Enriching a wordnet from a thesaurus

Sanni Nimb[1], Bolette S. Pedersen[2], Anna Braasch[2], Nicolai H. Sørensen[1], Thomas Troelsgård[1]

(1) Society for Danish Language and Literature, Denmark
(2) University of Copenhagen, Denmark

sn@dsl.dk, bspedersen@hum.ku.dk, braasch@hum.ku.dk, nhs@dsl.dk, tt@dsl.dk

ABSTRACT

Wordnets are traditionally built around synonym sets with the vertical hyponymy relations as the central structuring principle. The hyponymy relation, however, does not necessarily group concepts into synsets that are particularly close from a thematic or functional point of view, a phenomenon which is sometimes referred to as the "ISA overload", or if contemplated from a thematic view point: the "tennis problem". In this paper we present two experiments. The first one concerns a method for remedying these problems by transferring thematic information from a thesaurus to a wordnet (Danish Thesaurus to DanNet). Hereby we can automatically subdivide co-hyponyms thematically as well as relate synsets thematically across parts of speech. Since the thesaurus is not yet fully completed, the paper describes work in progress; nevertheless, with an error rate below 5% of the most coarse-grained transferred themes, the experiment appears to be very promising. Finally, the second experiment concerns extension of DanNet via the Danish Thesaurus: The thematic organisation of the thesaurus in near synonyms is further applied as a very precise method for automatically extending the lexical coverage of DanNet.

KEYWORDS: Wordnet, " tennis problem", ISA overload, thesaurus, thematic information.

Sanni Nimb, Bolette S. Pedersen, Anna Braasch, Nicolai H. Sørensen and Thomas Troelsgård 2013. Enriching a wordnet from a thesaurus. *Proceedings of the workshop on lexical semantic resources for NLP at NODALIDA 2013.* NEALT Proceedings Series 19 / Linköping Electronic Conference Proceedings 88: 36–50.

1 Wordnets, ISA overload and the "tennis problem"

Wordnets (Fellbaum 1998; Vossen 1998) are traditionally built around synonym sets with the vertical hyponymy relations as the central structuring principle. This paradigmatic structure is further supplemented by a set of horizontal relations such as antonymy and meronymy. Applying the hyponymy relation as the skeleton for word taxonomies is indeed very convenient, in particular in relation to computational applications since it first of all facilitates the strong inference mechanism of inheritance. However, wordnets as they generally stand appear to lack crucial relations among concepts if they are to be used efficiently as knowledge bases for language technology applications. These include in particular applications requiring some level of "deep" understanding such as information retrieval, question answering, text navigation and text mining.

So, even if hyponymy may include some very basic aspects of the way we organize and conceive concepts in our mental lexicon, and even if this structuring principle is convenient for computers because of its inheritance properties, it is far from sufficient to account for the central relatedness between concepts. First of all, hyponymy does not necessarily results into groups of concepts that are particularly close from a thematic or functional point of view, a phenomenon which is sometimes referred to as the "tennis problem" (cf. Fellbaum 1998, Sampson 2000), pertaining the fact that wordnets traditionally do not account for the relatedness of concepts such as tennis, ball, racquet and net.

Seen from the taxonomical perspective, this lack of expressivity relates to the so-called ISA overload, i.e. the situation where sets of unequal hyponyms are grouped as simple sister terms under the same superordinate, cf. among others Guarino (1998), Guarino & Welty (2002), Huang et al. (2008). To illustrate the problem, consider in the Danish wordnet, DanNet (cf. Pedersen et al. 2009), the hyponyms for concepts like stang ('bar', 'stick') and maske ('mask'). Stang subsumes heterogeneous sets of hyponyms like candy bars, slate pencils, candles, and rods on mens' bicicles, where maske refers to hyponyms like a mask for dressing up for a carnival, diving masks, smoke masks and facial treatments. Thus, the hyponyms subsumed by these synsets may share some very general dimension of form or functionality (i.e. covering the face), but they belong to all sorts of domains and would, in a thesaurus, basically be categorized in a completely different way. Some belong to the food domain, some to entertainment, and others to different professions. In some cases an additional hypernym which clearly refers to the domain is given to these concepts; indicating for instance that a candy bar is also a kind of candy which is again a kind of food, but this is not always possible unless you want to introduce artificial concepts.

In this paper we present two experiments of automatic information transfer from a thesaurus to DanNet: one concerned with thematic information transfer in order to remedy the problem sketched out above (Section 3.1 and 3.2) and one experiment concerned with an extension of the number of synsets on the basis of near synonyms encoded in the thesaurus (Section 3.3).

2 Related work

The suitability of wordnets in intelligent language technology applications has been examined with shifting intensity during the last two decades. In the nineties the Text Retrieval Conferences (TREC) gave rise to a series of thorough testing of Princeton WordNet (PWN) in information retrieval (Voerhees 1993, Voerhees 1994, Voerhees & Harman 1997, Mandala et al. 1998; Gonzalo et al. 1998) without, however, showing radical improvement of system performance. In

2007 the EU project KYOTO (Knowledge-Yielding Ontologies for Transition-Based Organization) was launched as an ambitious, multilingual testing of the wordnet framework meant for mining, structuring and distributing knowledge across languages (Vossen et al. 2008). The project signalled a renewed interest in the use of wordnets for advanced language technology applications such as text mining, question answering, and text retrieval.

The idea of extending standard wordnets with supplementary relations is well-known, see for instance Fellbaum & Miller (2006) for psycholinguistic experiments on associative relations or Veale & Hao's work on folk knowledge in wordnets (2008). The same goes for employing semi-automatic expansion methods from other resources. For instance, in languages with rich productive morphological derivation (such as the Slavic languages), several experiments have been performed in order to semi-automatically capture such morphological relatedness across word classes, as seen in the Czech Wordnet (Pala & Hlaváčková 2007), the Polish WordNet (Piasecki et al. 2010) but also in the Turkish WordNet (Bilgin et al. 2004). Further, the Polish WordNet 2.0 has been enriched with information about verb sub-categorization and semantic classification of aspectual verb pairs. Likewise, innovations in the Hungarian WordNet (Kuti et al. 2008) comprise both language independent and language dependent expansions to the wordnet for verbs and adjectives, and in the Portuguese WordNet (Amaro et al. 2010) an explicit description of argument and event structure is given.

Other wordnets increase their number of relations by inheriting them from the wordnets they link to. The Arabic WordNet Project (Black et al. 2006) which uses the base concept sets of BalkaNet and EuroWordNet as the starting point, obtained a number of semantic relations expressed in SUMO for the English synsets simply by transferring the links from Princeton WordNet to Arabic WordNet.

Similar to the domain-related methods that we are proposing here, are Montoyo et al. (2001) who describe a method to enrich PWN with domain information, arguing that such information provides a natural way to establish semantic relations among synsets. On this approach, wordnet senses are automatically identified in files containing arranged information within a classification system, and the domain information from the file is assigned to the synset in PWN. In addition, earlier work by Navigli and Velardi focuses on relations for domain concepts in the framework of OntoLearn (Navigli & Velardi 2002; Navigli et al. 2004). In SentiWordNet, PrincetonWordNet is semi-automatically extended with sentiment information expressed as polarity values (Baccianella et al. 2010). Further, Veale & Moueddeb (2010) exploit lexical distribution patterns in corpora and semantic similarity scores extracted from WordNet in order to gain more semantic knowledge.

Finally should be mentioned three approaches for extending wordnets on the basis of encyclopedic information. Ruiz-Casado et al. (2005) enrich PWN with encyclopedic definitions from a (rather small) online encyclopedia. In Veale (2006) and Veale & Butnariu (2010) an automated system to extend the number of synsets in PWN is described, building on the extraction and morphological analysis of new words in Wikipedia texts as well as on semantic knowledge from the same text. Finally, Navigli & Ponzetto (2010) describe a similar approach to produce a large, wide-coverage multilingual semantic network. In BabelNet, concepts and relations are automatically extracted from PWN and Wikipedia. The approach involves automatic mapping of Wikipedia pages to PWN synsets after a disambiguation process of candidates from both knowledge sources.

3 Transferring information from a Danish thesaurus to the Danish wordnet, DanNet

3.1 Transferring thematic information from three different levels

Since both thesauri and wordnets arrange concept data in a logical relationship, these resources resemble similar semantic properties to a considerable extent. However, they also differ in several ways. Where wordnets use the hyponymy hierarchy as the primary organizing principle as we have seen, often with an ontology-based division of the world at the uppermost level, thesauri tend to operate with less abstract main categories and a larger number of basic thematic units that go across hyponymies and across parts of speech. In other words, where wordnets are basically ordered vertically, thesauri are rather ordered horizontally – or by themes.

In the present experiment, performed by the Society for Danish Language and Literature (developer of DT and co-developer of DanNet) and the University of Copenhagen (co-developer of DanNet) in collaboration, we exploit the fact that a new Danish thesaurus (DT), which is being compiled at the moment (2009-2013) at DSL, shares common sense IDs with DanNet. Actually, both resources are derived from a third resource, Den Danske Ordbog, a medium-sized contemporary monolingual dictionary of Danish (Hjorth & Kristensen 2005). According to the plan, DT will when it is completed by the end of 2013, contain more than 100,000 word senses compared to DanNet's only 65,000 synsets. The senses are grouped according to a set of different types of relations, and formalized information on the group is annotated in a header. Figure 1 below exemplifies the structure of DT with thematic chapters, sections, subsections and clusters.

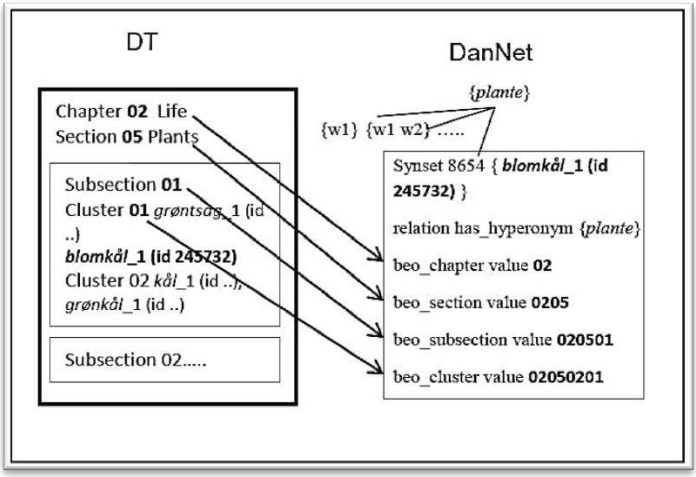
Chapter [section [subsection HEADER [word, cluster[word, word..], cluster[…], word ..] subsection..]]
Chapter2 ***Life*** [section 02.02 ***Plants*** [subsection1 HEADER: has_hyperonym *nytteplante _1* ('utility plant'), concerns *spise_1* ('eat') [cluster1 [*grøntsag_1* ('vegetable'), *blomkål_1* ('cauliflower'), *broccoli_1* ('broccoli'] cluster2 [*kål_1* ('cabbage'), *grønkål_1* ('collard')..] cluster3 [*kornsort_1* ('cereal'), *rug_1* ('rye')..]] [subsection2..]]

Figure 1: Structure of DT exemplified by the chapter 2, Life, the section 02.02 Plants, the subsection of utility plants and clusters of different types of cabbage, cereals etc. (in progress).

As briefly mentioned in the introduction, a well-known problem in the context of hyponymy is the ISA overload where heterogeneous groups of hyponyms are grouped as sisters under the same hypernym. To further illustrate this, person ('person') has more than 6,000 hyponyms in DanNet comprising both persons with inherent characteristics, as well as persons with persistent or temporary roles (Pedersen & Braasch 2009, we avoid artificial hypernyms (in contrast to e.g. PWN which operates with artificial nodes such as 'evil person' etc.). DT also groups a big part of the senses according to a common hyperonym, as the example of utility plants seen in Figure 1. But opposite DanNet, the thesaurus also allows for a flexible placement of concepts in different groups irrespective of whether a common, precise hypernym can be identified or not. In this way, thematic grouping can be provided into e.g. reference to persons having some particular feelings, persons involved in travelling, persons involved in music etc. The domain information in DT is therefore likely to be far more relevant when it comes to a subdivision of the cases of high number of co-hyponyms in DanNet.

The completed thesaurus will consist of 22 chapters divided into approx. 970 sections containing a total of more than 6,000 subsections which contain headers with different types of semantic features. We have carried out the transfer experiment when 1/3 of the thesaurus was completed, using information regarding 2,178 finished subsections (thematic level 3), their corresponding section (thematic level 2) and the chapter to which the section belongs (thematic level 1). Also information on the most detailed semantic level 4, the clusters in DT which typically group near synonyms or synonyms was transferred ; we will return to the use of this data in 3.3.

Since each "synonym" (i.e. lexical representation) in a DanNet synset share the unique id with the relevant sense or senses in DT, the transfer was carried out by assigning the level numbers to the synsets, cf. Figure 2.



Figur 2: The transfer of thematic and semantic information on synset members from DT to their corresponding synset in DanNet.

Four different level numbers (corresponding to chapter, section, subsection and cluster) were assigned the synsets in order to be able to experiment with the data in different ways. Only one member of a synset had to be represented in DT in order for the number values to be assigned to the whole synset. In case of conflicting values the level numbers for that synset were discarded. The transfer resulted in thematic assignment of level numbers to 17,816 synsets.

In the case of large groups of co-hyponyms in DanNet, we assume intuitively that the section level with its 970 divisions (corresponding to 'beo_section value 0205' in Figure 2) is quite informative when it comes to a subdivision. As an example, the completed third of DT contains 15 subsections with co-hyponyms of stang (bar, pole) distributed among 15 different sections. Pløk ('plug', 'peg') and teltstang ('tent pole') appear in the same section as the only ones and are thereby both grouped together and sorted out from the other 13 hyponyms.

In cases where a number of co-hyponyms belong to the same section but to different subsections, the subsection domain numbers (eg. the beo_subsection value 020501 in Figure 2) are the ones that reduce the ISA overload. For example, words denoting 'persons who dislike something or somebody' constitute only one of the two 'person' subsections in the same section ('Dislike') (the other one grouping different words for persons to be disliked). In such cases, the section domain information transferred from DT is not enough to delimit the relevant words in DanNet. A similar example is seen in Figure 3 for food words from the section 'Food and Dishes' in DT.

Figure 3: Hyponyms of food divided in three semantic groups in DT according to different meaning aspects such as **1) 'concerns': main ingredient**: fish dish [pasta dish, spaghetti dish] rice dish, vegetable dish [meat dish, dish of minced meat, venison dish, chicken dish] potato dish, egg dish, **2) 'concerns': consistency and quantity: cream**, goo [**puree**, mash, mousse] [blob, sprinkling], [**drop/spot**, splash] [sip, mouthful, swallow, gulp, bite, morsel] **[slice/cut**, hunk, chunk, lump, section] lump of meat, hunk of bread, [sprinkling of (chopped) parsley, sprinkling of chives, sprinkling of sugar] knob/nut of butter, blop of whipped cream] [section of a lemon, section of an apple] [**serving**, ration, bowl, bowlful, plate, plateful], and **3) 'concerns': time**[breakfast, lunch dish, dinner meal, midnight snack][summer dish, Christmas meal, Easter meal]. Bold words in DT function as keywords.

In all three subsections the hypernym is food, but the subsection information distinguishes between three semantic dimensions of food 1) the major ingredient, 2) the consistency or quantity and 3) the time when it is eaten.

Within the field of terminology a similar method to distinguish between co-hyponyms via semantic dimensions has been introduced (Madsen & Thomsen 2009; Madsen et al. 2004), but contrary to this method, the unique semantic criteria which connects the words of a certain group (e.g. concerns konsistens (consistency) in Figure 3) is not always made explicit in DT.

Furthermore, some subsection divisions are established in DT from purely thematic reasons (and marked as such in the header) when precise semantic relations, such as for example a common hypernym, are impossible to assign to the group. Actually, 12 % of the subsections in DT consists of words grouped together just for thematic reasons. E.g. in the section 1.2 Himmellegemer ('heavenly bodies') all words concerning the sun (korona ('corona'), solvind ('solar wind'), solbane ('path of the sun') etc.) are grouped together in one subsection of this type and simply assigned the relation concerns sol ('sun') in the header. Likewise, words concerning golf (golfklub ('golf club'), golfbane ('golf course'), par (par) etc.) are grouped together in the

same type of subsection and assigned concerns golf ('golf') in the header. Along the same line, Figure 4 shows examples of boxing terms from different part of speech in the same subsection.

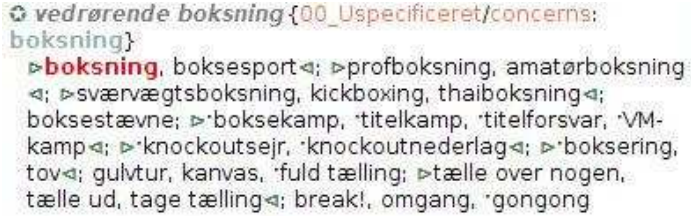


Figure 4: Boxing words in DT, such as 'boxing', professional boxing' 'amateur boxing', 'take the count', 'ring' etc.

The transferred data on subsection number for the words in these groups to their corresponding synsets in DanNet may constitute suitable answers to the tennis problem. However, in some cases it is probably more convenient to apply the more coarse-grained thematic groupings (section and chapter). To illustrate the classical group of concepts mentioned above, tennis, raquet, net, and ball, these belong to different subsections of 'Raquet sports' (badminton, tennis and squash) and are only related via the more broad section division.

To conclude, there is no doubt of the fact that the relevant information is indeed present in DT, but it should be investigated further and at a larger scale which detail of thematic information proves to be most appropriate to the majority of the cases.

3.2 Assessment of the coarse-grained divisions

At the current stage of work in progress, only the coarse-grained thematic groupings at chapter level have been systematically assessed. Two percent of these assignments were manually judged in order to see to which extent the subdivision of co-hyponyms on the basis of this raw material made sense. The result was that 4,6 % of the assigned themes were not considered fully intuitive. Among these, several, however, made sense when considering DT in more detail. For instance, all wooden materials were assigned the theme "Equipments, technology", simply because they were described as materials for producing furniture and buildings.

The transferred data on subsection number for the words in these groups to their corresponding synsets in DanNet may constitute suitable answers to the tennis problem.

All in all, the experiment showed a considerable enrichment of the data in spite of its coarse-grainedness. In Figure 5 is shown how some of the many direct hyponyms in DanNet of the synset {egenskab_1, beskaffenhed_1} ('property') are sorted in intuitively meaningful groups based on chapter information in DT. In Nimb & Pedersen (2012) we discuss in detail how information on properties in DT are transferred and used in the form of a new semantic relation in DanNet.

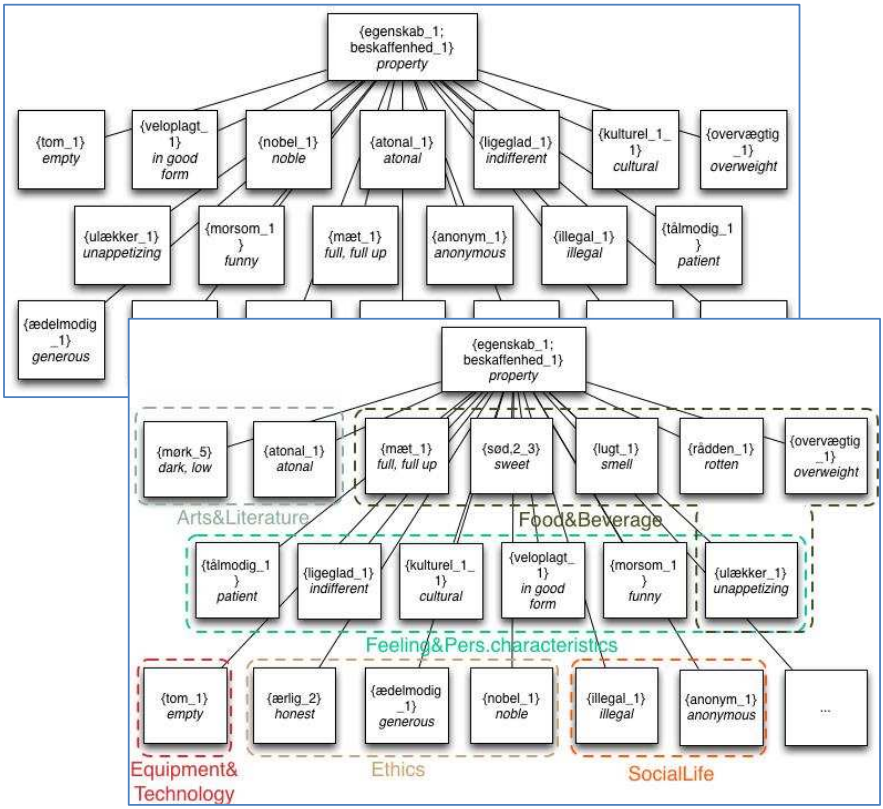


Figure 5: Hyponyms of the synset {egenskab} ('property') in DanNet, above presented without DT values, below sorted out in thematic groups based on the assigned DT chapter values.

Also Figure 6 shows some of the thematic groupings of co-hyponyms that appeared after the coarse-grained assignment based on chapter values in DT, in this case under persons, masks and bars.

Persons	Masks	Sticks/bars
Sports&leisure	**Equipment&technology**	**Food&beverage**
gæst_1 (guest)	*maske_1_1* (mask)	*stang_2* (bar)
bryllupsgæst_1 (wedding guest)	**Life**	*chokoladebar_1* (chocolate bar)
middagsgæst_1 (dinner guest)	*muddermaske_1* (mud mask)	*ostebjælke_1* (cheese stick)
sportspige_1 (sportsgirl)	*ansigtsmaske_1* (face mask)	**Location&change**
lilleput_1 (pre-teen)	**Sports&leasure**	*stang_1* (pole)
gymnastiklærer_1 (gym master)	*dykkermaske_1* (diving mask)	*bom_1* (balance beam)
Feelings&pers.characteristics	*fægtemaske_1* (fence mask)	*lossebom_1* (derrick)
ønskebarn_1 (planned child)	*fastelavnsmaske_1* (halloween mask)	*jernbanebom_1* (railway barrier)
kælledægge_1 (darling)		
møgunge_1 (kiddie)		
øjesten_1 (apple of ones eye)		
heltinde_1 (heroine)		
bøddel_1 (tormentor)		
Art&culture		
balletbarn_1 (ballet child)		
geisha_1 (geisha)		
mavedanser_1 (belly dancer)		
sanglærer_1 (song teacher)		
musikforsker_1 (musicologist)		
kubist_1 (cubist)		
skjald_1 (scald)		
Location&change		
bærer_1 (carrier)		
rumforsker_1 (space researcher)		
Equipment&technology		
håndværker_1 (workman)		
tømrer_1 (carpenter)		
murer_1 (brick layer)		
bygmester_1 (master builder)		
saddelmager_1 (saddler)		
arkitekt_1 (architect)		
Life		
skabsbøsse_1 (closet queen)		
elsker_1 (lover)		
muskelmand_1 (muscleman)		

Figure 6 : Three examples of automatic thematic groupings of synsets in DanNet with identical hyponyms (persons, masks and sticks), based on transferred chapter values from DT.

From the opposite perspective, Figure 7 exemplifies how the coarse-grained chapter level information now relates concepts that otherwise are unrelated in DanNet like instruments used for cooking, containers for containing food, food itself, properties of food, eating and cooking events as well as persons involved in eating.

Food&Beverage
bestik_1 (cutlery)
ravioli_1 (ravioli)
tærteform_1 (baking tin)
slikmund_1 (sweet-tooth)
slubre_1 (to slurp)
gennemstegning_1 (cooking to be well-done)

Figure 7: Some of the **t**erms in DanNet which have been related via the Food&Beverage theme information from DT

3.3 Automatic compilation of new synsets in DanNet based on DT

Based on the most fine-grained thematic level in DT, where we find clusters of near synonyms, we have further experimented with the automatic compilation of new synsets in DanNet. A well-known corpus-based method for extending the coverage of a lexical-semantic resource is to examine syntactic patterns such as enumerative noun phrases and look for unknown words in the phrases. In such investigations, the semantics of new words is guessed upon with some accuracy based on the information from the already known words in the phrase (see for instance Kokkinakis et al. 2000). Our approach is conceptually similar to this method; however, we base our compilation not on enumerations in a corpus, but on the more precise near synonyms given in DT. If a new word is listed in a cluster in DT where at least two DanNet synsets are already represented, then we consider the new word to be of the same synset type (i.e. with same ontological type and the same hyperonym) given that the two known synsets have identical hypernyms and identical ontological types.

The current experiment results in 440 new synsets, an excerpt of which can be seen in Figure 8. An assessment of 10 % of these indicates that the method is indeed very precise since all the evaluated synsets were assigned a correct hypernym as well as a correct ontological type. All in all we estimate that we can automatically generate 1,500 new synsets using this method on the completed DT data. In future, however, we plan to investigate further whether we can extend the DanNet coverage also on the basis of less precise data in DT. Approx. half of the subsections in DT consist of co-hyponyms, and we consider this encoding to be of significant value for automatically generating new synsets. In accordance with the experiment described above, information on existing synsets in the same subsections will be considered. Since DT already contains more than 20,000 concepts not represented in DanNet, the amount of potentially easy-accessible material for extending DanNet is considerable.

Near synonyms	New synset(s)
espresso_1 *café au lait_1,* *cappuccino_1*	*caffé latte_1*
es_1_1 (ace) *toer_3* (two/deuce) *treer_2* (three) *firer_3* (four) *femmer_2*(five) *sekser_1* (six)	*syver_1* (seven) *otter_2* (eight) *nier_1* (nine) *tier_2* (ten) *joker_1* (joker)
jaloux_1 (jealous) *skinsyg_1* (jealous) *syg_1_1* (compulsive)	*besidderisk_1* (possessive)
kuffert_1 (suitcase) *rejsetaske_1* (travel bag) *rygsæk_1* (backpacker)	*læderkuffert_1* (leather suitcase) *håndkuffert_1* (gripsack)

Figure 8: Examples of new synsets in DanNet based on near synonyms in DT

4 Conclusions

It is not an easy task to define which semantic relations are the crucial ones for automated, intelligent information handling. Each application can be seen as having its own very particular requirements regarding relevant cognitive associations. Put to the extreme, there seems to be infinite dimensions of meaning similarity and infinite ways in which concepts can relate to each other and it is unattainable to provide a full lexical semantic network which contains all relations of potential relevance.

Nevertheless, in this paper we have argued that classical hyponymy is often underspecified with regard to some very central meaning dimensions such as thematic context and particular use. Also, we have seen that many wordnets, including DanNet, lack important relations across part of speech. Classical thesauri have obvious resemblances with wordnets, but they differ with respect to the criteria used to carve up the conceptual world. For example, as we have shown, thematic relations are in fact well-represented in these resources.

Because of the close connection between the two resources DT and DanNet, both based on the original dictionary, DDO, and both maintaining the same sense IDs, transfer from one to the other is in fact technically feasible and profitable as our experiments have indicated. The fact that all word clusters in DT are XML tagged with one or more semantic relation types makes transfer directly practicable, especially in the cases where the relations are not currently present in DanNet. In contrast, supplementary, more functionally oriented hyponymy relations based on the DT have to be introduced semi-automatically to prevent clashes with the existing taxonomies.

Our experiment has shown that DanNet can be extended profitably by adding the different thematic levels given in DT in order to be able to distinguish between high numbers of co-hyponyms and thematically to relate concepts across the hierarchy. Further, the demonstrated method of adding new synsets to DanNet on the basis of near synonyms in DT has shown the potential for further research in this area.

References

Amaro, Raquel, Sara Mendes & Palmira Marrafa (2010). Encoding Event and Argument Structures in Wordnets. TSD 2010, LNAI 6231, 21–28. Berlin Heidelberg: Springer-Verlag. DOI:10.1007/978-3-642-15760-8.

Baccianella, Stefano, Andrea Esuli & Fabrizio Sebastiano (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. Proceedings of 7th LREC - Language Resources and Evaluation. Paris: ELRA (European Language Resources Association). http://www.lrec-conf.org/proceedings/lrec2010/index.html.

Bilgin, Orhan, Özlem Cetinoglu & Kemal Oflazer (2004). Building a Wordnet for Turkish. Romanian Journal of Information, Science and Technology, 7 (1-2), 163-172. Bucarest: Editura Academiei Române.

Black, William, Sabri Elkateb, Horacio Rodriguez, Musa Alkhalifa, Piek Vossen, Adam Pease, Christiane Fellbaum (2006). Introducing the Arabic Word Net Project. Petr Sojka, Key-Sun Choi, Chritiane Fellbaum, Piek Vossen (Eds.) Proceedings of the third International WordNet Conference (GWC-06). Brno: Masaryk University. http://NLPweb.kaist.ac.kr/gwc/ pdf2006/74.pdf

Braasch, A. & B.S. Pedersen (2010). Encoding Attitude and Connotation in Wordnets . In: The 14th EURALEX International Congress, Leeuwarden , The Netherlands.

Fellbaum, Christiane (ed) (1998). WordNet – An Electronic Lexical Database. Cambridge, Massachusetts, London, England: The MIT Press.

Fellbaum, Christiane, Georg A. Miller (2006). Whither Wordnets? Zampolli Prize Presentation at LREC 2006, Genova. http://www.lrecconf.org/lrec2006/IMG/pdf/ AZPrize.Christiane%20 Fellbaum%20Presentation.LREC06.pdf.

Fellbaum, Christiane & Piek Vossen (2008). Challenges for a Global WordNet. Online Proceedings of the First International Workshop on Global Interoperability for Language Resources (ICGL 2008), 75-82. Hongkong: City University of Hongkong. http://icgl.ctl.cityu.edu.hk/2008/html/resources/~proceeding_conference.pdf.

Gonzalo, Julio, Felisa Verdejo, Carol Peters & Nicoletta Calzolari (1998). Applying EuroWordNet to Cross-Language Retrieval. Computers and the Humanities. 32 (2/3), 185-207. The Netherlands: Kluwer Academic Publishers.

Guarino, Nicola (1998). Some Ontological Principles for Designing Upper Level Lexical Resources. Proceedings from the First International Conference on Language Resources and Evaluation, 527–534. Granada.

Guarino, Nicola & Chris Welty (2002). Identity and Subsumption. Green, R., Bean, C.A. & Myaeng, S. H. (Eds.), The Semantics of Relationships: An Interdisciplinary Perspective, Information Science and Knowledge Management. Springer Verlag.

Hjorth, Ebba & Kjeld Kristensen (eds.) (2005). Den Danske Ordbog. Copenhagen: Gyldendal & Det Danske Sprog- og Litteraturselskab. Online version: http://ordnet.dk/ddo.

Huang, Chu-Ren., I-Li Su, Pei-Yi Hsiao, Xiu-Ling Ke (2008). Paranymy: Enriching Ontological Knowledge in WordNets. Proceedings of the Fourth Global WordNet Conference, 221–228. Szeged, Hungary: Juhász Press Ltd.

Kokkinakis, Dimitrios, Maria Toporowska Gronostaj, Karin Warmenius (2000). Annotating, Disambiguating & Automatically Extending the Coverage of the Swedish SIMPLE Lexicon. Proceeding LREC 2000, 1397-1403. Paris, France: ELRA

Kuti, Judit, Károly Varasdi, Ágnes Gyarmati, & Péter Vajda (2008). Language Independent and Language Dependent Innovations in the Hungarian WordNet. Proceedings of the Fourth Global WordNet Conference. 254-268. Szeged, Hungary: Juhász Press Ltd.

Madsen, Bodil Nistrup, Hanne Erdman Thomsen, & Carl Vikner (2004). Comparison of Principles Applying to Domain-Specific versus General Ontologies. Ontolex 2004, 90-95. Paris, France: ELRA.

Madsen, Bodil Nistrup & Hanne Erdman Thomsen (2009). Ontologies vs. Classification Systems. Proceedings of the NODALIDA 2009 workshop WordNets and other Lexical Semantic Resources — between Lexical Semantics, Lexicography, Terminology and Formal Ontologies. NEALT Proceedings Series 7, 27-32. Tartu: Northern European Association for Language Technology (NEALT) and Tartu University. http://dspace.utlib.ee/dspace/handle/10062/9840.

Mandala, Rila, Takenobu Tokunaga, & Hozumi Tanaka (1998). The use of WordNet in Information Retrieval. Proceedings of the COLING-ACL workshop on Usage of Wordnet in Natural Language Processing, 31– 37. Montreal, Canada: ACL / Morgan Kaufmann Publishers.

Montoyo, Andrés, Manuel Palomar and German Rigau (2001). Method for WordNet Enrichment using WSD. Matousek, V. ,P. Mautner R. Moucek and Karel Tauser (eds.) Proceeding TSD 2001 Lecture Notes in Computer Science,Volume 2166 , 180-186. Springer.

Navigli, Roberto & Simone Paolo Ponzetto (2010). BabelNet: Building a Very Large Multilingual Semantic Network. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 216-225. Uppsala, Sweden. Association for Computational Linguistics.

Navigli, Roberto & Paola Velardi (2002). Automatic Adaptation of Wordnet to Domains. Proceedings of the Third International Conference on Language Resources and Evaluation (LREC), 1499-1504. Paris, France: ELRA.

Navigli, Roberto, Paola Velardi, Alessandro Cucchiarelli & Francesca Neri (2004). Extending and Enriching WordNet with OntoLearn. Proceedings of The Second Global Wordnet Conference - GWC 2004. Brno: Masaryk University. http://www.dsi.uniroma1.it/~navigli/pubs/GCW_2004_Navigli_al.pdf.

Nimb, S. & B.S. Pedersen (2012). Towards a richer wordnet representation of properties – exploiting semantic and thematic information from thesauri. In: LREC 2012 Proceedings pp. 3452-3456. Istanbul, Turkey.

Pala, Karel & Dana Hlaváčková: Derivational Relations in Czech WordNet (2007). ACL '07 Proceedings of the Workshop on Balto-Slavonic Natural Language Processing: Information Extraction and Enabling Technologies. Stroudsburg, PA, USA: Association for Computational Linguistics. http://portal.acm.org/citation.cfm?id=1567559.

Pedersen, Bolette.S. & Patrizia Paggio (2004). The Danish SIMPLE Lexicon and its Application in Content-based Querying. Nordic Journal of Linguistics 27 (1), 97-127. Cambridge University Press.

Pedersen, Bolette S, Sanni Nimb, Jørg Asmussen, Nicolai Sørensen, Lars Trap-Jensen & Henrik Lorentzen (2009). DanNet: The challenge of compiling a WordNet for Danish by reusing a monolingual dictionary. Language Resources and Evaluation, Computational Linguistics Series 43 (3), 269-299, doi:10.1007/s10579-009-9092-1.

Pedersen, B.S. & A. Braasch (2009). What do we need to know about humans? A view into the DanNet Database. In: K. Jokinen and E. Bick (eds.) Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009. NEALT Proceedings Series, Vol. 4, Odense, Denmark.

Pianta, Emanuele, Luisa Bentivogli & Christian Girard (2002). MultiWordNet – Developing an aligned multilingual database. Proceedings of the First International Conference on Global WordNet, 293-302. Mysore, India.

Piasecki, Maciej, Stanislaw Szpakovicz & Bartosz Broda (2010). Toward plWordNet 2.0. Proceedings of the 5th International Conference on Global Wordnet (GWC2010), 263-270. Mumbai: Narosa Publishers.

Ruiz-Casado, Maria, Enrique Alfonesca & Pablo Castells (2005). Automatic Assignment of Wikipedia Encyclopedic Entries to WordNet Synsets. Piotr S. Szczepaniak, Janusz Kacprzyk, Adam Niewiadomski (Eds.): Advances in Web Intelligence Third International Atlantic Web IntelligenceConference, AWIC 2005, Lodz, Poland, Proceedings. Lecture Notes in Computer Science 3528. Springer

Sampson, Geoffrey (2000). Review of WordNet: An Electronic Lexical Database. In International J. of Lexicography 13.54–9, 2000.

Veale, Tony (2006). Tracking the Lexical Zeitgeist with WordNet and Wikipedia. Proceedings of the 17th European Conference on Artificial Intelligence (ECAI 2006), IOS Press, 56-60. Amsterdam, The Netherlands.

Veale, Tony & Yanfen Hao (2008). Enriching WordNet with Folk Knowledge and Stereotypes. Proceedings of the Fourth Global WordNet Conference, 453-461. Szeged, Hungary: Juhász Press Ltd.

Veale, Tony & Cristina Butnariu (2010). Harvesting and understanding on-line neologisms. Alexander Onysko, Sascha Michel (eds.) Cognitive Perspectives on Word Formation. 399-420. De Gruyter Mouton.

Veale, Tony & Mourad el Mouedden (2010). Similarity, Comparability and Analogy in WordNet: Squaring the Analogical Circle with Mondrian.

Proceedings of the 5th International Conference on Global Wordnet (GWC2010). Mumbai: Narosa Publishers. http://afflatus.ucd.ie/Papers/Mondrian%20GWC%20paper.pdf

Voorhees, E.M. (1993). Using wordnet to disambiguate word senses for text retrieval. Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval, 171-180. New York, NY, USA: ACM.

Voorhees, Ellen M. (1994). Query expansion using lexical-semantic relations. Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, 61-69. New York: Springer-Verlag New York, Inc.

Voorhees, Ellen M. & Donna Harman (1997). Overview of the fifth text retrieval conference (trec-5). Proceedings of the Fifth Text Retrieval Conference, 1-28. NIST Special Publication 500-238. Gaithersburg: NIST. http://trec.nist.gov/pubs/trec5/t5_proceedings.html

Vossen, Piek, Eneko Agirre, Nicoletta Calzolari, Christiane Fellbaum, Shu-Kai Hsieh, Chu-Ren Huang, Hitoshi Isahara, Kyoko Kanzaki, Andrea Marchetti, Monica Monachini, Feririco Neri, Remo Raffaelli, German Rigau, Maurisio Tesconi & Joop CanGent (2008). KYOTO: A System for Mining, Structuring and Distributing Knowledge Across Language and Culture. Proceedings of the Fourth Global WordNet Conference, 474-484. Szeged, Hungary: Juhász Press Ltd.

Vossen, Piek (ed.) (1998). EuroWordNet: A Multilingual Database with Lexical Semantic Networks. Dordrecht: Kluwer Academic Publishers.