LEXADV - a multilingual semantic Lexicon for Adverbs

Sanni Nimb

Center for Sprogteknologi (CST) – Københavns Universitet
Njalsgade 80, Copenhagen, Denmark
sanni@cst.dk

Abstract

The LEXADV-project is a Scandinavian research project (2004-2006, financed by Nordplus Sprog) with the aim of extending three Scandinavian semantic lexicons building on the SIMPLE lexicon model (Lenci et al., 2000) with the word class of adverbs. In the lexicons of approx. 400 Danish, Norwegian and Swedish adverbs the different senses are described with a semantic type and a set of semantic features. A classification covering the many meanings that adverbs can have has been established and integrated in the original SIMPLE ontology. Similarly new features have been added to the model in order to describe the adverb senses.

The working method of the project builds on the fact that the vocabularies of Danish, Norwegian and Swedish are closely related. An encoding tool has been developed with the special purpose of permitting easy transfer of semantic types and features between entries in the three languages. The Danish adverb senses have been described first, based on the definition in a modern, comprehensive Danish dictionary. Afterwards the lemmas have been translated and the semantic data have been copied into the Swedish as well as into the Norwegian equivalent entry. Finally these copies have been evaluated and when necessary adjusted by native speakers.

1. The Scandinavian SIMPLE Lexicons

1.1. Background

The Danish and the Swedish SIMPLE lexicons are two in the suite of formal semantic lexicons developed in the late nineties for the 12 EU languages, harmonized with respect to the encoding formalism and containing approx. 10,000 noun, verb and adjective senses for each language (Lenci et al., 2000). In the years after, the Norwegian lexicon was established on national funding at the Department of Linguistics and Scandinavian (ILN), University of Oslo. It was building on the Danish lexical data and maintaining links between entries in the two languages. This method was possible due to the fact that the vocabularies of Danish and Norwegian are closely related. The transferred data simply had to be adjusted by Norwegian lexicographers, e.g. in the many cases of different spelling of the lemma and in the cases of false friends. Furthermore approx. 5000 Norwegian lexicon entries have been linked to the Swedish SIMPLE lexicon, and the first step to establish a multilingual lexicon for the three Scandinavian languages has thereby been taken.

LEXADV is a collaborative project between Center for Sprogteknologi, University of Copenhagen (Denmark), ILN, University of Oslo (Norway) and Språkdata, University of Gothenburg (Sweden), financed by Nordplus Sprog. The idea of the LEXADV project has been to 1) extend the size of the Scandinavian lexicons for all three languages at the same time by utilizing the fact that they are closely related and 2) extend the lexicon model as well as the three lexicons with the wordclass of adverbs which was not included in the original SIMPLE model. This paper will mainly deal with the extension of the lexicon model and the Danish lexicon.

1.2. A lexicon model for adverbs in SIMPLE

In order to cover adverb senses, the original top-ontology of the SIMPLE lexicon model consisting of the types 'Entity', 'Event', 'Intensional', 'Extensional', 'Telic', 'Agentive' and 'Constitutive' has been extended with three more semantic types.

The first one, 'Functional', covers adverb senses of which the meaning must be described by logic expressions referring to sets, either in the context or in the discourse. In (Quirk et al., 1972) these groups correspond to focusing adjuncts (like *only* and *also*), and negations and conjuncts (like *otherwise* and *therefore*). In the computational lexicon COMLEX (Macleod et al., 1998) conjuncts are categorized as a certain subtype of clause modifying adverbs of which the other four subtypes are a) attitude, b) viewpoint, c) performative and d) epistemic adverbs. We find that although conjuncts syntactically modify the whole clause, they are more similar to focusing adverbs when it comes to semantics since they get their meaning by reference and logic inference (see section 1.3.1), opposite to other clause modifying adverbs. Our semantic model for conjuncts and focus adverbs is strongly inspired by (Teleman et al., 1999).

The second new main type in the SIMPLE top ontology is 'SpeakerAttitude'. It covers adverb senses expressing the speaker's attitude to the content of the sentence. In (Quirk et al., 1972) these are called disjuncts and in COMLEX they constitute the four clause modifying groups a, b, c and d mentioned above. 'SpeakerAttitude' is an independent top node because the semantics of these adverbs has a pragmatic character opposite to all other types of adverbs. The semantics of these senses can only be described by features or relations which say something about the speaker's expectations, presuppositions and feelings.

Finally the top node 'Situational' covers adverb senses describing event circumstances such as time, place and manner as well as degree adverbs. In Quirk et al. (1972) these groups correspond to the group of adjuncts except the focusing ones. In COMLEX they correspond to the groups of clause modifying adverbs called manner, temporal and LOC&DIR. Degree adverbs constitute a group on their own in COMLEX, containing adjective and adverb modifying adverbs. In fact it could be discussed whether degree adverbs in our model should rather be categorized as Emphasize senses

instead, since the degree expressed is often a very subjective statement from the speaker.

The three top ontology types are subdivided into several subtypes. In Figure 1 the SIMPLE ontological types covering adverb senses is presented.

FUNCTIONAL [Top]
  Conjunct [Functional]
    Additive [Conjunct] *desuden* (furthermore)
    Adversative [Conjunct] *derimod* (however)
    Conclusive [Conjunct] *dermed* (thus)
    Explanative [Conjunct] *nemlig* (namely)
    Disjunctive [Conjunct] *ellers* (otherwise)
    InternalReference [Conjunct] *heraf* (of this)

  Focus [Functional]
    OnlyX [Focus] *kun* (only)
    PreciselyX [Focus] *præcis* (precisely)
    EspeciallyX [Focus] *især* (especially)
    AtLeastX [Focus] *i hvert fald* (at least)
    EvenX [Focus] *endog* (even)
    TimeFocus [Focus] *allerede* (already)
    SuchAsX [Focus] *eksempelvis* (such as)

  Negation [Functional] *ikke* (not)

SPEAKERATTITUDE [Top]
  Deontic [SpeakerAttitude] *forhåbentlig* (hopefully)
  Epistemic [SpeakerAttitude] *måske* (maybe)
  Aletic [SpeakerAttitude]
    Emphasize [Aletic] *simpelthen* (simply)
    Viewpoint [Aletic] *anstændigvis* (in all decency)
    Evaluation [Aletic] *desværre* (unfortunately)

SITUATIONAL [Top]
  Spatial [Situational]
   Place [Spatial] *derude, bagi* (out there; behind)
   Direction [Spatial] *bort, hjemmefra* (away; from home)
  Manner [Situational] *forsigtigt* (carefully)
  Time [Situational]
    PointofTime [Time] *nu* (now)
      Date [PointofTime] *i går* (yesterday)
    Set [Time] *tit* (often)
    Duration [Time] *længe* (long time)
  Degree [Situational] *aldeles* (completely)

Figure 1. The top ontology of SIMPLE extended with semantic types covering adverb senses. For each semantic type a Danish example is given.

1.3. The encoding of Danish adverb senses

415 Danish adverbs have been classified and described according to the semantic types presented in Figure 1. The senses have been taken from *Den Danske Ordbog* (2003) (henceforth DDO). To give an idea of the degree of lexical coverage, there are 992 adverb lemma entries in DDO, of which of course many are polysemous. In all they have approx. 2150 senses when counting both main senses, sub-senses and senses defined under lexicalised collocations. SIMPLE-DK mainly contains main senses from DDO. It should be notified that many adverb senses in DDO are described and defined in adjective entries, especially when it comes to the many manner adverb senses. Manner adverbs though, as well as negation adverbs, have been sorted out from the Danish SIMPLE lexicon since our lexicon model has not been fully developed with respect to these two groups.

The database editor SIMPLE-SCAN functions both as an encoding tool, as a linking tool and as a search tool where e.g. adverbs sharing the same feature value or the same semantic type can easily be drawn out and compared. See Table 1 for the number of entries per semantic type.

Semantic type	Number of Semantic Units
Conjunct	22
Focus	50
Deontic	25
Epistemic	37
Emphasize	33
Viewpoint	4
Evaluation	6
Spatial	76 (29 place, 44 direction)
Degree	65
Time	97 (63 point of time)

Table 1.Number of Semantic units per type

1.3.1. Functional adverb senses: conjunct and focus

Conjuncts are adverbs like *anyway* and *thus* which link sentences together. Apart from being assigned one of the six conjunct subtypes (see Figure 1), all conjunct adverb senses are anaphoric and therefore assigned the feature 'ReferentialStatus' (added to the original SIMPLE model) with the value 'Anaphoric' (the other values for this feature are 'Descriptive', 'Interrogative', 'Deictic' and 'Deictic-Anaphoric'). Futhermore the feature 'Logic' with the set of values 'AandB', 'AorB', 'AbutB', 'BbecauseOfA' and 'AthenB' is used to describe the sense. For illustration, the conclusive adverb *hermed* (by this, with this) is assigned the value 'AthenB'. The other group of Functional adverbs is the semantic group Focus. Focusing adverbs draw attention to the element that they modify while in the same time comparing it to some other set of elements in the discourse. In very few cases though they do refer to elements in the textual context, e.g. in the case of the Danish synonyms *endda, endog, sågar* (even), all three are of the semantic type EvenX. For these the feature 'ReferentialStatus' has the value 'Anaphoric', whereas for all other focus adverb senses the 'ReferentialStatus' is 'Descriptive'. The largest group of focus adverbs in the Danish lexicon is of type 'PreciselyX' (17 entries). These senses are furthermore described by the feature 'PrecisionLevel' with the values 'Totality' (e.g. *præcis, nøjagtig* (precisely), 'High' (e.g. *næsten, omtrent* (almost) and 'Middle' (e.g. *cirka* (approximately)).

1.3.2. SpeakerAttitude senses

The many SpeakerAttitude adverb senses (105) are assigned five different semantic types (see Figure 1; Aletic functions only as a head category for its three subtypes). When using a deontic or an epistemic adverb, the speaker presupposes that the sentence is not

necessarily true and expresses either how likely he finds it to be true (epistemic, e.g. *måske* (maybe)) or how much he e.g. hopes for it to be true (deontic, e.g. *forhåbentlig* (hopefully)). In contrast in the case of aletic adverb senses, he does not question the truth of the sentence. Aletic adverb senses are divided in three. The type Emphasize is by far the largest group. One example is *virkelig* (=really like in 'I am really tired of this'). By using such an adverb the speaker emphasizes (or tones down) the utterance. The second group is Viewpoint where the speaker refers to an external viewpoint, e.g. *retfærdigvis* (in fairness, in justice)). The third group is Evaluation (e.g. *desværre* (unfortunately)) where the speaker expresses his own viewpoint.

The SpeakerAttitude adverb senses are described by means of different features, e.g. the feature 'Evaluation' (Positive/Negative), and the feature 'Emphasize Level' (High/Middle/Low/Tonedown). One single relation has also been established, namely the relation between an emphasize adverb and the feeling which provokes the speaker to use the adverb. This information is very often part of the dictionary definition of the emphasize adverb in DDO. The relation 'SpeakerFeeling' relates the adverb sense to a noun sense expressing the feeling. E.g. for the adverb *dog* (really) in DDO, sense 2, the definition refers among others to the noun *overraskelse* (surprise. See Figure 2.

Semantic Unit	DA_dog_EMP (really, emphasize)
Definition (translated from DDO)	bruges forstærkende for at udtrykke at noget gør stort indtryk el. påvirker den talende følelsesmæssigt og fx giver anledning til overraskelse, forundring, begejstring el. bestyrtelse **(**used emphasizing to express that something makes a big impression or has an emotional influence on the speaker and causes e.g. surprise, wonder, enthusiasm or dismay)
Corpus example	*Hvor ser du dog affældig ud, sagde Ea* (You really look decrepit said Ea)
Semantic Type	Emphasize/ Aletic / SpeakerAttitude /Top
Constitutive features	Lexical_Reference= descriptive Emphasize_level = middle SpeakerFeeling: *dog* RELATES_TO *overraskelse* SpeakerFeeling: *dog* RELATES_TO *forundring* SpeakerFeeling: *dog* RELATES_TO *begejstring* SpeakerFeeling: *dog* RELATES_TO *bestyrtelse*
Swedish	*verkligen, sannerligen*
Norwegian	*Så, jamen*

Figure 2. The entry of the adverb *dog*

1.3.3. Situational adverb senses: time, spatial and degree

Time has been treated in the lexicon model according to the notation used by (Sauri et al., 2005) in the standard mark up language TimeML. Time senses have thus been divided into 1) Point of time (in (Sauri et al., 2005) corresponding to 'Time') with the subtype Date (for (the few) adverb senses which can be replaced by a date, such as e.g. *forleden* (the other day)); 2) Duration senses, and 3) Set senses (denoting repetition or frequency). Far the largest group in Danish is the Point of time adverbs. They constitute an interesting group where different features are needed in order to describe whether the adverb expresses a punctual point of time or not, or whether this is in fact underspecified and instead depending on the event type of the actual verb in the phrase (in the SIMPLE model 'event type' is used to express Aktionsart).

The encoding shows that there is a rather small group of Danish point of time adverbs which always denote a punctual point of time and are never able to denote an iterative sense, e.g. the adverbs *snart* (soon), *straks, øjeblikkelig omgående, prompte, fluks* (immediately), *netop* and *lige* (just). Normally they only combine with verbs expressing transition or process.

Another group of point of time adverbs are on the other hand never able to denote a punctual sense and their meaning depends fully on the textual context: they have to combine with process and transition verbs in this case (e.g. *altid* (always) in *hun snakkede altid meget med ham* (she always spoke a lot with him)). Some examples are *stedse* (forever), *fremover* (in future, from now on), *førhen* (formerly) and *tidligere* (earlier, before).

A last quite big group of point of time adverbs is underspecified regarding the meaning components. Their meaning is in fact vague and depends completely on the verb and the rest of the context. They denote a punctual point of time in combinations with transition verbs (like *senere* (later (on)) in *senere tog han bogen frem* (later on he took out the book). They denote a longer (not punctual) point of time together with process or state verbs: *senere spiste han* (later he ate)*; senere læste han i bogen* (later he read/was reading in the book)*; senere boede han i huset* (later on he lived/was living in the house). These adverbs can even have an iterative meaning, although it is not the default reading (*senere læste han i bogen hver gang han fik lejlighed til det* (later he read in the book every time he had the opportunity)). Some examples of this group of point of time adverbs are *da, dengang* (at that time), *før, forinden* (before), *forleden* (the other day) and *herefter, herpå* (afterwards).

In Figure 3 the semantic unit for *førhen* is shown. Apart from the two features 'Punctual' and 'Iterative' having the values 'no' and 'underspecified', the feature 'ReferentialStatus' is used to indicate that the sense of *førhen* is deictic, referring to the textual present time opposite to e.g. the sense of the adverb *forinden* (before) which always refers to a point of time described in the context. The feature 'Temporality_type' has the value 'anterior' since *førhen* denotes a point of time before the deictic reference point of time.

1690

Semantic Unit	DA_førhen_POT_1
Definition	(DDO) tidligere; før i tiden = forhen (before)
Example	(DDO) *Førhen kunne folk godt leve med et lille handicap, det kan de ikke i dag* (before people could accept a small handicap, today they can't)
Semantic type	PointofTime/Time/Situational/Top
Constitutive features	ReferentialStatus = deictic Temporality_type=anterior Punctual=no Iterative = underspecified
Synonym	tidligere; før_i_tiden
Swedish	SE_tidigare_POT_1
Norwegian	NO_før_POT_1

Figure 3.The entry of the adverb *førhen* (before).

If we turn to the spatial adverb senses of Danish, these are divided into place senses and direction senses. They are described by features such as 'ProximitySpeaker', 'GeographicalDirection' and 'Direction' (which already existed in the original SIMPLE model but has been extended with the values 'left' and 'right'). The spatial adverb *herover* illustrates very well the description of spatial adverbs. In one sense it is an anaphoric place adverb denoting a place which is higher in position than the place referred to (= above this, over this). In the lexicon this sense has the semantic type Place; the feature 'ReferentialStatus' has the value 'anaphoric' and finally the feature 'Direction' has the value 'up'. In the second sense of *herover*, it is a deictic directional adverb (meaning over here) expressing a direction towards the speaker. It is encoded with the semantic type Direction; the feature 'ReferentialStatus' has the value 'Deictic' and the feature 'ProximitySpeaker' has the value 'NearSpeaker'.

Degree is the last subtype of the situational senses in the Danish lexicon. These senses are mainly described by the feature 'DegreeLevel' with the values 'Totality','High', 'Middle' and 'Low'. Most degree adverbs (41) have the value 'High', e.g. *særdeles* (very)*, vældig* (immensely*, overmåde* (extremely) and *overordentlig* (extraordinarily).

1.3.4. Further information in the lexical entry

In the lexical entry the dictionary definition from DDO is given as well as a corpus example. Synonyms are linked in the database and it has been a principle to try to give as many synonyms as possible. Swedish and Norwegian entries are created as copies of the Danish entry, based on lists of adverb translations and linked at the same time to Danish. Afterwards these entries are evaluated and if necessary adjusted by the national lexicographic team in Sweden and Norway.

1.4. Results and conclusions

The establishment of the Danish lexicon has given interesting insights to the semantic distribution and behaviour of the word class of adverbs. When it comes to the linking between the three languages the work is still ongoing, and it has turned out that some semantic types are far more complicated than others. This goes especially for the group of SpeakerAttitude adverbs where there are many false friends between the three Scandinavian languages. In many cases of these etymologically related adverbs the modal meaning has changed from one language to another and the abstract character of the senses in general, very often depending on pragmatic phenomena, makes it hard to find precise equivalents.

In the case of degree adverbs the problem with finding equivalents is caused by the lexical constraints many of these adverbs impose on the adjective they modify.

For other groups of adverbs though, it is quite easy to find precise equivalents between the languages. This is e.g. the case for time adverbs. Work still needs to be done regarding the linking as well as the completing of the three lexicons based on Norwegian and Swedish lemma lists instead of only Danish lemmas.

1.5. Aknowledgements

2. References

Den Danske Ordbog (2003). Hjort, Ebba & Kjeld Kristensen (red.). Det Danske Sprog- og Litteraturselskab. Gyldendal, Copenhagen.

Lenci, Alessandro, N. Bel, F. Busa, N. Calzolari, E. Gola, M. Monachini, A. Ogonowski, I. Peters, W. Peters, N. Ruimy, M. Villegas and A. Zampolli (2000) SIMPLE: A General Framework for the Development of Multilingual Lexicons. In *International Journal of Lexicography, vol. 13, number 4*, Oxford University Press, England. Pp. 249-263.

Macleod, Catherine, Adam Meyers & Ralph Grishman (1998): The Syntactic Classification of Adverbs as an Update to COMLEX Syntax: An Addition to an On-line Ressource for Research in Syntax. In *Proceedings of the ALLC/ACH'98*, Hungary.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik (1972): *A Grammar of Contemporary English*, Longman, Essex.

Sauri, R., J. Littman, B. Knippen, R. Gaizauskas, A. Setzer og J. Pustejovsky (2005) "*TimeML Annotation Guidelines*": http://www.timeml.org/timeMLdocs.

Teleman, U. Hellberg, S. & Andersson, E. (1999): *Svenska Akademiens Grammatik,* Svenska Akademien, Stockholm.