

of the remaining job. We have also explicated some constraints particular to Southern Sámi which fill in the gaps in the overall theory. For example, we have seen that there is a constraint which prohibits material of a stem to be extrasyllabic, and how this constraint forces closed syllable shortening in bare consonant final stems. Also, a ban on geminates across foot boundaries and a requirement that syllables have onsets account for the patterns observable with respect to vowel final stems.

References

- Bergsland, Knut. 1994. *Sydsamisk grammatikk*. Karasjok, Norway: Davvi Girji.
- Bergsland, Knut & Lajla Mattson-Magga. 1993. *Áarjelsaemien-daaroen baakoegærja*. Alta, Norway: Idut.
- Blevins, Juliette. 1995. 'The syllable in phonological theory'. In J. Goldsmith (ed.), *The handbook of phonological theory*, 206-44. Oxford: Blackwell.
- Goldsmith, John. 1990. *Autosegmental and metrical phonology*. Oxford: Blackwell.
- Hayes, Bruce. 1989. 'Compensatory lengthening in moraic phonology'. *Linguistic Inquiry* 20, 253-306.
- Hayes, Bruce. 1995. *Metrical stress theory: principles and case studies*. Chicago: Chicago University Press.
- Jaahkenelkien, A. 1993. *Goltelidh jih soptsestidh*. Aarbortesne, Norway: Daasta berteme.
- Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Oxford: Blackwell.
- McCarthy, John & Alan Prince. 1995. 'Prosodic morphology'. In J. Goldsmith (ed.), *The handbook of phonological theory*, 318-66. Oxford: Blackwell.
- Nilsson-Mankok, Erik. 1974. *Svensk-sydsamisk ordlista: Vilhelmina vefsen*. Vilhelmina, Sweden: Independent publisher.
- Rice, Keren. 1992. 'On deriving sonority: A structural account of sonority relationships'. *Phonology* 9, 61-99.
- Vinka, Mikael. 1997. 'Southern Sámi ablaut and the role of the harmony foot', ms. McGill University, Montreal.

Implementing WordNet for Swedish adjectives

Caroline Willners

Introduction

A Swedish version of WordNet was created and around 300 Swedish adjectives, mainly from the semantic field of *strength*, were implemented. This paper is a documentation of the implementation. The purpose of the study was to investigate the possibilities of applying WordNet to Swedish and to illuminate general problems with WordNet as well as specific problems in the handling of adjectives. First, a short overview of WordNet is given, and then the WordNet categorisation of adjectives is reviewed. The section about the implementation gives hands-on knowledge of how to add a new adjectival lexical entry in WordNet. Then a description of the problems encountered and some general remarks follow.

WordNet

WordNet is a lexical reference system designed to reflect the organisation of human memory as well as to be a useful on-line dictionary. It contains approximately 95,600 English nouns, verbs, adjectives, and adverbs which are linked semantically. The system also contains syntactic and morphological information. Apart from the original English version developed at Princeton University, WordNets are being implemented for German, Spanish, and Dutch in the EuroWordNet project (Climent et al. 1996).

The most significant feature of WordNet is the semantic organisation. It supports synonymic, antonymic, hyponymic-hypernymic, and meronymic-holonymic relations. Because of this semantic approach the parts-of-speech categories can naturally be separated.

Data in WordNet are entered in 'lexicographers' source files', with somewhat different formats for the different parts of speech. The source files are then compiled by the program *Grinder* which generates the database that can be accessed by the window-based retrieval software. The

semantic relations are coded in the lexicographers' source files, while morphological rules are hard-coded in Grinder. Words with irregular inflection are listed in exception files. (For a thorough introduction see Miller et al. 1993.)

Antonymy

Antonymy is the basic semantic relation for organising the adjectives in WordNet.

Defining antonymy may seem trivial at first sight, but is not an easy task. Antonymy should not be thought of as the maximum degree of difference in meaning (in opposition to synonymy which stands for identity of meaning), but rather it presupposes that all dimensions but one are similar (Lyons 1977). When we judge two words as being good antonyms, we contrast them on the basis of their similarities. The antonymic pair *stark* 'strong' and *svag* 'weak' is good because the words are similar in all other dimensions but the one representing strength, while *kraftig* 'powerful/sturdy' and *svag* are near antonyms since *kraftig* says something about volume as well. They differ too much in their similarities to qualify as good antonyms.

However, why is it so hard to define antonymy, yet people know that *stark* and *svag* are antonyms? The assumption that the antonymous associative bond is learned from the contexts that the words share in ordinary text and discourse is presupposed by the co-occurrence hypothesis (Charles & Miller 1989). Charles & Miller show in their study that antonyms co-occur more often than near-antonyms.

Another unclarity is whether antonymy stands between word forms or between word meanings. Miller et al. 1992 quite controversially adopts the first solution, namely that antonymy is a relation between word forms. That is, the antonymic relation holds between the two word forms *stark-svag* 'strong-weak'. When antonymy holds for other variants or synonyms, the relation *near-antonym* is used instead. An example of near-antonyms is *kraftig-svag* 'powerful/sturdy-weak'.

Synonymy/similarity

Synonymy is the other semantic relation involved in the organisation of adjectives.

Cruse 1986 reports that "there is, unfortunately, no neat way of characterising synonyms". True synonyms are rarely found if one uses the

strong definition of synonymy attributed to Leibniz (two expressions are synonyms if the substitution of one for the other never changes the truth value of a sentence in which the substitution is made). Miller & Fellbaum 1992 therefore propose a weaker definition which makes synonymy relative to context, namely *semantic similarity* (two expressions are synonymous in a linguistic context C if the substitution of one for the other in C does not alter the truth value). The latter notion entails interchangeability of two words in a given context.

WordNet makes use of the weak definition of synonymy, i.e. semantic similarity.

Nouns in WordNet

This paper focuses on the adjectives implemented in WordNet, but as some adjectives are linked to nouns, a small introduction as to how they are organised is called for.

The basic semantic relation organising the nouns in WordNet is hyponymy. Nouns are organised in semantic hierarchies in such a way that a lexical inheritance system is created. For example *canary* @ → *finch* @ → *passerine* @ → *bird* @ → *vertebrate* @ → *animal*. The '@' marks that the relation is hyponymic. 'Animal' is the top of one such hierarchy and all together there are twenty-five noun hierarchies stored in separate files. The system offers the possibility of distinguishing three different types of meronymic features for the nouns: component-object (e.g. *trunk-tree*), member-collection (e.g. *tree-forest*) and stuff-object (e.g. *aluminium-aeroplane*). Other features such as modification and predication are discussed in Miller et al. 1993 but are not implemented.

Adjectives in WordNet

There are four classes of adjectives in WordNet: descriptive, relational, reference-modifying and colour adjectives. The largest group consists of the descriptive adjectives, the 'typical' adjective which ascribes a value of an attribute to a noun, e.g. *en stark man* 'a strong man'. The descriptive adjectives are grouped around antonymous pairs (e.g. *stark-svag*, 'strong-weak'), quite differently from nouns and verbs which are organised in hierarchies with separate files for each hierarchy. Each adjective in the antonymous pair has sets of synonyms, or so called 'synsets', linked to it.

Relational adjectives mean something like 'of', 'relating/pertaining to', or 'associated with'. Examples from English would be *fraternal* as in *fraternal*

twins, and *dental* as in *dental hygiene*. Swedish often uses compounding for this type of construction, e.g. *enäggstvillingar* and *tandhygien*. Examples of relational adjectives in Swedish are derivations from Greek or Latin nouns such as *oral* 'oral' and *manuell* 'manual'.

Reference-modifying adjectives is a term introduced by Bolinger in 1967. He opposed them to referent-modifying adjectives which in WordNet corresponds to the descriptive adjectives. For example, in the nominal phrase *den förre kungen* 'the former king', *förre* does not modify the referent, but rather its reference. Reference-modifying adjectives can only occur in attributive position and the nouns they modify generally denote a function or a social relation.

Chromatic colour adjectives are treated as a special case in WordNet, but there is no example of any colour terms implemented and I will not discuss them further in this paper.

There is actually a fifth category not documented in Miller et al. 1993, but implemented in WordNet 1.5, namely participles. Participles are verb derivations with adjectival functions. They are listed as adjectives in WordNet, but their close relation to the verb is maintained by linking them to their respective verb root.

Index of familiarity

Each word form in WordNet is associated with an index of familiarity. It is a measure of how common a word is, an attempt to represent the fact that words differ in accessibility as has been shown in, for example, reaction tests such as speedreading.

The familiarity indexes are stored in PolyCount-files, one file for each character in the alphabet. Thus, all words starting with *a* are stored in the file *Polya*, the ones with *b* as an initial in *Polyb*, etc. Each file lists the word together with its part of speech and the familiarity index. An extract from the *Polya*-file follows.

abstrakt	adj	1
accentuerad	adj	1
allvarlig	adj	4
ansenlig	adj	1
atletisk	adj	1
avsevaerd	adj	1

The spelling of *avsevärd*: *avsevaerd* 'considerable' is not a graphic error, but the first example of a transcription of the Swedish special

characters *å ä ö*. More such transcriptions will occur in the examples below and the problem will be dealt with in the section *Special characters* under *Problems implementing WordNet*.

Due to the lack of a large enough tagged corpus, which would give the best familiarity index, the WordNet developers use the correlation between frequency of occurrence and polysemy. The more frequently a word is used, the more different meanings it will have in a dictionary (Zipf 1945). Thus, they count word senses within each syntactic category in an on-line version of Collin's Dictionary of the English Language, assigning the index value of 0 to words not occurring in the lexicon and 1 or more according to the number of senses they find.

The implementation

An attractive feature of WordNet is the modularity which makes it possible to implement one part (of speech) at a time. This implementation covers mainly descriptive adjectives but some relational adjectives were implemented for testing.

A copy of the sourcefiles of the English WordNet was used as a starting point¹. The synonym relations coded are based on *Strömbergs synonymordbok* (1995) and the antonym relations, due to the lack of other resources, on my own intuition. An alternative lexical resource for the synonym relations is Bring's *Svenskt ordförråd ordnat i begreppsklasser* (1962). However, Bring lists not only clear synonyms, but words that belong to a specific class in a very broad definition. For a future larger scale of the Swedish WordNet, Bring's work should be considered, but for this pilot implementation *Strömbergs synonymordbok* will do.

Descriptive adjectives

The descriptive adjectives are represented in a lexicographers' file called *adj.all*. Each entry is divided in two halves, one for each adjective in the antonymic pair in question. Each half is headed by the antonymous pair, followed by pointers to the synsets of the first adjective in the pair. Then follows the synsets which point back to the first adjective. The second half of the entry follows the same syntax, but the order of the antonymic pair is

¹The Swedish WordNet implementation was performed on a SUN work station, thus using the UNIX-version of WordNet which was downloaded from <ftp://clarity.princeton.edu/pub/wordnet/1.3/wn1.5lexsrc.tar.gz>. Compiled versions of WordNet for English can be obtained from the same site as well as retrieval software.

switched around and the pointers to the synsets now belong to the other adjective in the pair. To clarify I will go through an entry step by step.

First I have to define synonym sets, or so called synsets. A *synset* is a list of synonyms. The list is enclosed in curly brackets and the last element of the list is a pointer to the head word of the synset.

```
{kraftig, kraftfull, muskuloes, stark&}
```

Returning now to the total representation which, as already mentioned, is divided in two halves. Each half of the representation is headed by a head synset which starts with the antonymous pair, e.g.

```
[STARK, SVAG, !]
```

where '!' represents the antonymic relation. The antonymous pair is capitalised and is followed by pointers to the synonym sets of the first word.

A pointer is represented by a word followed by a comma and a character denoting a type of relation. Synonym pointers are marked with '&' and the words themselves label the pointers, e.g. *stark&*. The head synset is surrounded by curly brackets.

```
{[STARK, SVAG, !] kraftig,& fyllig,&}
```

At the end of the head synset there is an optional space for explanations which has to be surrounded by parentheses.

```
{[STARK, SVAG, !] kraftig,& fyllig,& (stark fysiskt eller psykiskt)}
```

Then the synonym sets follows, one for each pointer in the head synset. Synonym sets appear in curly brackets and start with the pointer from the head synset, followed by a list of synonyms. The list ends with a reciprocal pointer back to the head word and the option of a bracketed explanation.

```
{[STARK, SVAG, !] kraftig,& fyllig,& (stark fysiskt eller psykiskt)}
{kraftig, kraftfull, muskuloes, stark&}
{fyllig, maettad, skarp, stark& (om smak eller substans)}
```

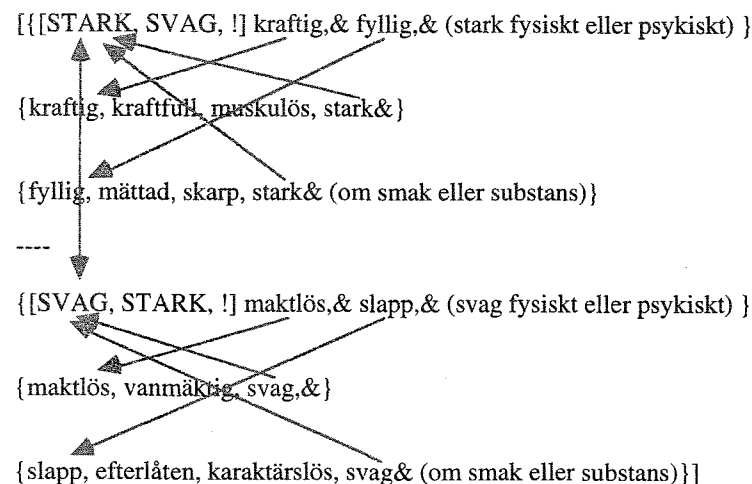
The antonym is coded similarly:

```
{[SVAG, STARK, !] maktloes,& slapp,& (svag fysiskt eller psykiskt)}
{maktloes, vanmaektig, svag,&}
{slapp, efterlaaten, karaktaersloes, svag& (om smak eller substans)}
```

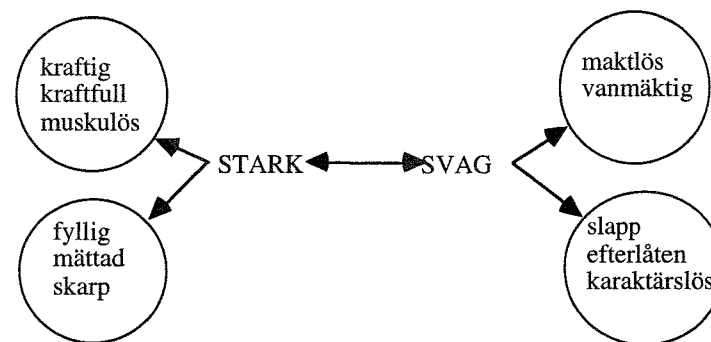
The antonyms' representations are separated from each other by a line of four hyphens, but held together by square brackets surrounding the entire expression.

```
[[[STARK, SVAG, !] kraftig,& fyllig,& (stark fysiskt eller psykiskt)}
{kraftig, kraftfull, muskuloes, stark&}
{fyllig, maettad, skarp, stark& (om smak eller substans)}]
----
[[[SVAG, STARK, !] maktloes,& slapp,& (svag fysiskt eller psykiskt)}
{maktloes, vanmaektig, svag,&}
{slapp, efterlaaten, karaktaersloes, svag& (om smak eller substans)}]]
```

Here follows a visualisation of the pointers encoded in the example above.



An equivalent, more interpretable visualisation would be the following.



Above was a reduced version of the entry for *stark-svag*, the full one is coded as follows.

```
[[[ STARK, SVAG,!] kraftig,& viljestark,& haallbar,& fyllig,&
{kraftig, kraftfull, muskuleos, atletisk, herkulisk, frisk, motstaands-
kraftig, senig, spaenstig, seg, handfast, stark.& (om fysisk styrka)}
{viljestark, karaktersfast, oboejlig, orubblig, okuvlig, stark.& (om
psykisk styrka och uthaallighet)}
{haallbar, stadig, bastant, solid, outslitlig, oemotstaandlig, ointaglig,
saeker, stark,& (om haallbarhet)}
{fyllig, maettad, intensiv, vaaldsam, skarp, fraen, pepprad, stark.& (om
smak och doft)}
----
[[[ SVAG, STARK,!] kraftloes,& maktloes,& slapp,& obetydlig,&
{kraftloes, medtagen, klen, matt, utmattad, nedsatt, vek, braecklig,
krasslig, daalig, skroepelig, darrig, skral, slak, debil, svag,& (saknar
vigoer eller energi)}
{maktloes, vanmaektig, underlaegsen, svag,&}
{slapp, efterlaaten, flat, karaktersloes, haallningsloes, menloes, svag,&}
{obetydlig, liten, laett, ringa, knappt_maerkbar, daempad, besloejad,
laber, foega, svag,&}]
```

It is possible to mark the adjectives according to what syntactic positions they can take, i.e. attributive or predicative placing. This is done with an *a* for attributive and a *p* for predicative, placed within parentheses immediately after the word. None of the adjectives in the field of strength have any such restrictions and are thus left unmarked, but an example from another semantic field would be

{nuvarande(a), foereliggende(a), paagaende}

The database of descriptive adjectives grows rapidly – after entering five antonymous pairs, close to 300 unique adjectives were represented in the database! Of course, many adjectives occur in several different synonym sets. For example *stark* not only heads the structure described above, but also occurs in synsets belonging to *hard* ‘hard’ (vs. *mjuk* ‘soft’), *stor* ‘big’ (vs. *liten* ‘little’) and *kraftig* ‘powerful’ (vs. *klen* ‘feeble’). In such cases of multiple occurrences, they will be listed as different senses, ranked according to the index of familiarity. If the word in question is a member of an antonymous pair, the word and its antonym will be listed, followed by the synsets connected to the word.

stark (vs. svag)

- => kraftig, kraftfull, muskuleos, atletisk, herkulisk, frisk, motstaandskraftig, senig, spaenstig, seg, handfast
- => viljestark, karaktersfast, oboejlig, orubblig, okuvlig
- => haallbar, stadig, bastant, solid, outslitlig, oemotstaandlig, ointaglig, saeker
- => fyllig, maettad, intensiv, vaaldsam, skarp, fraen, pepprad

Senses of the word that are not true antonyms are listed in reverse order, starting with the synset and followed by the heading antonymous pair.

Sense 2

ogenomtraenglig, fast, kompakt, stadig, kraftig, styv, **stark**, motstaandskraftig
=> haard (vs. mjuk)

The full answer given when one asks for the synonyms of *stark* follows.

Similarity of adj **stark**

4 senses of **stark**

Sense 1

stark (vs. svag)

- => kraftig, kraftfull, muskuleos, atletisk, herkulisk, frisk, motstaandskraftig, senig, spaenstig, seg, handfast
- => viljestark, karaktersfast, oboejlig, orubblig, okuvlig
- => haallbar, stadig, bastant, solid, outslitlig, oemotstaandlig, ointaglig, saeker
- => fyllig, maettad, intensiv, vaaldsam, skarp, fraen, pepprad

Sense 2

ogenomtraenglig, fast, kompakt, stadig, kraftig, styv, **stark**, motstaandskraftig
=> haard (vs. mjuk)

Sense 3

storvaext, hoegrest, laang, kraftig, **stark**, muskuloes, grov, fullvuxen
=> stor (vs. liten)

Sense 4

stark, kraftfull, robust, bastant, haardfoer, storvaext, bred, fyllig, yppig
=> kraftig (vs. klen)

This function to map multiple occurrences to different senses is an elegant feature of WordNet. The lexicographer does not have to take other senses of an adjective belonging to a synset into account when entering new data. Grinder resolves the pointers and ambiguous words will automatically be listed with its different senses.

Relational adjectives

Relational adjectives do not have antonyms and can therefore not be incorporated in the same type of database as the descriptive adjectives. What would, for example, be the opposite of *oral* 'oral' or *svensk* 'Swedish'? Their semantic properties are similar to those of nouns used as noun modifiers and thus they are listed in a file with pointers to the corresponding noun instead. The pointer includes the name of the file where the noun is entered which also coincides with the so called beginner (top) of the noun's hierarchy. The relational adjectives are stored in the file *adj.pert* (meaning 'pertaining to'). Each entry starts with the word in question. Then follows a pointer to the word it is related to, including the filename where the noun representation is stored, e.g. the pointer to *mun* 'mouth' in the file *noun.body* would be *noun.body:mun, *. Backslash '\ ' denotes relational adjective. Some examples follow.

```
{[oral, noun.body:mun, \]}
{[svensk, noun.location:Sverige, \]}
{[politisk, noun.act:politik, \]}
```

Since the relational adjectives make demands on the noun part of WordNet (which has not yet been developed for Swedish), only a couple of relational adjectives and their corresponding nouns have been implemented for testing.

Reference-modifying adjectives

Reference-modifying adjectives are treated similarly to the relational adjectives. Most of them occur only in attributive position and they are thus marked accordingly.

```
{foerre(a), foerutvarande(a), foeregaaende(a) }
```

Past participles

Past participles are listed as adjectives as well. The format of the entries is very much like the format for relational adjectives, but the pointer points to a verb instead of a noun. The entry for *sökande* 'searching' would be

```
{[soekande, verb.possession:soeka,<]}
```

No past participles have been implemented so far in the Swedish version of WordNet.

Problems implementing Swedish adjectives in WordNet*Special characters*

The Swedish special characters *å ä ö* are not accepted by Grinder (the utility compiling the lexicographers' files). In the present WordNet-implementation of Swedish adjectives the special characters are coded *å = aa*, *ä = ae*, and *ö = oe*. However, some small adjustments of Grinder would probably make it possible to use the special characters.

Morphology

WordNet has a function for morphological processing, Morphy, which handles morphological transformations. The user can enter an inflected word form into the system and obtain the base form and its senses. Morphy uses two types of processes to derive base forms. It has lists of suffixes and endings with which it deals with purely concatenated inflections. In the case of adjective morphology, the list of suffixes contains inflectional morphemes denoting noun agreement, e.g. *-t* and *-a* as in *starkt* 'strong-NEUT' and *starka* 'strong-PL'. The list of endings contain the root endings. For *stark*, a general rule will do and the elements in the endings list can be left empty. The lists are matched so that a suffix applies to an appropriate root. Words that cannot be inflected using the lists are listed in exception files, one for each part of speech.

The lists of endings and suffixes are hard-coded and thus not easily customised to other languages. One has to open the C-program file *morph.c*, edit the list, and then recompile. The program *morph.c* lies under *source/lib* and the suffixes involved in adjective inflection were listed with their corresponding stem endings in a separate list.

```
/* Adjective suffixes */
"t", "a", "e", "are", "ast", "aste"
```

```
/* Adjective endings */
"", "", "", "", "", ""
```

Note that the number of slots in the list of suffixes must equal the number of slots in the list of endings.

After this adjustment, the Swedish WordNet also accepts the inflected forms of for example *stark*:

stark-t (SG-NEUT)	stark-are (COMPARATIVE)
stark-a (PL)	stark-ast (SUPERLATIVE-INDEF)
stark-e (PL-MASK)	stark-aste (SUPERLATIVE-DEF)

Morphy can also deal with consonant doubling, e.g. *verksam-verksamma* 'active-SG-active-PL'. This is implemented by adding the suffixes with its doubled consonant to the list of suffixes and the corresponding final consonant of the root to the list of endings.

```
/* Adjective suffixes */
"t", "a", "e", "are", "ast", "aste", "ma", "me", "mare", "mast", "maste"
```

```
/* Adjective endings */
"", "", "", "", "", "", "m", "m", "m", "m", "m"
```

But this improvement also has negative consequences. Adjectives with consonant doubling of *m* will be correctly analysed, but there are also adjectives roots ending in *m* which do not double *m* when inflected, e.g. *tam* 'tame' and *lam* 'lame'. The first case, *tam* 'tame', is an example of how words not existing in Swedish can be accepted by WordNet. That is, *tammast*, which does not exist as a word in Swedish will be analysed as *tam* 'tame' and suggested to be an antonym of *vild* 'wild'. The other example *lam* 'lame' illustrates how a word from another part of speech can be incorrectly analysed as an adjective. The verb *lamma* 'to lamb' will be analysed as *lam* 'lame' according to the system above. There is a clear need for a marking system of which morphological pattern an adjective is inflected by. Swedish is rich in inflectional patterns. Consonant doubling is actually governed by phonological rules, but there are no means to express this in WordNet. Hellberg 1978 distinguishes 17 different morphological patterns for Swedish adjectives, and a system for marking the words according to what morphological pattern they are inflected is needed.

Morphological processes that are not purely concatenative, such as deletion and alternation, e.g. *ädel-ädla* 'noble-SG-noble-PL', *högljudd-högljutt* 'loud-SG-loud-PL' cannot be accounted for through the suffix list. These adjectives are listed together with the ones with irregular inflection patterns in *adj.exc*. Each irregular form (here including also 'regular patterns' such as *ädla* above) is listed together with its base form, e.g.:

smaa	liten	'little-PL'
mindre	liten	'less'
minst	liten	'least'
aedel	aedla	'noble-PL'
hoegljutt	hoegljudd	'loud-NEUT'

Morphy works quite well if one takes for granted that the user only enters existing word forms, but it is not psycholinguistically plausible that

such a big portion of the word forms are listed in exception files. Over 1,300 adjective forms are listed in the English version, of which most are instances of the transformation <y> --> <ie>, e.g. *breathy-breathiest*. The reason for this is that Morphy only can deal with purely concatenative processes, that is it can 'paste', but not 'cut'. It should not be too hard to implement a 'cut function' in Morphy and that would make it possible for Morphy to handle all regular inflection. The number of entries in the exception file would decrease considerably and it would only consist of purely irregular word forms.

Index of familiarity

Of course, Swedish is not better off than English in the supply of tagged corpora. At present, there are about 350,000 words of tagged Swedish text available (the Stockholm-Umeå Corpus). The approach of polysemy-counts seems to be a better approach than frequency-count also for Swedish.

I have two on-line dictionaries available: *Svenska Akademiens Ordlista (SAOL)* and Hedelin's dictionary with phonetic transcriptions (Hedelin et al. 1987). They are about the same size, about 115,000 entries, but only the latter includes parts of speech in the entries. Hedelin, however, does not list different senses for polysemous words to any larger extent. There are other dictionaries on the market but they have interfaces facilitating human-computer interaction but obstructing any other use of the data. Without a dictionary that lists different senses and that lists the data explicitly, it is not possible to obtain the familiarity indexes automatically this way either. I hope it will be possible to obtain such a dictionary in the future, it would certainly be useful also for other tasks.

The index of familiarity in the Swedish implementation is therefore obtained semiautomatically by counting the different senses listed in the synonym function of MSWord 5.1 and entering them in the Poly-files manually. MSWord 5.1 also uses Strömberg's synonymordbok.

Necessary files

The lexicographers' files and the software generating the database, e.g. Grinder, is a file package stored separately from the database and the retrieval software. In fact, it is possible to download only the database (for English) and the retrieval software via ftp from the address mentioned in footnote 1.

The files that were created or altered when implementing WordNet for Swedish adjectives are listed below.

dbfiles	source
adj.all adj.pert adj.exc noun.body noun.act noun.location	morphy.c

Conclusion

There are in principle no theoretical obstacles to the implementation of a WordNet for Swedish. The system has been proven to work for Swedish adjectives and there will probably not be any problems with the other parts of speech either.

WordNet is a very good system for coding semantic relations between lexemes. The lexicographers' files have a fairly simple syntax, they are easy to read, and it is easy to add and extract information. However, there are some problems concerning customisation to other languages. For Swedish, the character set has to be altered to include å ä ö, and when it comes to entering morphological information, several improvements can be made. Morphological information is not entered in the lexicographers' files. Suffixes are hard-coded in the C-program morph.c and exceptions are listed in the exception list adj.exc. Editing the listed suffixes in morph.c does not really involve any programming, but finding the proper place (and file) to edit requires such knowledge. Modularising, so that the morphological rules end up in a separate file among the lexicographers' source files, would be a more elegant solution, being easier to customise to other languages as well as to add new information to. There are also many morphological rules that cannot be dealt with by Morphy, eg. phoneme alternation and deletion. The possibility of entering more complex rules is desirable as well as a system to mark the words according to what morphological pattern they follow in the inflection.

WordNet is a very ambitious project. A dictionary is never complete, and there will always be types of information that could be added. As for the adjectives, it would be desirable to have information about what nouns, or group of nouns, are modified by a specific adjective. Some attributive relations are coded in the English WordNet, e.g. *warm* is linked to the noun *temperature*, which in turn is linked to *property*. But nothing is said about who or what has such a property. An idea would be to link the adjectives directly to the noun or group of nouns it can modify. For example *hårig*

'hairy' can modify concrete things such as in *en hårig man* 'a hairy man', *ett hårigt blad* 'a hairy leaf', but not abstract phenomena like **en hårig dag* 'a hairy day'. Another idea (and a faster way) is to macrocode the lexicon with co-occurrence information, but then one is left with the problem of how to interpret the co-occurrence index. The number does not really say anything about the type of relation between the two words co-occurring. It seems like the manual approach is the most suitable for implementing attributive relations in WordNet, yet another task for the lexicographer.

Building a WordNet of the calibre of the existing English one is a time-consuming task which demands collaboration and many man-hours. But as a devoted end-user of the English WordNet, I know how useful it is and it would definitely be worth the trouble to build a Swedish WordNet. It would not only add to the sparse collection of machine-readable Swedish dictionaries, but also give new insights into the Swedish vocabulary.

References

- Bolinger, Dwight. 1967. 'Adjectives in English: attribution and predication'. *Lingua* 18, 1-34.
- Charles, Walter & George Miller. 1989. 'Contexts of antonymous adjectives'. *Applied Psycholinguistics* 10:3, 357-375.
- Climent, Salvador, Horacio Rodríguez & Julio Gonzalo. 1996. *Definition of the links and subsets for nouns of the EuroWordNet project*. URL: <http://www.let.uva.nl/~ewn/D005.ps>.
- Cruse, D.A. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.
- Hellberg, Staffan. 1978. *The morphology of present-day Swedish: word-inflection, word-formation, basic dictionary*. Stockholm: Almqvist & Wiksell International.
- Lyons, John. 1977. *Semantics*. Vol. 1. New York: Cambridge University Press.
- Miller, George, Richard Beckwith, Christiane Fellbaum, Derek Gross & Katherine Miller. 1993. *Five papers on WordNet*. URL: <http://www.cogsci.princeton.edu/~wn>. Cognitive Science Laboratory, Princeton University.
- Miller, George & Christiane Fellbaum. 1992. 'Semantic networks of English'. In Beth Levin & Steven Pinker (eds.), *Lexical and conceptual semantics*, 197-229. Amsterdam: Elsevier Science Publishers.

Zipf, George K. 1945. 'The meaning-frequency relationship of words'. *Journal of General Psychology* 33, 251-256.

Lexical resources

A fragment of SUC: the Stockholm-Umeå Corpus of modern written Swedish. In European Corpus Initiative Multilingual Corpus 1 (ECI/MC1).

Bring, Sven Casper. 1962. *Svenskt ordförråd ordnat i begreppsklasser.* Stockholm: Nike-Tryck AB.

Hedelin, Per, Anders Jonsson & Per Lindblad. 1987. *Svenskt uttalslexikon* (3rd ed.). Technical Report, Chalmers University of Technology.

SAOL = *Svenska Akademiens ordlista.* 10. uppl. 1973. Stockholm: P. A. Norstedt & Söner.

Strömberg, Alva. 1995. *Strömbergs synonymordbok.* Stockholm: Strömbergs bokförlag.

Impersonation: a phonetic case study of the imitation of a voice

Elisabeth Zetterholm

1. Introduction

In language acquisition it is important to imitate the native speakers of the language. For the young child it is natural to imitate both the language and the behaviour of the culture. Imitation, or adaptation (Markham 1997), is also useful in second language acquisition to learn how to pronounce the words and to learn the prosody of the language. For most people it is difficult to learn to speak a second language in a native-like way after puberty (Larsen-Freeman & Long 1991). The normal young child does not fail in its acquisition and some people seem to have that ability even after puberty. Imitation can also be used for entertainment. Following Markham 1997, I call this type of imitation, when a speaker reproduces another speaker's voice and speech characteristics, impersonation. For the impersonator it is necessary to be aware of the target speaker's speech behaviour and characteristic features.

Some experiments have been done with animals, birds, and monkeys, in trying to teach these animals a human language by imitating (Klatt & Stefanski 1974; Linell 1978). These experiments have not been completely successful, probably depending on the anatomy of the vocal tract of the animals and since the human brain is much more complex. However Klatt & Stefanski have done some analysis with an Indian mynah bird. They observe that the imitation made by the bird was quite good in the speechlike utterances as evidenced by the acoustic analysis.

2. The present study

This paper present a phonetic case study of impersonation, focusing on the imitation of the voice and the speech behaviour. Only *one* impersonator and how he works with *one* of his impersonations has been studied. The study is restricted to phonetic aspects and ignores other aspects such as non-verbal,