ISA-overload

The ISA-relation is, together with synonymy, one of the most important relations in wordnets (Miller 1998:24), because it places the lemmas in a hierarchical structure. However, Huang et al. (2008:220) find that the way of defining hyponomy in wordnets (with a simple "is a" relation, as in "a car is a vehicle") is inadequate to deal with the complex conceptual relations between co-hyponyms. When unequal hyponyms are placed as co-hyponyms, the result is ISA-overload (Pedersen et al. 2009:277). For instance would *christmas tree* and *pine* as hyponyms of *tree* constitute ISA-overload, because *christmas tree* and *pine* are subordinates to *tree* in different ways. Unequal hyponyms are hyponyms of different kinds, following Cruse (1986) and his framework with 'simple' or less restrictive hyponymy and taxonymy. Taxonomy is a strict form of hyponymy, where a word needs to fit the sentence "an X is a kind of / type of Y" in order to qualify as a taxonym. If the word does not fit this sentence, but fits the sentence "an X is a Y", the word is a hyponym, but not a taxonym. Therefore, *christmas tree* is not a taxonym to *tree*, but it is a hyponym. *Pine* on the other hand is a taxonym of *tree*. A less restrictive hyponym and a taxonym are unequal hyponyms. One of the reasons for this is that all co-taxonyms are incompatible. Both *pine* and *oak* are taxonyms of *tree*, and it is impossible for a *pine* to also be an *oak*. It is a common assumption that co-hyponyms should be incompatible (Cruse 1986:136), but this is not the case with the less

restrictive kind of hyponymy. A *christmas tree* could be a *pine* or a *fir*, so these co-hyponyms are not incompatible.

The problem of ISA-overload in wordnets is closely connected to the less restrictive type of hyponomy. However, the two phenomena are not the same. ISA-overload is a consequence when hyponyms and taxonyms are put side by side in wordnets. The less restrictive kind of hyponyms do not make sound taxonomies, and the problem becomes especially visible when a hyponym is placed as a co-hyponym with a taxonym. I will return to this issue in section 3. ISA-overload shows that the relationship between hyperonym and hyponym is not a straight forward relation, and that problems arise when this relationship is not treated with extra care in wordnets.

3 ISA-overload in NorNet

We find ISA-overload in all areas of NorNet, also in such areas as terms for animals and plants, even though these semantic fields are seen as relatively easy to construct taxonomies of. Still, there are several instances of hierarchies of plants and animals with no ISA-overload, like *dyr > virveldyr > fisk > torsk > møretorsk* (animal > vertebrate > fish > cod > cod from Møre). However, not all of the animal hierarchies are this sound. There are also examples of the less restrictive kind of hyponymy in the animals of NorNet, which do not construct sound taxonomies. An example here is *bjørn* (bear). *Bjørn* has nine hyponyms. Some of these are taxonyms (*alaskabjørn* (American black bear), *brunbjørn* (brown bear), *grizzlybjørn* (grizzly bear), *gråbjørn* (grizzly bear), and *isbjørn* (polar bear)), but there are also hyponyms of the less restrictive kind (*grasbjørn* (bear that does not eat meat), *mannbjørn* (dangerous bear), *skadebjørn* (bear that kills cattle), and *slagbjørn* (bear that kills cattle)), and these are not incompatible. Such a variety creates ISA-overload and poses a problem in NorNet, as the taxonyms and the less restrictive hyponyms have the same hyperonym. The fact that we find ISA-overload in the field of animals is perhaps not that surprising after all, as animals plays many different roles in human life. It is for instance not surprising that it is useful to separate dangerous bears (like *skadebjørn* and *slagbjørn*) from non-dangerous bears (like *grasbjørn*), and not just divide bears into different subspecies.

The example of *bjørn* shows a case of ISA-overload among hyponyms. In NorNet we also find instances of ISA-overload among hyperonyms. That is the case with *erteblomst* (pea flower). *Erteblomst* have several potential hyperonyms in the material from Bokmålsordboka, such as *hageblomst* (garden flower), *klatreplante* (climbing flower), *kulturplante* (domesticated plant), *prydplante* (ornamental plant), *slyngplante* (liana) and *plante* (plant). The first five hyperonyms are of the less restrictive kind, in contrast with *plante*, which is taxonomic.

4 Possible solutions

One possible solution of ISA-overload in wordnets is presented in Huang et al. (2008), namely the relation of *paranymy*. They define paranymy as a relation between a set of coordinate terms that share the same classificatory criteria, and illustrate the relation with the difference between *spring/summer/autumn/winter* and *dry season/rainy season*, which all have *season* as their hyperonym. This is an example of ISA-overload, since the hyponyms belong to different semantic fields. Their solution is to say that *spring* is a paranym of *summer*, *autumn*, and *winter*, and that *dry season* is a paranym of *rainy season*. The paranymy

relation could reduce the ISA-overload in NorNet as well. In the *bear*-example *grasbjørn*, *mannbjørn*, *skadebjørn*, and *slagbjørn* would be paranyms, and thus differentiated from *alaskabjørn*, *brunbjørn*, *grizzlybjørn*, *gråbjørn*, and *isbjørn* (see figure 1). This is because they have the same classificatory criterion. For the users of the wordnet, both humans and computers, this would lead to an understanding that an *isbjørn* is a different kind of hyponym to *bjørn* than *grasbjørn*.
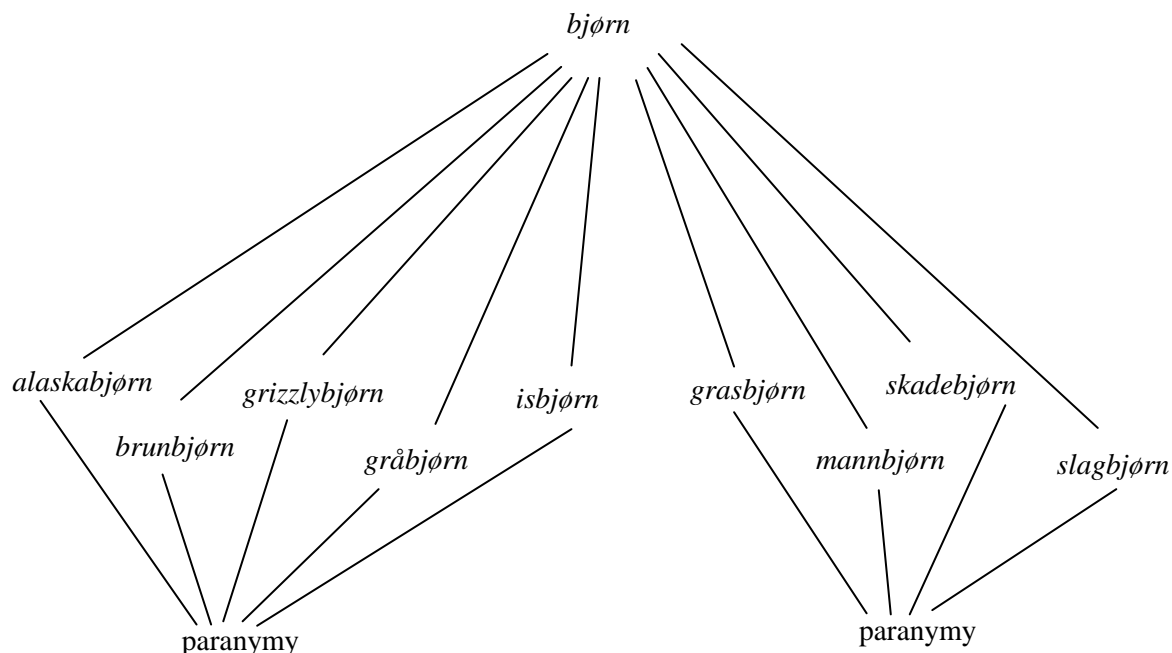

Figure 1. Hyponyms of *bjørn* differentiated with the paranymy relation.

However, paranymy does not seem to be a good solution for cases like *erteblomst*. It is difficult to find a classificatory criteria for the hyperonyms *hageblomst*, *klatreplante*, *kulturplante*, *prydplante*, *slyngplante* and *plante*, seeing as the difference is biggest between the taxonym (*plante*) and the non-taxonyms (see figure 2).
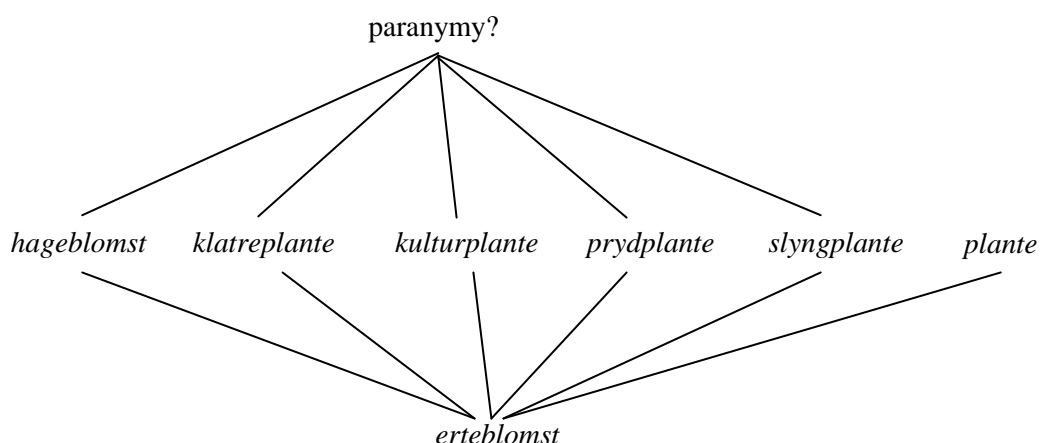

Figure 2. Hyperonyms of *erteblomst* grouped with the paranymy relation. Paranymy is marked with a question mark, as it is uncertain if the relation between *hageblomst*, *klatreplante* and so on can be called paranymic.

Pedersen et al. (2009) mentions paranymy as a solution, but end up choosing a different approach. In DanNet the hyponyms of the less restrictive kind are separated from the taxonomic hyponymy by means of the feature *ortho*. This feature is seen as a specification of the hyponym, for instance a *fidusmaleri* (pseudo-art) might be both an *akvarel* (water colour) or an *oliemaleri* (oil painting) (Pedersen et al. 2009:280). *Fidusmaleri* is therefore a specification of the hyponym *maleri* (painting). This solution works well in cases like *erteblomst*, where a hyponym has several hyperonyms of different kind (see figure 3). *Hageblomst*, *klatreplante* and the other orthogonal hyponyms can be seen as a spesification of *erteblomst*, and *plante* is the only regular hyponym. With this solution, there is no ISA-overload.
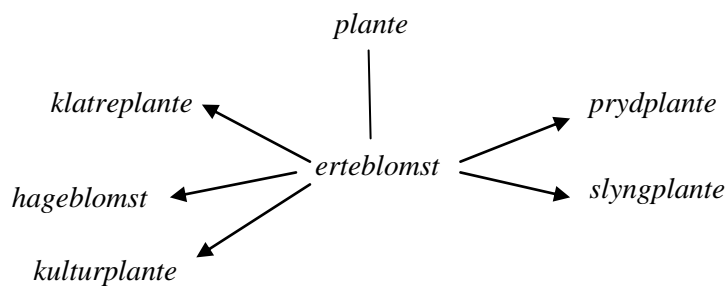
plante

klatreplante

prydplante

erteblomst

hageblomst

slyngplante

kulturplante

Figure 3. Hiearchy of *erteblomst* with normal and orthogonal hyponymy. Orthogonal hyponymy is marked with arrows.

On the other hand, this solution is not suitable in cases like *bjørn*, seeing as the less restrictive hyponyms of *bjørn* (*grasbjørn*, *mannbjørn*, *skadebjørn*, and *slagbjørn*) can't all be orthogonal hyponyms to *bjørn* at the same time. For instance, a bear can not both eat meat and not eat meat at once (see figure 4).
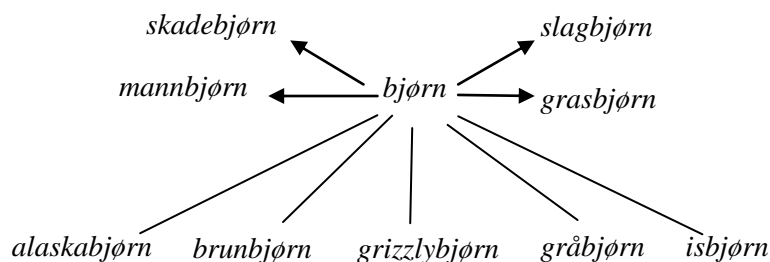
skadebjørn

slagbjørn

mannbjørn

bjørn

grasbjørn

alaskabjørn   *brunbjørn*   *grizzlybjørn*   *gråbjørn*   *isbjørn*

Figure 4. Hiearchy of *bjørn* with normal and orthogonal hyponymy. Orthogonal hyponymy is marked with arrows.

Where there are ISA-overload among the hyperonyms, orthogonal hyponymy is a good solution. With the example of *erteblomst* you could say that all the different hyperonyms shows different qualities of a pea flower, something you can't say in the case of *bjørn*.

5. Conclusion

Both paranymy and orthogonal hyponymy have their strengths and weaknesses. Paranymy is a good solution where a hyperonym have different kinds of hyponyms, and orthogonal hyponymy seems to be a good solution i cases where a hyponym have several hyperonyms. It seems as both paranymy and orthogonal hyponymy may be possible solutions to the ISA-overload in NorNet, even though there are problems with both of the strategies. However, no solution has yet been chosen for NorNet.

References

A. Dictionaries
Wangensteen, B. 2005. *Bokmålsordboka: definisjons- og rettskrivningsordbok*. Oslo: Kunnskapsforlaget.

B. Other literature
Cruse, D. A. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.
Fjeld, R. V. and L. Nygaard 2009. 'NorNet - a monolingual wordnet of modern Norwegian.' In B. S. Pedersen, A. Braasch, S. Nimb and R. V. Fjeld (eds.), *Proceedings of the NODALIDA 2009 workshop - WordNets and other Lexical Resources*. Odense: NEALT, 13–16.
Huang, C.-R., P.-Y. Hsiao, I.-L. Su, and X.-L. Ke 2008. 'Paranymy: Enriching Ontological Knowlegde in WordNets.' In *Proceedings from the Fourth Global WordNet Conference*. Szeged, Ungarn, 221–228.
Miller, G. A. 1998. 'Nouns in WordNet.' In C. Fellbaum (ed.), *WordNet: an electronic lexical database*. Cambridge, Massachusetts: MIT Press.
Pedersen, B. S., S. Nimb, J. Asmussen, N. H. Sørensen, L. Trap-Jensen, and H. Lorentzen 2009. 'DanNet: The Challenge of compiling a wordnet for Danish by reusing a monolingual dictionary.' *Language Resources and Evaluation* 43.3: 269–299.