

Adverbs in Semantic Lexica for NLP – The extension of the Danish SIMPLE lexicon with Time Adverbs

Sanni Nimb

Center for Sprogteknologi, Njalsgade 80, DK-2300, Denmark
sanni@cst.dk

Abstract

In this paper we will discuss the treatment of adverbs in semantic lexica for NLP. On the basis of a semantic classification of Danish lexical time adverbs as well as a test carried out wrt. their ability to combine with different tenses and types of Aktionsart, an ontology on time adverbs is established. We will discuss which semantic characteristics exposed by the test that should be included in a computational lexicon, and propose how the ideas can be incorporated in the SIMPLE lexicon model, partly by reusing already implemented features from the model, partly by an extension of the set of features. Furthermore we will show how some adverbs will inherit information from several nodes in a SIMPLE ontology for adverbs, and how semantic relations, e.g. synonymy and antonymy, is relevant also in the case of adverbs. Finally we will give some examples on lexical entries of adverbs. The result can easily be applied on other adverbials with a time sense, and will therefore in fact also cover a large group of lexicalised multiword entities.

1. Introduction

The aim of a triennial project initiated September 2001 at Center for Language Technology, Copenhagen, is to establish a formal, syntactic and semantic lexicon covering the most common Danish lexical adverbs – approx. 900 lemmas. The semantic part of the lexicon will be developed within the SIMPLE model - an ontology-based semantic model (Lenci et al., 2000; Pedersen, B. S., S. Nimb, 2000) - and will thus constitute an extension of the SIMPLE-DK lexicon with the word category of adverbs. The SIMPLE-DK lexicon already contains encodings of approx. 10,000 senses deriving from nouns, verbs and adjectives.

Lexical semantic information on adverbs is useful in many cases in language technology products. In all kind of automatic text production or recognition, semantic information on adverbs is needed in order to be able to place them correct in the sentence. Also in the case of recognition a semantic lexicon is needed, since the meaning of polysemic adverbs is often reflected in their syntactic distribution. When it comes to time adverbs, these are furthermore often decisive for the choice of verb wrt. aspect and tense, e.g. in machine translation from Germanic languages into Roman languages.

We have initially in the project focused on the 120 adverbs with a time dimension of the approx. 900 Danish lexical adverbs. In order to develop a semantic lexicon on this group of adverbs, the 120 lemmas were in the first place categorised into different subtypes. Secondly a set of test sentences were established in order to expose the distributional behaviour of the different types of adverbs as well as their ability to combine with different kinds of Danish verbs wrt. Aktionsart and tempus. The task is then to combine the already implemented semantic relations and features from the description of nouns, verbs and adjectives in the SIMPLE lexicon, and extend the set where it is necessary in order to be able to describe time adverbs.

2. A semantic classification of Danish time adverbs

Following Klein (1994) in order to test his categories of time adverbs in the first place, the 120 adverbs were initially roughly categorised in the following groups:

1) Point in time. Positional temporal adverbs specifying a point in time in relation to another point in time, the last one being either deictic (the time of utterance) or anaphoric (a point in time given somewhere in the linguistic context). Examples are *i går* (yesterday) (deictic), *senere* (later) (anaphoric).

2) Frequency. These adverbs indicate the frequency of temporal entities. Examples are *ofte* (often), *sommetider* (sometimes), *altid* (always), *sjældent* (rarely).

3) Duration. These adverbs specify the duration of temporal entities. Examples are *i dagevis* (for days), *siden* (since).

4) Adverbs describing inherent temporal properties of a situation. Examples are *hurtigt* (quickly), *langsomt* (slowly).

5) Adverbs which can indicate the position of a situation in a series of situations. Examples are *først* (to begin with), *allersidst* (at the very end).

6) A group of quite different adverbs which fit neither of the other five classes. Here Klein (1994) mentions the English adverbs *still*, *already* and *again*.

This categorisation constitutes the basis upon which an ontology for Danish time adverbs will be established.

3. Testing distributional behaviour and interaction with aktionsart and tense

A set of test sentences was created in order to expose the distributional behaviour of each time adverb, as well as its ability to combine with different kinds of Danish verbs wrt. Aktionsart and tempus.

The description of time adverbs in SIMPLE is to be integrated with the part of the SIMPLE ontology describing events, since time adverbs and events share some of the same kind of semantic characteristics, namely those concerning boundaries or telicity, and

those concerning tense (Klein, 1994). E.g. time adverbs can interact with the Aktionsart of the verb in a sentence, having as a consequence that the aspectual sense changes, and certain time adverbs tend to combine more easily with e.g. process verbs than others. Three event subtypes in the SIMPLE ontology – namely ‘State’ (e.g. *stay, live*), ‘Process’ (e.g. *run, talk*) and ‘Transition’ (e.g. *arrive, buy*), inspired by Pustejovsky (1995) (see figure 1), describe the part of semantic information concerning Aktionsart, and a semantic unit describing an event is always assigned one of these three subtypes.

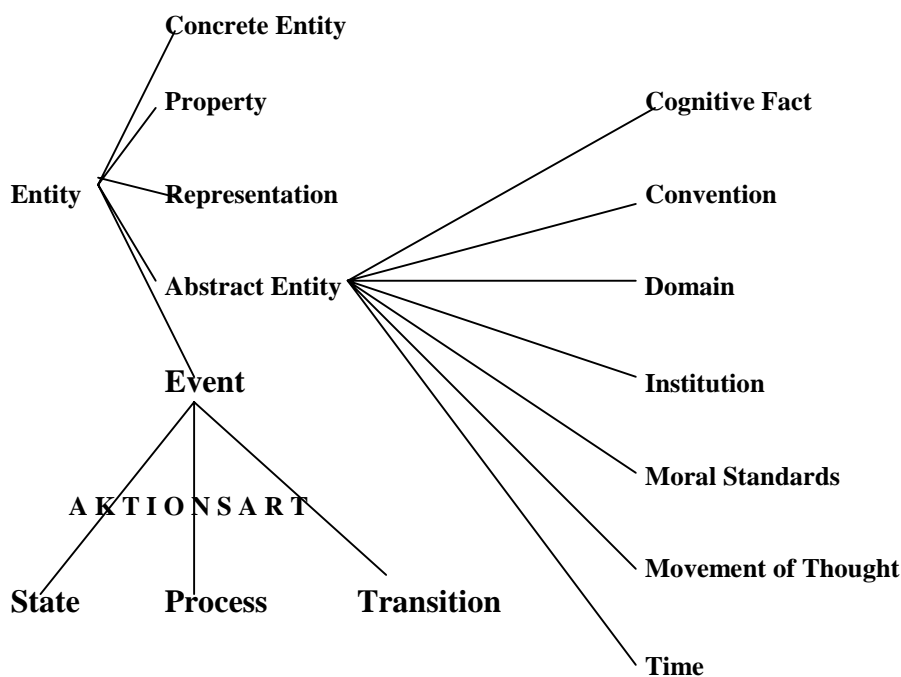


Figure 1: Part of the SIMPLE ontology: main types, the Abstract Entity types with the TIME subtype, and the Event types: State, Process and Transition.

In the test, adverbs from each semantic group were placed in a number of sentences with verbs representing the three different types of Aktionsart. The test was carried out with verbs in the following Danish tenses: present, past, present perfect, past perfect, future and future perfect.

We operated with the following positions in the sentences, according to (Diderichsen, 1968; Petterson, 1983):

Central position 1 (C1):

Han havde længe lidt af hovedpine

(Litt. He had for a long time suffered from headache)

Final position (F):

Han havde lidt af hovedpine længe

(He had suffered from headache for a long time)

Initial position (I):

Længe havde han lidt af hovedpine

(For a long time he had suffered from headache)

Central position 2 (C2), after main verb, before prepositional phrase or particle:

Han havde lidt længe af hovedpine

(He had suffered for a long time from headache)

The test showed that in many cases an adverb can only occur in some of the 4 positions, and that synonyms don't necessary behave in the same way. E.g the adverbs *igen* and *atter*, both meaning 'again', show the following distributional pattern (meaning of sentences a – h: He had suffered again from headaches)

a) *han havde igen lidt af hovedpine*

b) *han havde lidt af hovedpine igen*

c) *igen havde han lidt af hovedpine*

d) ? *han havde lidt igen af hovedpine*

e) *han havde atter lidt af hovedpine*

f) * *han havde lidt af hovedpine atter*

g) *atter havde han lidt af hovedpine*

h) * *Han havde lidt atter af hovedpine*

In some cases an adverb changes meaning according to the position, e.g. this is the case for the adverb *først* (first). Thus, the following sentences i) – l) have different

meanings, namely either sense 1: *The first thing he did was to run out*, or sense 2: *He ran out as the first person*, or both:

- i) *Først er han løbet ud*: Sense 1
- j) *Han er først løbet ud*: Sense 1
- k) *Han er løbet først ud*: Sense 2
- l) *Han er løbet ud først*: Senses 1 and 2

The test also showed that in many cases an adverb only occurs with verbs in specific tenses and with a specific Aktionsart. An example is the adverb *for længst* (a long time ago), which only occurs with verbs in either present perfect tense or past perfect tense in C1 and F, and with Aktionsart process or transition. The meaning of the sentences m) - s) is 'We had talked about it a long time ago':

- m) *Vi havde for længst snakket om det*
- n) *Vi havde snakket om det for længst*
- o) * *Vi havde snakket for længst om det*
- p) * *For længst havde vi snakket om det*

- q) * *Vi snakkede for længst om det*
- r) * *Vi snakkede om det for længst*
- s) * *For længst snakkede vi om det*

For all the different kinds of time adverbs, we want to be able to give information in the lexical entries on the semantic characteristics described above: sense/position relations and tense/Aktionsart restrictions. We also want to establish semantic relations between the different adverbs as well as to create links to other temporal expression, like nouns or verbs with a time dimension. Furthermore, for each main group of adverbs, the description of more specific characteristics is necessary.

3.1. Point in time adverbs

If we first consider the group of point in time adverbs, it is necessary to describe in the lexicon whether they are deictic relating adverbs or anaphoric relating adverbs - in (Quirk, 1972) anaphoric relating adverbs are thus considered a separate group of adverbs 'expressing relationship in time'. We also want to give information on whether the 'time' of the adverb itself belongs to the future, the past or the present - or whether it is unspecified, according to Nilsen (1972). This is, among other reasons, due to the fact that the adverb is able to determine the tense of the whole phrase, e.g. in cases where the verb is in present tense, which in Danish can also be used in a future sense. An adverb with a precise time component, e.g. the future adverb *snart* (soon) or the present adverb *nu* (now), will determine this ambiguity.

This kind of information is needed in order to assure the ability to identify and analyse temporal information in a variety of natural language applications, such as machine translation, information extraction, question answering, and multi-document summarisation.

3.2. Duration adverbs and frequency adverbs

In the case of duration adverbs, according to Quirk (1972) a distinction should be made between those denoting length of time (as *længe* (long time)), and those denoting a period from (or to) some point in time -

e.g. *siden* (since), which indicates a period beginning at a certain time in the past but which has not ended yet. These features have an influence on both tense and aspect of the verb in a sentence. More detailed studies on this subject has to be done in the project.

Also frequency adverbs tell a lot about aspect (Nilsen, 1972), implying that lexical entries on these have to contain information on whether they denote a repetition or not - whether they express the measurement in number of times (ex. *twice*) (Quirk, 1972), or whether they denote the period of time by which the frequency is measured (ex. *daily*). In Danish, though, this group of lexical adverbs is small and frequency is mostly expressed by adverbial phrases.

Information on aspect of the adverb - whether it tells something about repetition, termination or continuation of an event - is necessary in machine translation from Danish into languages expressing aspect in the verb form, as is the case for English and Romance languages. E.g. indefinite iterative adverbs (e.g. *ofte* (often)) always cause imperfective aspect in French (Pedersen et al., 1980).

3.3. A subontology on time adverbs

In figure 2 we show a first attempt of a subontology on time adverbs to be incorporated in the SIMPLE model, established on the basis of Klein (1994), Quirk (1972) and our results from the test carried out.

It is interesting that some time adverbs in the ontology share characteristics with a completely different kind of adverbs, namely adverbs which are characterised by expressing the speakers attitude to what is said (disjuncts, like *desværre* (unfortunately) and *sikkert* (probably)). This is the case for time adverbs as e.g. *endelig* (at last), *omsider* (at last), *allerede* (already) and *pludselig* (suddenly), where the speakers expectations about the time of the happening of the event is expressed. Also anaphoric relating time adjuncts share characteristics with another group of adverbs, namely the conjuncts (adverbs like *therefore*, *anyway*), in the fact that they link sentences together - this aspect is also to be studied more precisely in the project.

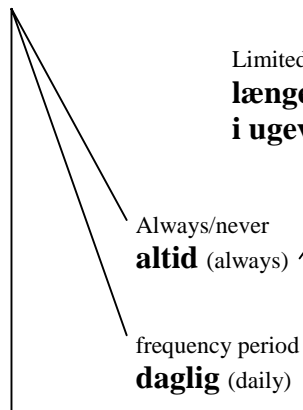
The idea of building this kind of ontology for time adverbs is to be able to assign to each adverb in the computational lexicon the ontology subtype to which it belongs, and thereby assign the characteristics it therefore automatically inherits. Afterwards the lexical entry is further supplied with different semantic relations and features characterising the adverb more specifically.

4. The SIMPLE model: already implemented lexical features describing temporal meaning components

The different meaning components of time adverbs will be described in the lexical entry by using features. In the SIMPLE model a number of features describing temporal meaning components for nouns, verbs and adjectives are already implemented, namely the following:

- Constitutive features:
- Duration = Temporary/Persistent/Underspecified
- Iterative = yes/no
- Punctual = yes/no
- Temporality_type = Past/Present/Future

• **Frequency** (**tit** (often))



• **Duration**



frequency, number of times

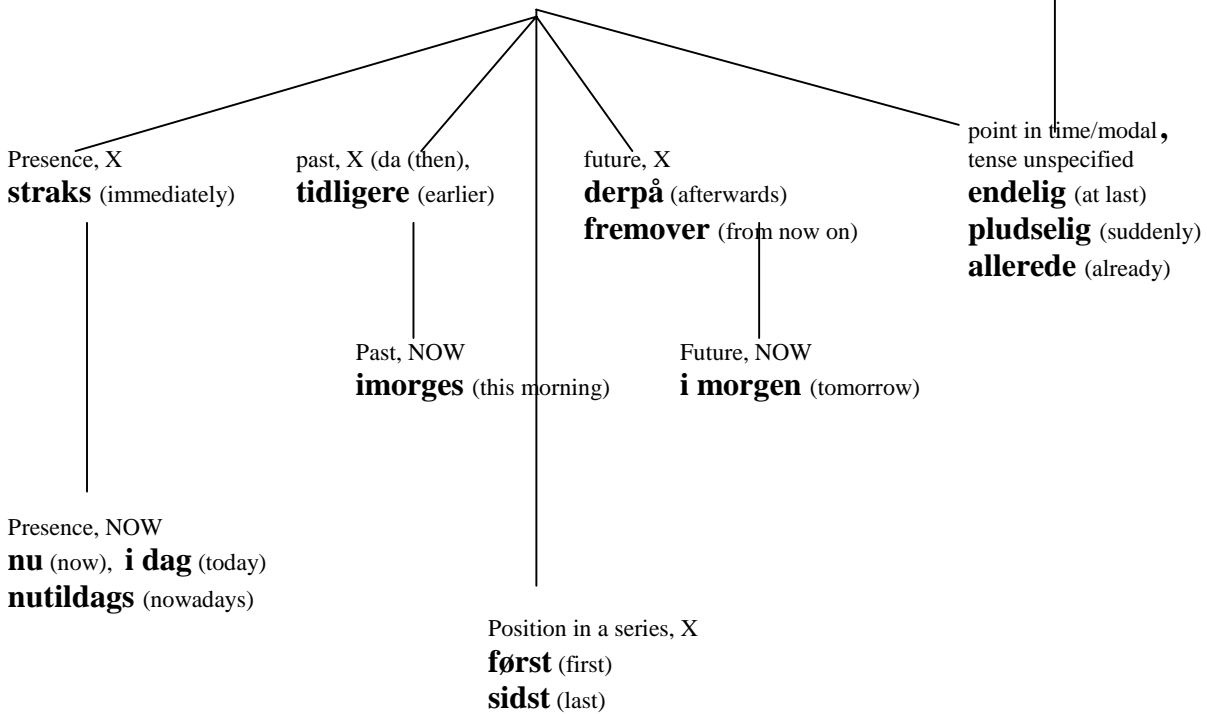
to gange (twice)

igen (again)

• **Point of time**

tense unspecified
 X **så længe** (in the meantime)
 NOW **engang** (once)

• **modal adverbs/disjuncts**



• **Inherent temporal properties of event**

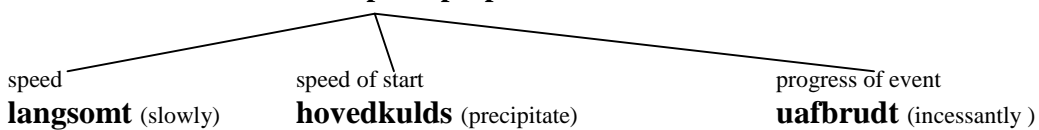


Figure 2: The proposed subontology for Danish time adverbs (X=anaphoric point in time, NOW = deictic point in time)

Furthermore, in the SIMPLE ontology model on nouns, the abstract 'Time' ontology type is intended for nouns with a time dimension (see figure 1). Here it is possible to assign constitutive qualia information like 'iterative' and 'punctual' to the semantic unit, as well as to assign semantic relations between lexical units like 'is_a' (hyperonym), *has_as_part* (e.g. 'month' *has_as_part* 'day'), *is_a_part_of* (e.g. 'day' *is_a_part_of* 'month'), *successor_of* ('Tuesday' *successor_of* 'Monday'). Also synonym and antonym relations can be expressed in the SIMPLE model. These kinds of relations are also relevant in the case of the time senses of adverbs, and in fact quite common. E.g. a temporal expression like *i morges* (this morning) could be described by *is_a_part_of i dag* (today). In figure 3 is shown an extract of the 120 time adverbs with their synonym and antonym relations.

Lemma	Antonym	Synonym
straks (immediately)		øjeblikkelig
førhen (formerly)	i fremtiden	tidligere
derefter (thereafter)	inden	så
almindeligvis (normally)		sædvanligvis
undertiden (sometimes)		indimellem
øjeblikkelig (immediately)		straks, fluks
fluks (immediately)	om lidt	straks
stadig (still)		stadigvæk, endnu
stadigvæk (still)		stadig, endnu
endnu (yet)		stadig
fortfærende (still)		stadig
snarlig (soon)	om lang tid	snart
snart (soon)	om lang tid	snarlig
sent (late)		tidligt
allersidst (at last)	allerførst	sidst
allersenest (the latest)	allertidligst	senest
jævnligt (regularly)		regelmæssigt
endelig (finally)	allerede	omsider
forbi (over)	i gang	omme
tiere (more often)	sjældnere	oftere
tit (often)		ofte
nutildags (nowadays)	før i tiden	nu
sommetider (sometimes)		nogle gange
just (just)		netop
efterhånden(as time goes on)		med tiden
indimellem (sometimes)		lejlighedsvis
længe (for a long time)	kort	lang tid
regelmæssigt (regularly)		jævnligt
lejlighedsvis(occasionally)		jævnlig, indimellem
netop (just)		just
før (before)	bagefter	inden
momentvis (momentary)		indimellem

Figure 3: Synonyms and antonyms of adverbs

5. An extension of the set of features describing temporal meaning components

The test described above, of which the purpose was to expose the distributional behavior of the time adverbs as well as its ability to combine with different kinds of Danish verbs wrt. Aktionsart and tempus, has shown, as mentioned before, that it is necessary to give information on these phenomena in the lexicon.

Also a feature describing whether at point in time adverb is deictic or anaphoric related is needed. Furthermore it has to be possible to describe the connection between the meaning in question, that is to say the semantic unit, and a specific syntactic position of the adverb.

Finally, we think that selectional restrictions are relevant for time adverbs, though this is a subject which has to be studied more carefully. E.g. some time adverbs are able to modify other time adverbs, but the modification is often restricted to a certain type. As an example, the adverb *allerede* (already) can modify adverbs of the type 'point in time':

Adverbial phrase: *allerede i morgen*
(already tomorrow)

but not adverbs of the type 'frequency':

Adverbial phrase: **allerede ofte*
(already often)

All things considered, we propose the following extension of features in the SIMPLE model, concerning the description of time adverbs:

Time_Relation = Deictic / Anaphoric

Preferable_Aktionsart = State/ Process / Transition

Syntactic_Position = C1 / C2 / F / I

Selectional_restriction: Ontological type

Further research has still to be done in order to prepare a complete model on time adverbs, but following the SIMPLE encoding principles for the description of aktionsart and of time entities, extended with the features described above, the entries of *nutildags* (nowadays), *i dag* (today), and *allerede* (already) could, as a first attempt, be composed as shown in figure 4.

The first slot 'Semantic Unit' refers to the word described; e.g. *nutildags*. The next slot, 'Definition', is preferably taken from a Danish medium-sized dictionary and helps define the actual sense. 'Corpus example' is taken from a Danish corpus (Berlingske Korpus of 20 mill. words) and is also meant as a help to the user.

'Semantic type' refers to the concept in the ontology e.g. in the case of *nutildags* 'Point of time'. 'Unification Path' gives the unification path for the unified type; in the case of the entry of *allerede* (already) the type inherits from both 'Point in time' and 'Modal'.

As regards 'Selectional restrictions', which is relevant for the entry of *allerede* (already), the concepts of the ontology are applied; thus 'Frequency', 'Duration' and

'Point in time' are concepts of the ontology applied for adverbs.

Selectional restrictions are followed by the four Qualia Roles, (i) the formal role, which provides information that distinguishes an entity within a larger set (today 'Is_a' point in time), (ii) the agentive role, which concerns the origin of an entity (not relevant for time adverbs) (iii) the telic role, which concerns the typical function of an entity (not relevant for time adverbs), and finally (iv) the constitutive role, which expresses a variety of relations concerning the internal constitution of an entity (e.g. for the case of *i dag* that temporality type is present). 'Synonymy' relations should be self explanatory.

'Syntactic Position' refers to the position in a sentence in which the adverb can have the actual meaning.

Word relations between the different time adverbs in the lexicon are in these three entries established by the semantic relations 'Synonym' and 'Has_as_part'.

Semantic Unit:	ADV_nutildags (nowadays)
Definition:	Angiver en længere tidsperiode omkring nutiden (states a longer time period in the present)
Example	Penge er i forreste række nutildags (nowadays money comes first)
Semantic type:	Point of time Presence
Formal quale:	is_a = tidspunkt (point of time)
Agentive quale:	Nil
Telic quale:	Nil
Constitutive Quale	Time_Relation = Deictic Durative=yes Temporality_Type = Present Preferable_Aktionsart = State, Process
Syntactic Position:	Central1=no Central2=no Final=yes Initial=yes
Synonymy:	Synonym= ADV_i_dag_2

Semantic Unit:	ADV_i_dag_1 (today)
Definition:	States a point of time within the same day as it is now (in the text)
Corpus example:	Statsminister Poul Nyrup Rasmussen holder kl . 18.00 i dag sin nytårstale (Prime minister Poul Nyrup Rasmussen delivers his new years speech at 6.p.m. today)
Semantic type:	Point in time Presence
Formal quale:	is_a = tidspunkt (point of time)
Agentive quale:	Nil
Telic quale:	Nil
Constitutive Quale:	Time_Relation = Deictic Temporality_Type = Present Has_as_part= <i>i_morges</i> (this morning) Has_as_part= <i>i_aften</i> etc.) (tonight) etc. Preferable_Aktionsart =Process, State, Transition
Syntactic Position:	Central1 = yes Central2= no Final = yes Initial = yes
Synonymy:	Nil

Semantic Unit:	ADV_allerede (already)
Definition:	Angiver et tidspunkt der er tidligere end forventet (States a time point which was earlier than expected)
Corpus example:	Flyet landede allerede næste dag (the plane landed already the next day)
Semantic type:	Point in time
Unification Path:	Point in time modal
Selectional restrictions:	modifies 'point in time' adverbs
Formal quale:	is_a = tidspunkt (time point)
Agentive quale:	Nil
Telic quale:	Nil
Constitutive Quale:	Punctual = yes Preferable_Aktionsart = Process, State, Transition Tense=underspecified
Syntactic Position:	Central1 = yes Central2= no Final = yes Initial = No
Synonymy:	Nil

Figure 4: The entries of *nutildags* (nowadays), *i dag* (today), and *allerede* (already).

6. Conclusion

In this paper we have proposed a subontology for Danish time adverbs, as well as a method of encoding them in the Danish SIMPLE lexicon, partly by reusing the already established semantic descriptions in the SIMPLE model, partly by using new features established in order to describe the temporal meaning component of time adverbs. The project is not finished yet and more detailed studies have to be carried out before a complete description and method is established. The aim of these investigations is of course an extension of the Danish SIMPLE lexicon with the word group of adverbs as well as the testing of the lexical entries in an application.

7. References:

- Diderichsen, P. (1968). *Elementær Dansk Grammatik*, 3. Udgave. Gyldendal, København.
- Klein, Wolfgang (1994). *Time in Language*, Routledge, London
- Nilsen, Don Lee Fred (1972). *English Adverbials*, Mouton, The Hague
- Lenci, Alessandro, Nuria Bel, Federica Busa, Nicoletta Calzolari, Elisabetta Gola, Monica Monachini, Antoine Ogonowski, Yvonne Peters, Wim Peters, Nilda Ruimy, Marta Villegas and Antonio Zampolli (2000). *SIMPLE: A General Framework for the Development of Multilingual Lexicons*. *International Journal of Lexicography*, Volume 13, Number 4. Oxford University Press.
- Pedersen, B. S., S. Nimb (2000). Semantic Encoding of Danish Verbs in SIMPLE -- Adapting a verb-framed model to a satellite-framed language. In: *Proceedings from 2nd Conference on Language Resources and Evaluation, LREC*. Athens 2000, pp.1405-1412.
- Pedersen, John, Ebbe Spang-Hanssen og Carl Vikner (1980): *Fransk Grammatik*, Akademisk Forlag, København
- Petterson, Thore (1983). *Fria Adverbial*. *Nys 13*, Akademisk Forlag, København
- Pustejovsky, J. (1995). *The Generative Lexicon*, Cambridge, MA, The MIT Press.
- Quirk R. et. al. (1972). *A grammar of Contemporary English*, Longman, England.