Semantic Encoding of Danish Verbs in SIMPLE

- Adapting a verb-framed model to a satellite-framed language

Bolette Sandford Pedersen & Sanni Nimb

Center for Sprogteknologi, Njalsgade 80,
DK-2300 S, Denmark
e-mail: bolette@cst.ku.dk, sanni@cst.ku.dk

Abstract

In this paper we give an account of the representation of Danish verbs in the semantic lexicon model, SIMPLE. Danish is a satellite-framed language where prepositions and adverbial particles express what in many other languages form part of the meaning of the verb stem. This aspect of Danish – as well as of the other Scandinavian languages - challenges the borderlines of a universal, strictly modular framework which centralises around the governing word classes and their arguments. In particular, we look into the representation of phrasal verbs and we propose a classification into compositional and non-compositional phrasal verbs, respectively, and adopt a so-called *split late* strategy where non-compositional phrasal verbs are identified only at the semantic level of analysis.

1 Introduction

The aim of the EU-project SIMPLE (Semantic Information for Multifunctional Plurilingual Lexica) is to provide harmonised semantic lexicons for Natural Language Processing for 12 of the European languages. The project is an extension of the LE-PAROLE lexicons, which contain 20,000 entries with corresponding morphological and syntactic information for each of the 12 languages that participated in the project (cf. Ruimy *et al*, 1998).

The language specific encodings in SIMPLE are performed on the basis of a unified, ontology-based semantic model - the so-called SIMPLE model - representing an extended Qualia Structure based partly on Pustejovsky (1995), partly on experiences in preceding lexical projects such as Genelex, WordNet and EuroWordNet.

In this paper we focus on the problems encountered during the encoding of *Danish verbs* in SIMPLE, an encoding process which has been continuously supported by corpus data and where some principled solutions have been required in order to adapt the universal SIMPLE model to the empirical data of a Scandinavian language like Danish.

Speaking in Talmy terms (Talmy 1985), Danish is a typical satellite-framed language, meaning that prepositions and adverbial particles express what in many other languages form part of the meaning of the verb (cf. Harder, Heltoft & Thomsen 1996, Durst-Andersen & Herslund 1996, Herslund 1993 and Pedersen 1999). Thus, several of the most frequent verbs in Danish are relatively neutral with respect to semantic affiliation in the ontology as well as regarding event type; their affiliation being determined rather by the particle or the preposition than by the verb stem itself. In fact, from our corpus examinations we estimate that more than half of the verb senses relevant for SIMPLE (relevance is here solely based on frequency) is constituted by phrasal verbs which cannot be uniquely assigned a semantic type on the basis of the verb stem alone.

Representing this aspect in a lexicon is a challenge not only for traditional lexicography but even more for computational lexicography, which has a long tradition of a modular composition of the lexicon distinguishing strictly between morphology, syntax and semantics; and which is traditionally centralised around the governing word classes, nouns, adjectives and verbs and the arguments that they take. Such a model seems intuitively better suited for a verb-framed language which encodes the core meaning components in the verbal stem.

Two questions need to be answered in order to propose a treatment of Danish phrasal verbs in the PAROLE/SIMPLE model:

- are phrasal verbs to be considered a morphological, a syntactic or a semantic phenomenon ?
- how do we represent the semantics of phrasal verbs ?

In this paper, we recognise the fact that phrasal verbs challenge a strictly modular view of the lexicon and that they are an excellent example of why such a strict modularity is in essence not ideal for the intuitive treatment of a satellite-framed language like Danish. The compromise that we suggest to handle this fact is a so-called *split late* strategy where phrasal verbs are only represented as such at the semantic level irrespective of whether they are compositional or non-compositional in meaning.

This has as consequence a somewhat controversial analysis at the syntactic level, where particles that are in fact incorporated in the verb receive a treatment parallel to weakly bound prepositional objects. In the following we first give an introduction to the event ontology in SIMPLE (Section 2); then we proceed to a description and classification of Danish phrasal verbs (Section 3), and finally we propose a treatment of these in the modular PAROLE/SIMPLE model (Section 4).

2 Verb encoding in SIMPLE

One of the fundamental assumptions behind the SIMPLE model is that word senses differ in terms of their internal complexity and that this complexity can be described on the basis of an ontology established along different dimensions (Lenci *et al.* 2000). Some word senses can be described by means of *simple* types, which means that they inherit their information from only one mother node in the ontology; others are more complex and thus inherit information from several mother nodes following the principle of orthogonal inheritance[1].

These multiple dimensions of meaning are represented in SIMPLE by means of an extended Qualia Structure model based on Pustejovsky (1995) encompassing a set of semantic relations such as *is_a*, *part_of*, *has_as_parts*, *resulting state* etc. for each qualia. Furthermore, regular polysemous classes are represented in SIMPLE via the additional type: *complex* which establishes a link between systematically related senses. In the case of verbs, complex types are in particular applied for causative/decausative alternations, as in *jeg triller bolden* (I roll the ball) vs. *bolden ruller* (the ball rolls).

In the SIMPLE ontology, verbs and event nouns are affiliated under the node *event* which again dominates a whole sub hierarchy of types to be used when classifying different kinds of events (cf. Lenci *et al.* 2000:pp. 29-30). The ontology for events is influenced by several sources including in particular WordNet (Miller *et al.*1990), EuroWordNet (Alonge *et al.* 1998) and Levin verb classes (Levin 1993). One of the aims has been to find a number of event classes which is richer than that of WordNet comprising 15 classes and less detailed than Levin's 234 classes. Thus, the SIMPLE event ontology comprises 59 classes grouped into 7 core categories as seen in Figure 1.

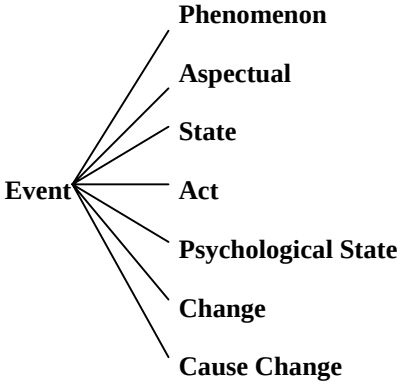

Figure 1: Core event types in SIMPLE

Three fundamental aspects have been considered in the classification:

- event type, i.e. basically whether a verb sense denotes a state, an act or a transition
- argument structure; i.e. arity and type of arguments subcategorised for by the verb sense
- causativity; i.e. whether a verb sense is causative or non-causative; the former always being represented by a unified type.

When a verb is described in SIMPLE, the appropriate set of senses to be assigned to the verb is first established, preferably on the basis of other dictionary sources as well as on corpus examination. Each sense constitutes what is labelled a semantic unit (a *SemU*) which is then assigned a semantic type according to the ontology. Each SemU is further linked to its corresponding syntactic and morphological units. Consequently, the model permits to distinguish different syntactic behaviours on pure syntactic criteria and independently of whether they share the same meaning or not (see Ruimy *et al* 1998). Figure 2 illustrates the linking of units at the different levels (*Mus* = morphological unit, *SynU* = syntactic unit, *Semu* = semantic unit).

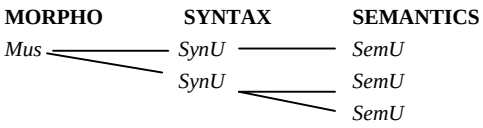

Figure 2: Linking of units at the three levels

As can be seen, one syntactic unit can very well link with two semantic units which are then maybe assigned two different types in the event ontology.

As regards the internal structure of the semantic unit; consider Figure 3 below which gives the contents of the semantic unit of one of the senses of *krydse* (cross) in the Danish SIMPLE lexicon:

Semantic Unit	*krydse_CHL* (cross)
Definition:	*Bevæge sig tværs over et åbent område* (Nudansk Ordbog) (move across an open area)
Corpus example:	*Drengen krydsede sporet ved stationen, men så ikke toget* 'The boy crossed the rails at the station but he didn't see the train'
Semantic type:	Change of location
Unification Path:	Change/Agentive
Domain:	General
Argument Structure	ARG1 ARGDirection
Selectional Restrictions	ARG1= Human OR Animal OR Vehicle Direction = Concrete
EventType	Transition
Formal quale:	Is_a = *ændring* (change)
Agentive quale:	Agentive = *bevæge_sig* (move)
Telic quale:	Nil
Constitutive quale:	Resulting_State = *være* (be) Direction= forwards
Systematic Polysemy	Nil
Synonymy	Nil

Figure 3: The semantic unit for *krydse* (cross)

[1] By 'orthogonal inheritance' we understand multiple inheritance with the restriction that a feature can only inherit its value from *one* mother node from the same partition. Thus, in SIMPLE each meaning dimension (each qualia role) establishes its own partition.

The first slot 'Semantic Unit' refers to the word described; in this case *krydse* .The suffix CHL refers to the semantic type 'Change of location' and is practical in order to keep track of other eventual readings of the same word which could require multiple linking from syntax.

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 collection of corpora (Berlingske Korpus of 20 mill. words and Bergenholtz Korpus of 4 mill. words) and is also meant as a help to the user. The example chosen should be typical for the sense and should exemplify argument realisation and typical selectional restrictions.

'Semantic type' refers to the concept in the ontology 'Change of location' which is placed as a subtype to 'Change'. 'Unification Path' gives the unification path for the unified type; in this case the type inherits from both 'Change' and 'Agentive'. 'Argument structure' should be self-explanatory; however, it should be noted that each language group has here been relatively free regarding how to analyse predicates. The Danish lexicon is based on the linguistic specifications developed within an EU-grammar project (LINDA – Linguistic Specifications for Danish (cf. Underwood et al. in press)).

As regards 'Selectional restrictions', the concepts of the ontology are applied; thus 'Animal', 'Human' and 'Vehicle' are concepts of the ontology applied for concrete nouns. 'Event type' can be either 'state', 'process' or 'transition' and is meant to refer to the 'neutral' interpretation of the verb in question – a somewhat problematic issue for Danish which we shall come back to in Section 3. In the case of *krydse*, however, there are no problems in assigning the value 'transition'.

Event type is followed by the four Qualia Roles, (i) the formal role, which provides information that distinguishes an entity within a larger set (*krydse* 'Is_a' *ændring* (cross 'Is_a' change), (ii) the agentive role, which concerns the origin of an entity (in this case *bevæge_sig* (move)) (iii) the telic role, which concerns the typical function of an entity (*krydse* has no such function), and finally (iv) the constitutive role, which expresses a variety of relations concerning the internal constitution of an entity (in this case 'Resulting State' which is to be another place (*være)* and 'Direction' 'forwards'). 'Systematic polysemy' and 'Synonymy' relations are not relevant for the encoding of *krydse.*

3 Danish verbs and adverbial particles

3.1 Unit accentuation as a linguistic test

When analysing Danish verbs and their satellites, unit accentuation proves to play a central role (see Scheuer 1995 and Harder, Heltoft & Thomsen 1996). In fact, unit accentuation can be used as a linguistic test in order to distinguish phrasal verbs from other verbs combined with adverbial particles. Consider the examples below:

(1) *Han 'blev 'væk*
'he stayed away'
(2) *Han blev 'væk*
'he got lost'

Example (1) has stress on both the verb and the particle indicating the fact that we are dealing with a simplex verb *blive* (stay) combined with an adverbial modifier *væk* (away).

In contrast, in (2) absence of full stress on the verb indicates that the verb does not constitute a clausal predicate on its own but that the element that receives full stress (the particle) should be interpreted as part of the semantics of the predicate. In the case of (1), *blive* can be described as a state verb - i.e. as non-transitional - subcategorising for a locational argument; in the case of (2), we must consider *blive væk* as one semantic unit; i.e. a phrasal verb of the type transitional with a 'change of location' assignment.

3.2 Compositional vs. non-compositional phrasal verbs

When looking closer into the category of verbs which can be categorised as phrasal verbs according to the test described above, a blurred picture emerges between those phrasal verbs that are compositional in meaning and those that are not. With compositionality we understand that both the host verb and the particle retain their core meaning as is normally the case when directional particles are combined with motion verbs as in:

(3) *han løb 'ud*
'he ran out'
(4) *han gik 'op*
'he walked up'

Compositionality is an important parameter when deciding how to represent a phrasal verb in the computational lexicon, and actually; in the traditional lexicon this distinction is also usually maintained although several Danish dictionaries are not completely clear on this point. Normally, however, only the non-compositional phrasal verbs find their way into a traditional lexicon as *sublemma* as for instance in the case of *vaske op*:

(5) *han vaskede 'op*
he washed up
'he did the dishes'

whereas the compositional phrasal verbs, which are predictable in meaning and often productive wrt. to the directional particle to be connected with, are rather described by means of valency patterns in the 'core' entry, as in *løbe op/ned/ud...* ('he ran up/down/out..) resulting in the following valency pattern description: SUBJECT+DIRECTIONAL[2].

[2] We must remark here, however, that frequency also plays a role in the construction of most modern lexica; thus very frequent phrasal verbs do tend to figure as sublemma even if they are predictable in meaning.

It should be noted that such an analysis may coincide with the description given for some non-phrasal verbs, resulting in the confusing fact that predictable phrasal verbs and non-phrasal verbs combined with directional particles are described in the same fashion in the lexicon. Consider the example below, which cannot be considered a phrasal verb according to the accentuation test, and where *ned* is thus not incorporated in the verb:

(6) *han 'kiggede 'ned*
'he looked down'

Such an example would usually be treated similarly to examples like (3) and (4) above.

3.3 Motion verbs as a special semantic class

The best way to solve this apparent contradiction, is to recognise that motion verbs constitute a very special semantic class. Motion verbs occurring as phrasal verbs are mostly predictable in meaning since the directional particle is usually to be understood in its core sense; motion verbs thus constitute a unique semantic class by admitting the directional marker to be understood as *incorporated* in the verb as is shown by the accentuation test. This relates well to the fact that other motion verbs inherently express direction without the need for a directional particle; (cf. *krydse* (cross) in Section 2). It also opens for the possibility of viewing the expression of *direction* as a regular syntactic pattern which alternates with directional prepositional phrases:

(7) *han løb ud / til bageren / hen til skolen*
'he ran out/ to the Baker's / over to the school'

In the group of motion verbs that can occur as compositional phrasal verbs should also be included 'Cause change of location' verbs such as *flytte* (move), which have the same characteristics:

(8) *han flyttede ud/ ind/ hen til skolen*
'he moved out/ in/ over to the school

Furthermore, this group includes verbs from other domains that are 'coerced' into motion verbs when a directional particle is added:

(9) *han humpede hjem på sit forslåede ben*
'he limped home on his battered leg'
(10) *han skramlede ned ad vejen i den gamle bil*
'he rattled down the road in his old car'

3.4 Particles and prepositions acting as aspectual markers

It is a well-known characteristic of most satellite-framed languages that the particle acts as an aspectual marker typically changing a process verb into a transition verb as seen in examples (3) and (4), where process motion verbs are changed into becoming change of location verbs. In some cases the aspectual marking is the only function of the particle as in the examples below where process verbs like *spise* (eat) and *drikke* (drink) are changed into transition verbs:

(11) *han spiste 'op* TRANSITION
'he ate up'
(12) *han drak 'ud* TRANSITION
he drank out
'he drank up'

But there also exist directional particles that are *not* transitional; in the examples below the meaning is that of a process even if a directional particle is added:

(13) *han gik hjemad* PROCESS
'he walked homewards'
(14) *han daskede rundt* PROCESS
'he sauntered about'

Likewise if two directional particles with opposite meaning are applied together as in:

(15) *pendulet svingede frem og tilbage* PROCESS
'the pendulum swung forwards and backwards'

Note also that the preposition *på* (on) can have an aspectual marker function (marking a process rather than a transition) when combined with a verb like *bygge* (build) which normally denotes transitions when combined with a nominal phrase:

(16) a. *hun byggede et hus* TRANSITION
'she built a house'
b. *hun byggede på et hus* PROCESS
she built on a house
'she was building on a house'

Several other features can also influence the event type, such as definiteness of the object, addition of a resulting state etc.

3.5 Concluding remarks

In the preceding sections we have identified three kinds of constructions involving directional particles:

- simplex verbs combined with adverbial modifiers ((1), (6))
- phrasal verbs which are compositional, i.e. predictable in meaning ((3)(4))
- phrasal verbs which are non-compositional, i.e. idiosyncratic ((2), (5))

We have also seen that the second group of verbs is constituted by motion verbs; a unique semantic class in the sense that it admits the directional marker to be understood as *incorporated* in the verb in spite of the fact that the verb itself contributes to the meaning of the expression. Finally, we have looked at particles and at the preposition *på* as aspectual markers which switch the event type of a verb into another type.

4 Representing phrasal verbs in the modular PAROLE/SIMPLE framework

4.1 Lemma identification and *split late* philosophy in PAROLE/SIMPLE

The question is now how to represent these aspects in a verb-framed model like the SIMPLE event ontology where one of the basic underlying classification criteria has been exactly the event type. Leaving the simplex verbs aside in this discussion, it seems obvious that the idiosyncratic phrasal verbs must be fully lexicalised at the semantic level since their meaning is unpredictable and therefore requires a semantic description of its own. For the compositional phrasal verbs on the other hand, we can opt for either a fully lexicalised representation or for a directional slot representation of some kind. In any case we need to consider more thoroughly the whole PAROLE/SIMPLE structure in order to decide for a convenient strategy since the representation of the *lemma* as such, as well as the representation of syntactic information, prove to be highly relevant for this discussion.

When deciding how to represent the semantics of phrasal verbs in PAROLE/SIMPLE, we need not only consider how to represent them at the semantic level, but also to find a principled solution wrt. their representation at earlier levels; i.e. the morphological and the syntactic level. An interesting aspect of the PAROLE/SIMPLE model is that there exists no *lexical* unit or *lemma* as such in the traditional sense of the word. In order to identify what in traditional lexicography is conceived as a lemma, one has to start from the semantic unit and work back through the syntactic and morphological levels and gather all the relevant information.

This is due to the fact that the Danish PAROLE/SIMPLE lexicon consistently applies the so-called *split late* strategy. A split late strategy implies that only what can be identified at a particular level of analysis as two different units (morphology, syntax or semantics) should result in the entry being split into separate units. Thus, for the two homographs of *love* (promise, praise) for instance, we find the following representation in the Danish PAROLE/SIMPLE lexicon:

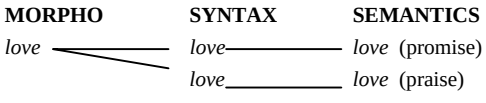

Figure 4:The representation of *love* (promise, praise)

In other words, even if we speak of homographs with different etymology and completely unrelated meanings, only one representation is given at the morphological level since the two are identical from a purely morphological point of view. The split into two units is realised at the syntactic level since the valency patterns of the two verbs differ; the former being ditransitive and the latter transitive.

For the representation of phrasal verbs several approaches can be adopted. We can either lexicalise a phrasal verb like *vaske op* at the morphological level and thus treat it as a completely different lexeme than *vaske*:

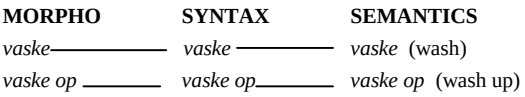

Figure 5: 'Split early' representations of *vaske* and *vaske op* (wash, do the dishes)

Or we can split at the syntactic level and lexicalise the phrasal verb *vaske op* at this level:

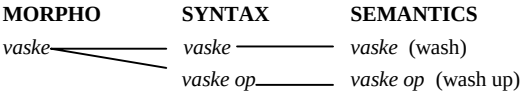

Figure 6: Splitting in syntax of *vaske* and *vaske op* (wash, do the dishes)

It does, however, match the split late strategy better to make the distinction at the semantic level and let the particle be treated as an optional complement at the syntactic level:

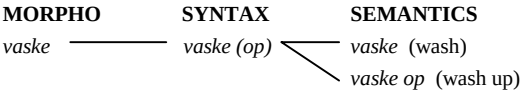

Figure 7: 'Split late' representations of *vaske* and *vaske op* (wash, do the dishes)

To consider *op* an optional complement to *vaske* may be controversial from a syntactic point of view but it has the advantage of leaving the decision of whether we are dealing with a compositional or a non-compositional phrasal verb for the semantic level where it actually belongs since it is basically a semantic distinction. Especially in cases of *ambiguity* (i.e. where both a compositional and a non-compositional interpretation is possible, as in *gå op* which can either be compositional meaning 'walk upwards' or non-compositional meaning 'cancel out') this is convenient since it prevents unnecessary overgeneration at earlier levels and allows for a unified syntactic description of directionals at the syntactic level irrespective of whether these are expressed as particles or as prepositional phrases[3]:

[3] Two aspects should be noted here: firstly, the strategy proposed here obviously introduces an element of non-compositionality between syntax and semantics; secondly, for a lexicon meant for speech recognition we would probably prefer a *split early* strategy (at the morpohological level) since ambiguity would then be eliminated by information on stress.

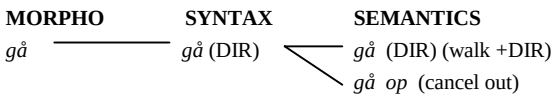

Figure 8: 'Split late' representation of *gå op* (walk upwards, cancel out)

4.2 The semantic unit

The non-compositional phrasal verb *gå op* (cancel out) is given a particular, fully 'lexicalised' representation in the semantics and is assigned the type 'Identificational State' from the ontology denoting states which ascribe a property to an entity and e.g. identity and resemblance with another entity:

Semantic unit:	*gå op_IDS* (cancel out))
Definition:	*(om regnestykke og kabale) løses så der ikke bliver nogen rest* (Nudansk Ordbog) '(about calculations and patiences) solve so that there is no remainder'
Corpus example:	*..men for at få regnestykket til at gå op måtte han inddrage naboens grund* 'but in order to make the calculation cancel out he had to include the neighbour's garden'
Semantic type:	Identificational state
Sem. Supertype:	Relational state
Event type:	State
Domain:	Mathematics
Predicative rep:	ARG1
Selectional restrictions:	ARG1= Representational
Formal quale:	isa = *relation* (relation)
Agentive quale:	Nil
Telic quale:	Nil
Constitutive quale:	Relates = numbers
Systematic Polysemy:	Nil
Synonymy:	Nil

Figure 9: Semantic representation of *gå op* (cancel out)

As for the compositional phrasal verbs, the representation in Figure 8 has the advantage that these are treated in the same way as constructions with weakly bound preprositions (see example 7). Such a generalisation opens for a representation at the semantic level where only one semantic unit is established for *gå* (walk) as depicted below:

Semantic unit:	*gå_MOV* (walk - move reading)
Definition:	*komme frem ved at sætte den ene fod foran den anden* (Nudansk Ordbog) 'proceed by putting one foot in front of the other'
Corpus example:	*Vi skal **gå** hen til telefaxen , vente mens den kalder op osv.* 'we have to **walk** over to the fax machine, wait while it makes the call etc.'
Semantic type:	Move
Sem. Supertype:	Act
Event type:	Process
Domain:	General
Predicative rep:	ARG1 (DIR)
Selectional restrictions:	ARG1= Human OR animal DIR= Concrete
Formal quale:	isa = *bevæge sig* (move)
Agentive quale:	Nil
Telic quale:	Nil
Constitutive quale:	Manner = yes
Systematic Polysemy:	Nil
Synonymy:	Nil

Figure 10: Semantic representation of *gå* (walk)

One semantic unit can now account for *gå* in phrases like *han gik 2 km, han gik ud, han gik op, han gik hen til telefaxen* etc. (he walked 2 km, he walked out, he walked up, he walked over to the fax machine) since no lexicalisations of the phrasal verbs are involved. The problem of event structure assignment, is however *not* solved by this approach. In the SIMPLE Specifications (Lenci *et al.* 2000) it is stated that event type is meant to abstract away from the possible effects by complements and adjuncts – therefore the semantic unit in Figure 10 is assigned the event type 'process'. Thus, in this representation we ignore the fact that *gå* in some cases denotes a process and in other cases a transition and that there holds a regular pattern between the two that could have been represented as systematic polysemy between 'move' constructions and 'change of location' constructions. Again, the modular organisation of the SIMPLE model centralised around the governing word classes is not completely well-suited for this kind of phenomenon.

In the present representation it is thus up to a particular grammar or lexical rule set to capture this pattern which – as is well-known under the term *path-manner divergences* (cf. Talmy 1985, Slobin 1996, Pedersen in press and many others) - is crucial for instance in case of translation between Germanic and Romance languages. This is caused by the fact that Romance languages usually lexicalise the direction component (or 'path' in Talmy's terms) in the verb stem whereas Germanic languages lexicalise the manner component in the verb stem. So, *gå i to timer* (walk for two hours) would in Spanish translate into *caminar durante dos horas* whereas *gå ud* (walk out) would translate into a 'Change of location' verb: *salir* (exit). In other words, in Spanish the two verbs *caminar* and *salir* are placed in two different parts of the ontology; under 'Move' and 'Change of location', respectively, whereas for Danish we would only have one semantic unit affiliated under 'Move'.

5. Conclusions

Strict modularity as well as centering around the governing word classes and their arguments is an obvious approach for a computational, multipurpose lexicon where the idea is that the lexicon should be usable for different kinds of NLP applications requiring different levels of linguistic information.

In this paper, we have discussed some of the linguistic problems encountered when adapting such a model to Danish verbs, and in particular to Danish phrasal verbs, which are not easily described from a modular point of view since they incorporate particles and therefore are discontinuous. In other words their analysis lies in the borderline between morphology and syntax on the one hand, and syntax and semantics on the other. In order to overcome this problem we have suggested a split late approach where at the morphological and syntactic levels simplex verbs with directionals, compositional phrasal verbs as well as non-compositional phrasal verbs are treated alike: the particle is treated as a valency slot filler even if from a semantic point of view it is incorporated in the verb.

This approach leaves it for the semantic level to differentiate between the different kinds of particle constructions since it is at this level the proper disambiguation can take place between the ones that are non-predictable in meaning and should therefore be lexicalised and those that are predictable in meaning and can therefore be described in a regular fashion.

Attempts to harmonise linguistic descriptions of different European languages into a universal model constitutes a challenging task but they also bring linguistic research further. Thus, the scope of the SIMPLE project makes it a truly pioneering project for Danish and considering the current status of language technology for the 'small' European languages, the development of these harmonised large-scale semantic lexicons is a first step in the right direction for creating advanced language technology also for less widely spoken European languages.

References

Alonge, A., N. Calzolari, P. Vossen, L. Bloksma, I. Castellon, M.M. Marti, W. Peters (1998). 'The Linguistic Design of the EuroWordNet Database', In P. Vossen (ed.) EuroWordNet, A Multilingual Database with Lexical Semantic Networks.Kluwer Academic Press, Dordrecht, Boston, London.

Braasch, A. & B. Pedersen (1999) 'En stor sprogteknologisk ordbog for dansk - med særligt fokus på håndtering af flertydighed i en niveaudelt ordbog', in: *7. Møde om Udforskning af Dansk Sprog*, Århus University.

Durst-Andersen, P. & M. Herslund (1996). 'The syntax of Danish verbs: Lexical and syntactic transitivity', in: E. Engberg-Pedersen et al. (eds.) *Content, Expression and Structure. Studies in Danish Functional Grammar*. John Benjamins, Amsterdam.

Harder, P., L. Heltoft & O.N.Thomsen (1996).'Danish directional adverbs, content syntax and complex predicates: A case for host and co-predicates', in: E. Engberg-Pedersen et al. (eds.) *Content, Expression and Structure. Studies in Danish Functional Grammar*. John Benjamins, Amsterdam.

Herslund, M. (1993). 'Transitivity and the Danish Verbs', in *LAMBDA no. 18*, Copenhagen Business School, Copenhagen.

Lenci, A. F. Busa, N. Ruimy, E. Gola, M. Monachini, N. Calzolari, A. Zampolli, J. Pustejovsky, E. Guimier, G. Recourcé, L. Humphreys, U. Von Rekovsky, A. Ogonowski, C. McCauley, W. Peters, I. Peters, R. Gaizauskas, M. Villegas, O. Norling-Christensen (2000). *SIMPLE Linguistic Specifications*, University of Pisa.

Levin, B. (1993) *English Verb Classes and Alternations, A Preliminary Investigation*. The University of Chicago Press, Chicago and London.

Miller G., R. Beckwith, C. Fellbaum, D. Gross, K.J. Miller (1990). An On-line Lexical Database, in: *International Journey of Lexicography, 3(4), p.235-244.*

Pedersen, B.S (1999).'Systematic Verb Polysemy in MT: A Study of Danish Motion Verbs with Comparisons to Spanish', in H. Somers (ed.): *Machine Translation Vol.14, Iss. 1 p 39-86.*, Kluwer Academic Publishers, Dordrecht.

Pustejovsky, J. (1995). *The Generative Lexicon*, Cambridge, MA, The MIT Press.

Ruimy, N. O. Corazzari, E. Gola, A. Spanu, N. Calzolari, A. Zampolli (1998). 'The European LE-PAROLE Project: The Italian Syntactic Lexicon', in: *First International Conference on Language Resources & Evaluation*, Granada, Spain.

Scheuer, J. (1995). *Tryk på Danske Verber*, RASK Supplement, Vol. 4, Odense Universitetesforlag, Odense.

Slobin, D. (1996). 'Typology and Rethoric: Verbs of Motion in English and Spanish', in: Shibani and Thomsen (eds.) *Grammatical Constructions: Their Forms and Meanings*, Oxford University Press.

Talmy, L. (1985). 'Lexicalisation Patterns: Semantic Structures in Lexical Forms', in T. Shopen (ed.) Grammatical Categories and the Lexicon, Vol. 3, Press Syndicate of the University of Chicago, Chicago.

Underwood, N., C. Poulsen, P. Paggio, A. Neville, B.S. Pedersen, B. Ørsnes, A. Braasch (forthcoming) 'LINDA - Linguistic Specifications for Danish', in: *CST Working Papers*, Center for Sprogteknologi, Copenhagen.