Annotation of morphology and NP structure in the Copenhagen Dependency Treebanks (CDT)

Henrik Høeg Müller

Copenhagen Business School
Department of International Language Studies and Computational Linguistics
E-mail: hhm.isv@cbs.dk

Abstract

This paper provides an overview of the annotation design for morphological structure in CDT. The structure of words and phrases is encoded as a dependency tree which can be specified in two different ways: either as an ordinary dependency tree or by means of an abstract operator specification. The dependency notation encodes the internal structure of phrasal compounds and regular NPs, while the operator notation encodes dependency structure within solid orthography compounds and derivationally constructed words. Finally, the paper discusses the semantic labeling system used in CDT and some specific issues related to the annotation of NPs.

1 Introduction

The Copenhagen Dependency Treebank (CDT) is an ongoing project which seeks to create a parallel treebank for Danish, English, German, Italian, and Spanish with 80,000 words in each language. The CDT treebanks are based on dependency, but the annotation includes not only syntax, but also analyses of morphological, discourse, and anaphoric structure. This multilevel annotation distinguishes CDT from other treebank projects which tend to focus on a single linguistic level[1], and it has the advantage of not obliging us to limit the kind of linguistic relations that can be annotated, and not having to draw precise, and often arbitrary, boundaries between morphology, syntax, and discourse (for an outline of discourse annotation, see, e.g., Webber [20] or Buch-Kromann et al. [3]). Our main claim is that by means of a primary tree structure supplemented by an inventory of secondary relations we will be able to give a unified account of morphology, syntax and discourse which is theoretically appealing while also providing a good basis for building automatic parsers and MT-systems. However, it is not possible here to

---

[1] For instance, the Penn Treebank (Marcus et al. [9]) and the Prague Dependency Treebank (Böhmová et al. [1]) mainly concentrate on syntax; the Penn Discourse Treebank (Prasad et al. [13], [14]) and the RST Treebank (Carlson et al. [4]) focus on discourse, and the GNOME project (Poesio [11]) on coreference annotation. The TuBa-D/Z treebank (Hinrichs et al. [6]), however, includes both morphology and coreference annotation and has thus multiple levels of annotation.

account for all the general design principles behind the CDT, and, therefore, as indicated in the title, the centre of attention will be morphology and NPs.

This paper is structured as follows. In Section 2, it is explained how morphological structure is annotated in CDT. In Section 3, focus is on the marking-up of NP structure, and, finally, the most central points are summed up in Section 4, which also includes a short comment on the annotators' evaluation of the system.

2   Morphological annotation

2.1 Operator vs. dependency annotation
The morphological annotation in the CDT treebanks is only concerned with derivation and composition, since inflectional morphology can be identified and analysed automatically with a high degree of accuracy for all the languages involved in the treebanks.

The complex internal structure of words, word-like phrases and regular NPs is encoded as a dependency tree which can be specified in two different ways: either as an ordinary dependency tree, i.e. similar to syntactic dependency annotation, cf., e.g., Buch-Kromann [2], Buch-Kromann et al. [3], Kromann [8], (the *dependency notation* in Figure 1), or by means of an abstract specification of how the dependency tree for a morphologically complex word is constructed from roots in combination with morphological operators (the *operator notation* in Figure 2).
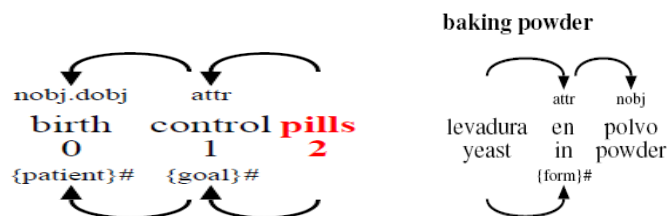

Figure 1.   Dependency annotation of the phrasal compounds *birth control pills* (left) and *levadura en polvo* [baking powder] (right).[2]

*Krigsskib***: skib –[krig]s/GOAL**   *Træbord***: bord –træ/CONST**
[war ship]   [wooden table]

Figure 2.   Operator annotation of the solid orthography compounds *krigsskib* [war ship] (left) and *træbord* [wooden table] (right).

In other words, the dependency notation specifies the tree directly, whereas the operator notation indicates how the tree can be constructed from a set of operators. The motivation for having these two annotation principles is that we use the dependency notation to encode dependency structure between

---

[2] The color code (red) and numbers (0, 1, 2, …) are tagging marks and not relevant in this context.

tokens (NPs and word-like phrases) in the automatically produced word tokenisation, while the operator notation is employed to encode dependency structure within tokens (derivations and compounds).

The analyses of the phrasal compounds in Figure 1 can be explained in the following way: The head of *birth control pills* is *pills*. The relation between the head and the non-head *birth control* is non-argumental, i.e. what we understand as one of attribution – basically because the head is non-predicative or non-relational. This relation is indicated by the arrow pointing from *pills* to *control* above the text, with the relation name written at the arrow tip. The other top arrow indicates that *control* functions as governor to the non-head *birth*, which is a noun object equivalent to a corresponding sentence level direct object.

The arrows below the text indicate semantic structure. The non-head activates the telic quale of the head – we refer to it as a "goal" relation – being the general assumption that the qualia of the head can be triggered by different modifiers, in this case a noun phrase (Pustejovsky [15], [16]). The head of *birth control* is predicative/deverbal, and *birth* fulfils the patient role of the head's argument structure (see, e.g., Grimshaw [5]).

In Figure 1 (right), the prepositional phrase headed by *en* functions attributively – being the head unable to project an argument structure – and the noun *polvo* [powder] is syntactically a noun object. The semantic relation established is "form", indicated again by the arrow at the bottom. The hash symbols following the semantic relation labels in both constructions indicate that the phrases in question show composite structure.

The operator annotations in Figure 2 show analyses of minimally complex Danish compounds. *Krigsskib* [war ship] is composed of the modifier *krig* [war], the head *skib* and the linking consonant or interfix -*s*. The annotation should be read in the following way: The minus sign indicates the pre-head position of the modifier, the lexical material of the modifier itself appears in square brackets, then comes the interfix which is a phonetically induced morpheme whose only function is to act as a glue between the head and the modifier, and finally, following the oblique slash, the meaning facet of the head noun selected by the non-head modifier, here a telic meaning relation.

The analysis of *træbord* [wooden table] follows the same principles, but here the meaning component prompted by the modifier is constitutive.

Figure 3 below shows a dependency and an operator annotation of the same Danish compound. The two types of annotation look very different, but they are merely two notational variants for the same underlying abstract dependency tree. So, you could say that the operator notation maps on to a dependency structure with equivalent principles to the ones governing syntactic expansions.
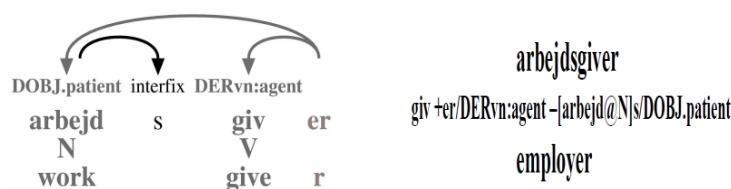
Figure 3. Morphological analysis of the compound *arbejdsgiver* [employer] annotated in dependency notation (left) and operator notation (right).

The example in Figure 3 is slightly more complex than the ones in Figure 1 and 2 because in this case the annotation of compounding is combined with that of derivational morphology. The analysis is as follows: The head of the compound is *giver* [giver], which is a derivationally complex lexeme. The operator "+er/DERvn:agent" indicates that the head is an agent nominalization of the verb *give* [give] triggered by the suffix *-er*. The annotation of the non-head, i.e. "–[arbejd@N]s/DOBJ.patient" indicates its pre-head position, that the lexical material is a noun with the interfix *-s*, cf. [arbejd@N]s, and that it corresponds syntactically to a direct object with the semantic function of Patient. The indication of word class with the specification "@*word-class*" is optional, but it should be indicated when the form is ambiguous, as in this case between a noun and a verb. The governor is the suffix which takes the root as dependent, and the non-head functions as dependent to the root. Generally, the root is governor (head) and the element activating the morphological operation functions as dependent. However, when the operator/affix is transformational or transcategorial, the operator functions as governing head and the root/stem as its dependent.

2.2 Operator annotation of different word-classes
In CDT, the three word-classes nouns, adjectives and verbs are marked-up according to the operator annotation scheme.

As illustrated below, nouns can be morphologically expanded by pre-head modifiers and/or post-head modifiers. The position of the modifier is indicated simply as a minus sign for pre- and a plus sign for post-modification. The modifier itself can be a traditional prefix or suffix, or it can be a lexical root in the form of the non-head of a compound. The positional indication, i.e. plus/minus, says nothing about that.

Prefixed noun:
  (1) *antihero*: hero –anti/NEG:contr
Suffixed noun:
  (2) *payment*: pay +ment/DERvn:core
Noun compound:
  (3) *brødproducent*: producer ! +nt/DERvn:agent –brød/DOBJ.patient
      [bread producer]

154

The adjectives in (4)-(7) are annotated according to the same annotation principles as the nouns, but the semantic categories for adjectives differ from those of nouns with respect to the languages covered by CDT, cf. Table 1 below.

Prefixed adjective:
    (4) *inactive*: active –in/NEG:contr
Suffixed adjectives:
    (5) *folkelig*: folk +e[lig]/DERna:rel.norm
        [folksy/popular]
    (6) *historic*: history ! +ic/DERna:rel.norm
Adjectival compound:
    (7) *good-sized*: size +d/DERna:rel.norm –good-/EVAL

The annotation of verbs is slightly different in the sense that they cannot carry derivational suffixes because the post-head position is restricted to inflectional endings, at least in the languages dealt with in CDT.

Prefixed verbs:
    (8) *enjabonar*: jabón –+[en][ar]/DERnv –en/AGENT
        [in-soap = do the lathering]
    (9) *dislike*: like –dis/NEG:contr
Verbal compound:
    (10) *lungeoperere*: operer –lunge/DOBJ.patient
        [lung-operate]

Summarizing, an operator has the form "*pos affix/type*". The field *pos* specifies whether the abstract affix is attached to its base in prefix position ("−") or suffix position ("+"), or a combination of these (e.g., "−+"). The field *type* specifies the derivational orientation (e.g., "DERvn", {fig. 3}), either in the form of a categorial shift, i.e. a word-class transformation, or not. Moreover, the field *type* semantically and functionally identifies the type and, where relevant, the subtype, of the dependency relation that links the base with the abstract affix (e.g., "NEG:contr", {ex. 1}). The field *affix* specifies the abstract affix and its possibly complex internal structure. The abstract affix may be encoded either as a simple string representing a simple affix or a simple root (e.g., "*er*", "*arbejd*", {fig. 3}), or as a complex string of the form "[*stem*]" or "[*stem*]*interfix*", where "*stem*" encodes the internal structure of the abstract affix in operator notation (e.g.," −[arbejd@N]s/DOBJ.patient", {fig. 3}).

    Finally, the number of exclamation marks used (e.g., "*historic*: history **!** +ic/DERna:rel", {ex. 6}) indicates how many letters have been removed from the derivational base in order to add the suffix, and the separation by square brackets (e.g., "*folkelig*: 'folksy/popular': folk +e[lig]/DERna:rel", {ex. 5}) indicates that the suffix "*-lig*" is connected to the base via the thematic vowel "*-e*". With this system of exclamation marks and brackets we are capable of separating linking elements such as thematic vowels, infixes and interfixes, on the one hand, from what is the suffix proper, on the other hand, and it allows

CDT to regenerate the word form in question on the basis of the operator instructions.

A sample of the most important relation types in the morphological annotation is listed in Table 1 below. The different relation types have taken inspiration from the works on morphological categories by Rainer [18] and Varela & Martín García [19]. All the relations can be annotated as either prefixes or suffixes or non-head roots in case of compounds; here they are just listed as they typically appear in the CDT languages. However, it is evident that some derivational meanings are typical for, or perhaps even restricted to, a specific word-class, but in principle any of the semantic relations can be used to describe derivation or compounding within all three word-classes. So, in that sense the system is flexible.

Relations that typically appear with prefixes
SPACE:loc (location: *intramural = mural −intra/SPACE:loc*)
SPACE:dir (direction/origin: *deverbal = verbal −de/SPACE:dir*)
TIME:pre (precedency: *prehistorical = historical −pre/TIME:pre*)
TIME:post (posteriority: *postmodernism = modernism −post/TIME:post*)
NEG:contr (contrast: *antihero = hero −anti/NEG:contr*)
NEG:priv (privation: *desalt = salt −de/NEG:priv*)
AGENT (causative: *acallar 'silence' = callar −a/Agent*)
TELIC (telic: *oplåse 'open' = låse −op/TELIC*)
MOD:quant (quantification: *multicultural = cultural −multi/MOD:quant*)
MOD:eval (evaluation: *maleducado* [mal-behaved] *= educado −mal/MOD:eval*)

Relations that typically appear with suffixes
AUG (augmentative: *perrazo 'big dog' = perro +azo/AUG*)
DIM (diminutive: *viejecito 'little old man' = viejo +ecito/DIM*)

*Verb derivation***:**
DERnv (noun→verb derivation: *salar 'to salt' = sal +ar/DERnv*)
DERav (adjective→verb derivation: *darken = dark +en/DERav*)
DERvv (verb→verb derivation: *adormecer 'lull to sleep' = dormir −+[a][ecer]/DERvv*)
*Noun derivation***:**
DERvn:agent (verb→noun derivation: *singer = sing +er/DERvn:agent*)
DERvn:core (verb→noun derivation: *exploitation = exploit@V +ation/DERvn:core*)
DERan:qual (adjective→noun derivation: *bitterness = bitter +ness/DERan:qual*)
*Adjective derivation***:**
DERva:pas.poten (deverbal adjective: *transportable = transport +able/DERva:pas.poten*)
DERna:rel.norm (denominal adjective: *presidential = president +ial/DERna:rel.norm*)

Relations that typically appear with compounds
CONST (constitutive: *træbord 'wooden table' = bord −træ/CONST*)
AGENT (agent: *politikontrol 'police control' = kontrol −politi/AGENT*)
SOURCE (source: *rørsukker 'cane sugar' = sukker −rør/SOURCE*)
FUNC (function: *krigsskib 'war ship' = skib −[krig]s/FUNC*)
LOC (location: *loftlampe 'ceiling lamp' = lampe −loft/LOC*)

Table 1.   Exemplification of relation types in the morphological annotation (relation types with head-switching are italicised).

3 Annotation of NP structure

This part of the paper discusses how NP structure compared with sentence level structure is annotated in CDT, concentrating on analogies and differences between these two linguistic levels (Grimshaw [5]). Figure 4 below is a simple example of a syntactic dependency annotation of a sentence. The complements *He*, *her* and *a kiss* are lexically licensed by the head *gave*, i.e. they function as arguments to the governor, while on the phrasal level *kiss* is a dependent of the indefinite article *a*. The arrows point from governor to dependent and the relation name is written at the arrow tip.
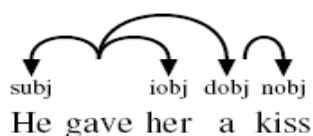

Figure 4.   Basic CDT dependency annotation of sentence.

In general, on the sentence level semantic features are not annotated, i.e. a type system for verb-based annotation has not yet been introduced, the CDT does not make use of a semantic labeling system for arguments, and neither do we attempt to identify qualia-relations in a verb-argument context. However, all free adjuncts are labeled semantically according to which semantic relation they establish with the predicate, as illustrated in Figure 5 with a "manner" relation (left), and a relation of "contrast" (right), i.e. *instead of fruits*.
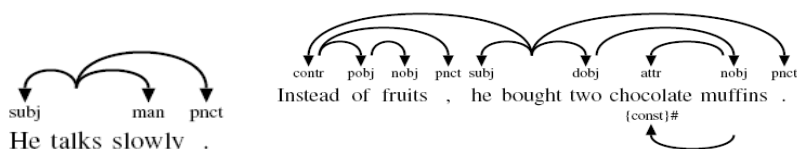

Figure 5.   Annotation of sentence level free adjuncts expressing *manner* (left) and *contrast* (right).

With respect to NP-structure, we take our point of departure in the assumption that NPs with deverbal head noun project a dependency structure similar to the corresponding verb, as the top arrows of Figure 6 illustrate. In the dependency annotation above the text, we distinguish between "pobj" and "nobj", on the one hand, and "attr", on the other hand. The syntactic labels "pobj" and "nobj" indicate that the modifying noun or PP is lexically governed by the head, whereas the "attr"-label indicates that this is not the case. "nobj" is also used more widely when a noun is governed by an article or a preposition.
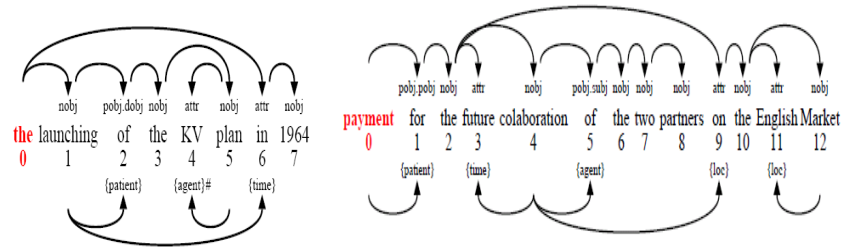
157

Figure 6.   Full syntactic and semantic annotation of NPs.

The arrows at the bottom illustrate how we on the NP-level – contrary to the sentence level – use a system of semantic labeling for both lexically governed arguments (when the head noun is deverbal, relational or deadjectival, and, hence, projects an argument structure) and free adjuncts (when the head noun is non-predicative, and, hence, establishes a descriptive or qualia-type relation). The inventory for argument labels (deverbal, relational, deadjectival) and adjunct labels (descriptive, qualia) is listed in Table 2. There is a substantial overlap between sentence level and nominal level adjunct labels, but on the sentence level CDT makes use of a number of special semantic relations, such as certain pragmatic adverbials, and, e.g., the contrast adverbial in Figure 5 (right), which for various reasons do not seem to occur on the nominal level. Generally, i.e. both in the analysis of sentence level adjuncts, NP modification and with respect to derivational morphology, we have sought to let the qualia-structure be a guiding principle for the organization of the semantic inventory in CDT. This goes also for the anaphoric relations and discourse structure, whose annotation falls outside the scope of this paper. However, this strategy does not imply that it is possible to account for any semantic relation with point of departure in the qualia-structure, as also indicated in Table 2.

Table 2.   Semantic relations for annotating NPs.

158

Evidently, the head of an NP is not always derived from a predicate, and in CDT we calculate with two other types of head nouns, i.e. relational head nouns and absolute head nouns.

Relational nouns can be divided into, on the one hand, partitive and quantitative expressions which denote arbitrary parts of something and only exist due to the whole of which they form part (such as *top, piece, liter, centimeter*, etc.), and, on the other hand, role and kinship terms (such as *member, president, mother, brother*, etc.), which have independent existence and can be employed in an absolute, non-relational manner (such as *He saw a president on the street/I am a father*).

When the head is of the first type, i.e. denotes arbitrary parts of something, we use the label "apart"[3], cf. Figure 7 (left), and the semantic relation goes from the non-head to the head, which is a consequence of split headedness in the sense that the morpho-syntactic head, N1, functions as a specifier and N2, the second noun, is the semantic head. When the head is of the second type, in case of role terms for instance, we use the label "arg", cf. Figure 7 (right), – without further intents of semantic qualification – and the arrow goes the normal way from head to non-head. This label is also used when the head noun is deadjectival.
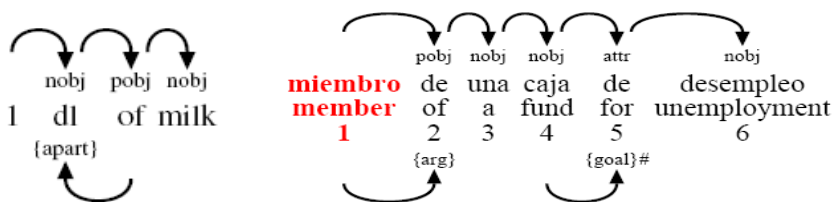

Figure 7.　　　Annotation of NPs with relational head nouns.

When the head noun is absolute, i.e. it has no connection to relational or deverbal nouns in the sense that it does not select or imply reference to any other element, cf. Figure 8, its predicative force is identified through a slightly expanded set of qualia-like relations. Our assumption is that one of the qualia-roles listed in Table 2 is activated by a modifier, which has the form of a noun or a PP.

[3] The "apart" relation is listed under the constitutive quale (CONST), cf. Table 2, which normally only applies when the head in non-relational. However, because of the special "partitive" nature established by nouns denoting arbitrary parts, the "apart" relation is categorized under the qualia-structure.
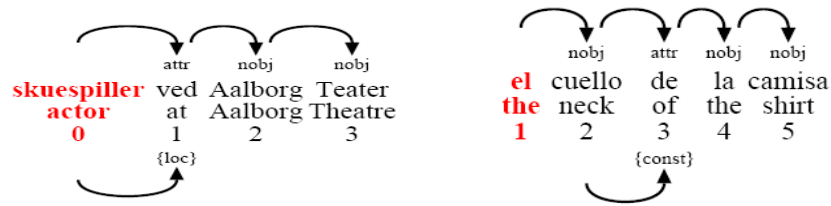
Figure 8.   Annotation of NPs with non-predicative/relational head noun.

4 Conclusion

The main conclusions and perspectives of the design principles behind the CDT annotation of morphological and NP structure are the following.

The operator annotation and the dependency annotation build on the same underlying principles. They are merely two manifestations of the same system. We need the operator annotation system to account for the internal structure of tokens, in the form of derivations and solid orthography compounds, and the dependency system to tackle relations between tokens.

When building the morphological component of CDT, we sought to establish an intimate analogy between the original dependency based, sentence level framework and the morphological analysis principles. Both systems part from the basic assumption that coherent linguistic units, in the form of sentences or words, are determined by a dependency structure in which each word or morpheme is assumed to function as complement or adjunct to another word or morpheme, called the governor. By their lexical make-up or content, governors license the complements which function as arguments, whereas the adjuncts function as free modifiers, i.e. their presence is not lexically determined by the head. This distinction between arguments and modifiers, between lexically bound and unbound elements, applies at all levels of CDT.

On the sentence level, only the free adjuncts have been annotated with respect to semantics, i.e. every adverbial modifier has been tagged with a semantic label indicating its relation to the predicate. However, in the annotation of morphology and NPs, we have gone one step further, you could say, by introducing a semantic labeling system with which we seek to identify the relations triggered by different affixes when they are attached to their lexical bases, including, not least, argument roles inside NPs and the head noun qualia-values activated by noun and PP modifiers.

Both in the analyses of sentential adjuncts, NP modification and with respect to derivational morphology, the qualia-structure has been a guiding principle of how the semantic component of CDT is organized. Many relations we know from one linguistic level are reproduced or somehow imitated on other levels, and, therefore, it is theoretically appealing to try to unify the inventory. In that respect the qualia-structure is attractive because it provides a template which is sufficiently general for structuring the relations.

The combination of morphological annotation, in its broadest sense, and alignment of parallel texts – an aspect of CDT which has not been described in this paper – will provide a good basis for doing multilingual language processing in the form of building machine translation systems. Just to mention one aspect, it is crucial to know the nature of the semantic relations that hold between NP-constituents in the source language in order to construct an analogous and well-formed nominal concept in the target language, e.g., with respect to the use of prepositions, constituent order, linking vowels or consonants, etc. (see, e.g., Johnson & Busa [7]). It is also expected that the rule-based, non-automated, hand-annotation approach, which is the actual practice of CDT, over time, and on the basis of statistical models, can develop into a more or less semi-automatic annotation system, especially taking into consideration that we do annotation on all linguistic levels. Apart from providing a basis for building automatic parsers and MT-systems, the combination of morphological annotation and alignment of parallel texts will facilitate specific inquiries into morphological cross-linguistic contrasts.

Despite the semantic granularity and complexity of CDT, the annotators generally evaluate the morphological component positively in terms of functionality and user friendliness. They especially emphasize that the hierarchical organization of the system facilitates a relatively smooth narrowing down of options to a few of the best available. Also, the high degree of specificity of the labels is mentioned as a factor which eases the final, detailed assessment. On the more critical side, the annotators find that it has been complicated and time-consuming to learn the system. In comparison with, e.g., the annotation of anaphora and discourse, the marking-up of morphological structure seems to require a deeper understanding of the languages in question both in terms of morphological structure, etymology, and (non)-productivity of certain derivational patterns, again according to the annotators. In comparison with, e.g., GLML-annotation (Generative Lexicon Markup Language) of lexical semantic structure (Pustejovsky et al. [17]), which can be done by any (native) speaker of English without prior training or too much instruction, the annotation of derivations, compounds and NPs in CDT requires a certain level of linguistic and systemic expertise.

References

[1]   Böhmová, A., Hajič, J., Hajičová, E. & Hladká, B. (2003). The Prague Dependency Treebank: a three-level annotation scenario. In A. Abeillé (ed.). *Treebanks: Building and Using Parsed Corpora*. Dordrecht: Kluwer Academic Publishers.

[2]   Buch-Kromann, M. (2006). Discontinuous Grammar. A dependency-based model of human parsing and language learning. Doctoral dissertation. Copenhagen: Copenhagen Business School.

[3]   Buch-Kromann, M., Korzen, I. & Müller, H. H. (2009). Uncovering the 'lost' structure of translations with parallel treebanks. In I. M. Mees, F. Alves, & S. Göpferich (eds.). *Methodology, Technology and Innovation in Translation Process Research. Copenhagen Studies in Language* 38: 199-224.

[4]   Carlson, L., Marcu, D. & Okurowski, M. E. (2001). Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue.*

[5]   Grimshaw, J. (1990). *Argument structure*. Cambridge: MIT press.

[6]   Hinrichs, E., Kubler, S., Naumann, K., Telljohann H. & Trushkina, J. (2004). Recent developments in linguistic annotations of the TuBa-D/Z treebank. In *Proceedings of the Third Workshop on Treebanks and Linguistic Theories*. Tübingen, Germany. 51-62.

[7]   Johnston, M. & Busa, F. (1999). The compositional interpretation of compounds, In E. Viegas (ed.). Breadth and Depth of Semantics Lexicons. Dordrecht: Kluwer Academic. 167-87.

[8]   Kromann, M. T. (2003). The Danish Dependency Treebank and the DTAG treebank tool. In *Proceedings of the Second Workshop on Treebanks and Linguistic Theories (TLT 2003), 14-15 November, Växjö*. 217–220.

[9]   Marcus, M. P., Marcinkiewicz, M. A., Santorini, B. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19(2). 313–330.

[10]   Mladová, L., Zikánová Š. & Hajičová, E. (2008). From sentence to discourse: building an annotation scheme for discourse based on Prague Dependency Treebank. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LCREC 2008)*. 2564–2570.

[11]   Poesio, M. (2004). Discourse annotation and semantic annotation in the GNOME corpus. In *Proceedings of the ACL Workshop on Discourse Annotation.*

[12]   Prasad, R., Dinesh, N., Lee, A., Joshi, A. & Webber, B. (2006). Annotating attribution in the Penn Discourse TreeBank. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text.* 31–38.

[13]   Prasad, R., Miltsakaki, E., Dinesh, A, Lee, A., Joshi, A., Robaldo L. & Webber, B. (2008a). *The Penn Discourse Treebank 2.0. Annotation Manual*. (IRCS Technical Report IRCS-08-01). University of Pennsylvania: Institute for Research in Cognitive Science.

[14]   Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A. & Webber, B. (2008b). The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08).*

[15]   Pustejovsky, J. (1991). The Generative Lexicon. *Computational Linguistics* 17: 409-441.

[16]   Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge (Mass.). London, England: MIT Press.

[17]   Pustejovsky, J., Rumshisky, A., Moszkowicz, J. L. & Batiukova, O. (2008). GLML: A Generative Lexicon Markup Language. https://sites.google.com/site/genlexml/papers

[18]   Rainer, F. (1999). La derivación adjectival. In I. Bosque. & V. Demonte (eds). *Gramática Descriptiva de la Lengua Española*. Madrid: Espasa. Volume 3, chapter 70, 4595–4643.

[19]   Varela, S. & Martín García, J. (1999). La prefijación. In I. Bosque. & V. Demonte (eds.). *Gramática Descriptiva de la Lengua Española*. Madrid: Espasa, volume 3, chapter 76, 4993–5040.

[20]   Webber, B. (2004). D-LTAG: extending lexicalized TAG to discourse. *Cognitive Science* 28: 751–779.