Tagging What Isn't There:

Enriching CG Annotation With Implicit Information

Eckhard Bick
University of Southern Denmark
eckhard.bick@mail.dk

Abstract

This paper examines ways to make existing Constraint Grammar (CG) annotation grammatically more explicit, allowing corpus users and application programs, such as machine translation (MT), to refer to context-implied grammatical features in a more direct fashion. Two types of categories are addressed. First, morphological categories are propagated to words that leave them under-specified (e.g. number and definiteness for Danish adjectives) or unexpressed (e.g. person-number for Danish verbs). Second, we also introduce new categories, such as aspect and future tense for Danish, that may be morphologically explicit in a given MT target language, but do not exist in the source language. In a pilot evaluation of four categories in the context of Danish-Greenlandic MT, the implemented enrichment grammar for Danish achieved F-scores of 97% for propagated categories and 85% for new categories. In addition to feature tagging, structural annotation is also made more explicit, adding secondary dependency links for e.g. the subjects of relative and infinitive clauses, or attribute links between subject complements and subjects.

1 Introduction

Arguably, every annotated corpus has specific uses and target groups in mind, and the choice of to-be-annotated categories, tag set and granularity may hamper research the corpus creators had not thought of. Missing information may well be present in implicit form, but difficult or impossible for the user to query. Based on feedback from users of a Portuguese treebank, Freitas et al. (2008) propose the introduction of so-called *searchables* - secondary tags that would allow the corpus equivalent of a 1-click order, subsuming in one new tag information that would otherwise be distributed across several tokens, e.g. complex tenses or np definiteness.

The same rationale can be extended beyond the corpus arena, to NLP pipelines where grammatical annotation supports applications such as proofing tools, computer-aided language learning or machine translation, all of which may be in need of specific information not explicitly provided by the underlying parser.

The add-on feature enrichment grammar presented here departs from a standard DanGram[1] CG annotation (Bick 2001) and systematically addresses under-specified and implicit information, progressing from simple morphological categories to more complex categories and dependency syntax. In sections 2 (morphology) and 4 (syntax), existing categories are treated. Section 3 is about adding new categories from distributed context clues, and section 5 addresses dependency issues. Rules and examples are for illustration purposes only. In the actual grammar, there are up to 20 rules, and/ or additional context restrictions, for the more difficult features.

[1] For an online demo and documentation, cf. http://visl.sdu.dk/da/

2 Inflection categories

2.1 Inflection morphemes

Traditional morphological analysis breaks down a word into morphemes and assigns feature-attribute pairs such as singular/plural for the number category of Danish nouns. In some cases, however, categories can be systematically underspecified and need to be disambiguated based on context. An example is the -e ending on Danish adjectives[2], wich is used independently of gender (nG=no gender), for either singular (S) definite (DEF) noun phrases (np's) or for plural (P) np's, in which case definiteness is left underspecified (nD):

store ('big')
"stor" ADJ nG P nD NOM ('[the] **big** cars')
"stor" ADJ nG S DEF NOM ('the **big** car')

One could say that the morphological categories of gender, number and definiteness come in two feature-bundles for the adjective ending -e. A tagger with a minimalist disambiguation approach would be content to choose one of the two reading lines in the cohort from context. However, a Danish np as a whole does have explicit gender (a lexical feature of the head noun, UTR[3]/NEU) and definiteness (marked either as an inflexion feature on single nouns, or by a lexically definite modifier in np's):

de P **DEF** *store* **nG** P **nD** *ulve* **UTR** P IDF ('the big wolfves)

In a first step, the new add-on grammar propagates the definiteness information from the article and the gender information from the noun, resolving the adjective's *nG* as *UTR*, and *nD* as *DEF*, making it possible to search for *common gender definite plural* adjectives in an annotated corpus.

2.2 Generalized inflection categories

However, it would still be difficult to search for *all definite/indefinite np's*, because there is no single np token that safely carries this information as an inflection morpheme. The logical candidate for a search target would be the head of the np, since it is the only obligatory part, but Danish nouns are only inflected for definiteness, if there are no (pre)modifiers. Therefore, in multi-word np's, an IDF tag on the noun head is merely a morphological zero-morpheme, while the definiteness information is distributed across other constituents of the np. This is a problem not only for corpus searches, but also for other tasks, such as syntactic tagging, topic/focus tagging, semantic role tagging and machine translation (MT). For instance, definiteness is one of several clues allowing a parser to distinguish between subjects and objects, or between agents and non-agents. In Danish-Greenlandic MT, the case of direct objects depends on definiteness (absolute changed into instrumental case for indefinite objects), and transitive Greenlandic verbs add a special *half-transitive* affix, if the direct object is indefinite.

In this case, because the noun already has a *morphological* tag for definiteness (here: IDF), our grammar adds a secondary <def> or <idf> tag referring to the definiteness of the entire np:

SUBSTITUTE (N) (<def>) TARGET (N IDF)
*-1 DEF-EDGE BARRIER NON-ATTR *(add <def> to indefinite nouns, if there is a definite np-edge word to the left with nothing but attributes in between, where DEF-EDGE is a set containing definite articles, demonstratives, possessives, genitive nouns etc.)*

2.3 Category propagation

In a further step, categories can be propagated to words that do not have them in Danish. A case in point are Danish verbs that only allow (1) tense and (2) participle morphemes, but completely lack person-number inflection common in many other European languages and Greenlandic. The example rule below harvests a person-number variable (e.g. 1S, 3P) from subject pronouns, unless they are conjuncts <cjt>, exploiting the c (child/daughter) dependency relation between subject and tense-carrying (finite) verb.

SUBSTITUTE (V) (V /$1/v)
TARGET V-TENSE
(c @SUBJ + (/\([123][SP]\)/r) - <cjt>)

Similar rules add *3S* and *3P* if the subject is a singular or plural noun, respectively. Co-

[2] Adjectives with a baseform ending in -e, and all comparative forms, do not inflect at all, and are thus ambiguous in all categories.

[3] Danish has a 2-gender system, NEU (neuter) and UTR ("utrum", common gender, historically a fusion of masculine/feminine genders)

ordinated subject trigger a plural marker, clausal subjects a singular.

A special case are the relative pronouns *som* and *der* that do not inflect in Danish, but may do so in an MT target language. Here, the dependency link from the relative clause to the antecedent can be used to recover number, gender or even semantic features from a noun, allowing MT transfer rules to "see" the necessary slot filler information in the relative clause itself.

2.4 Cross-level category transfer

Sometimes a morphological category in one language is entirely absent in another (or drastically under-specified), but still represented at the syntactic or semantic level. For instance, Danish cannot match the 6 cases used in Greenlandic, but case can still be assigned by identifying corresponding syntactic or semantic tags on the Danish side.

```
(a) SUBSTITUTE (NOM) (REL)
      TARGET N + @SUBJ
      (p VFIN LINK *1S <mv> LINK c @ACC)

(b) SUBSTITUTE (NOM) (LOK)
      TARGET §LOC
```

Thus, rule (a) assigns relative[4] case (REL) to subjects, if the parent (p) vp has a child daughter dependent (c - child) that is a direct object (@ACC). Rule (b) is an example of converting Danish semantic role tags[5] like §LOC (location) into Greenlandic case (LOK - locative).

3 Distributed information

Category mapping gets more complex, if the necessary information is distributed across several words. In Danish, this is the case for aspect, future tense and aktionsart, all of which are difficult to determine and have to be inferred from auxiliaries, framenet and semantic role tags (Bick 2011), adverbial particles and other clues. In our MT system, we introduced the secondary tags <fut> (future tense) and <iter> (iterative) in order to match special Greenlandic affixes, *SSA* and *TAR*, respectively.

```
(a) SUBSTITUTE (V) (<fut> V)
      TARGET ("ville" PR &AUX)
      (*1 @ICL-AUX< LINK 0 ("få") OR <ve>
OR V-NONCONTROL LINK *1 @<ACC
CBARRIER VV)
      (NEGATE *1 @ICL-AUX< LINK 0
("have") LINK *-1 @SUBJ> + HUM-person)

(b) SUBSTITUTE (V) (<fut> V)
      TARGET ("ville" PR &AUX)
      (c ROLE-NONCONTROL + @SUBJ) ;
```

The two rules above select the futures tense meaning of the Danish auxiliary *"ville"* over its other meaning 'want_to'. (a) looks for main verbs with frames that are -CONTROL (e.g. <fn:bodystate>, <fn:undergo>, <fn:worsen>)[6], with a safety condition of having a direct object, and an exception for *"vil have"* ('wants to have') with a human subject.

(b), on the other hand, looks for a -CONTROL subject, e.g. semantic roles like §TH (theme), §EXP (experiencer) or §STI (stimulus). At the time of writing, the add-on grammar contains about 20 rules about future tense, using hints like the following:

- always <fut> with <fn:become_be>, *blive* ('become'), *komme (komme til at* - 'shall')
- <fut> with future-triggering adverbs, dates, weekdays, months, unless the latter are modified by *hver* ('each') or the containing clause is headed by a preposition + *at* ('that')
- never <fut> with <fn:be_attr>, *omfatte* ('include'), *tilhøre* ('belong to'), *være* ('be'), *kunne* ('can'), *burde* ('should'), *måtte* ('must')
- never <fut> with generic present tense (e.g. substances or celestial bodies as subjects)

Another difficult category is aspect, since Danish does not explicitly mark any aspect categories. Telicity has a strong lexical bias and for many verbs it is possible to infer a default tag from a given verb frame. In our

[4] Greenlandic is an ergative language, and uses the neutral case (ABS) for subject of intransitive verbs and for objects of transitive verbs, changing subject case into (REL) in the latter scenario.

[5] "Adverbial" roles, e.g. time and space, are often realized by pp's in Danish. Greenlandic has no prepositions, but because DanGram tags roles on the semantic head of the pp rather than its syntactic head (the preposition), there is a simple one-on-one correspondence between Danish semantic role and Greenlandic case

[6] Some 30 frames in all

telicity scheme we use a 5-way distinction, with ±static, ±telic and ±time. In the table, ±control (±C) is added.

			-Time (0)	+Time (1)
-Static	Telic (t)	+C	goal-action	goal-activity
		-C	result-event	result-process
	Atelic (a)	+C	do-action	do-activity
		-C	pass-event	pass-process
+Static	state (s)			

Table 1: Telicity

The 5 lexical aspect categories are tagged as <aa:t0>, <aa:t1>, <aa:a0>, <aa:a1> and <aa:s>[7]. Because of the (partial) overlap between tense and aspect, and because Greenlandic verbs do not mark tense, these categories can be used to choose a Danish translation tense in the absence of more specific clues (such as time adverbs). For instance, <aa:t0> and <aa:a0> verbs like '*ramme' (hit)*, *eksplodere' (explode)* and *'opdage' (notice)* are much more likely to occur in the past tense than in the present tense.

An example of grammatical aspect is the Greenlandic morpheme category of iterative, which corresponds to the suffix *TAR* and the Danish support verb *'pleje at' (use to).* In most cases, however, the category is unmarked in Danish and has to be inferred from context:

```
(a) SUBSTITUTE (V) (<iter> V) TARGET V
      (c §LOC-TMP LINK c ("hver")) ;

(b) SUBSTITUTE (V) (<iter> V) TARGET V
      (c ("om" PRP) LINK c §LOC-TMP
      LINK 0 (<weekday>) OR (<season>)) ;
```

The first example marks a verb (V) as iterative (<iter>) if it has a dependent (c)[8] with a temporal semantic role (§LOC-TMP) modified by the determiner pronoun *'hver' (each).* Rule (b) asks for the preposition *'om' (at/on/about)* with a temporal argument (weekdays or seasons).

[7] 'aa' stands for aspect/aktionsart

[8] The CG3 implementation of Constraint Grammar uses 'c' (child) rather than the traditional 'd' (dependent/daughter).

4 Secondary syntactic tags

Pronouns are often subdivided into syntactic or semantic sub-classes such as relative, determiner, interrogative and quantifier. However, these are not necessarily lexeme classes. Thus, in Danish, the syntactic category of reflexive is only lexeme-bound in the 3rd person forms *'sig'* (accusative 'him-/herself') and *'sin'* (his/her own), and otherwise identical with ordinary personal object pronouns and possessives. For 1./2. person the <refl> mark can be safely added (a) and for 3. ps. plural it can be guessed (b):

```
SUBSTITUTE (<poss>) (<refl> <poss>)
      TARGET (<poss> @>N)
      (0 (<\([12][SP]\)>r))
      (p (*) LINK *p VFIN
        LINK c @SUBJ LINK 0 (VSTR:$1)) ;

SUBSTITUTE (<poss>) (<refl> <poss>)
      TARGET (<poss> 3P @>N)
      (*p VFIN LINK c @SUBJ
        LINK 0 (3P) OR (P) OR <cjt-head>) ;
```

In a feature propagation step, the <refl> tag can then be exploited to mark reflexivity on transitive verbs, in the presence of a <refl> @ACC tag. This mechanism is part of a more general method: valency instantiation. From its lexicon, the DanGram parser draws tags for valency potential, such as <vt> for monotransitive, <vdt> for ditransitive or <vr> for reflexive. The add-on grammar "instantiates" these tags by adding a '¤'-sign to it, e.g. <¤vt> if there, in fact, is a direct object corroborating the monotransitive tag. In the case of reflexive verbs (<¤vr>) this is useful in our MT setting, because Greenlandic verbs need to be inflected for reflexivity.

In Danish, with the exception of object-elliptic relative clauses and non-interrogative object clauses, all subclauses must begin with a subordinator. This is an obvious MT advantage with Danish as source language (SL), because the subordinator serves as a surface clue classifying the subclause and for choosing the right target language (TL) conjunction (e.g. English) or mood (e.g. Greenlandic). Thus, the conjunction *'hvis' (if)* translates into conditional mood inflection in Greenlandic. However, a little-known quirk in Danish syntax does allow conjunction-less conditional clauses, if they are fronted and SV is inverted to VS: *Kommer han ikke, må vi udskyde mødet.* ('If he doesn't come,

we will have to postpone the meeting.') Here, because word order is the only clue, a secondary marker tag is needed:

```
SUBSTITUTE (V) (<if> V) TARGET VFIN
1:  (0 @FS-ADVL>)
2:  (*-1 >>> BARRIER NON-KC)
3:  ((*1 KOMMA BARRIER CLB OR VV -
@ICL-AUX< LINK 1 VFIN)
4:  OR (*1 @<SUBJ-ALL BARRIER NON-
PRE-N/ADV LINK *1 @FMV BARRIER CLB-
ORD LINK *1 @<SUBJ-ALL BARRIER NON-
PRE-N/ADV)) ;
```

The rules asks for an adverbial subclause tag (1), beginning-of-sentence (2), and - to the right (*1) - either (3) a comma followed directly by a finite verb (VFIN) or (4) a left-pointing subject followed by a finite main verb (@FMV) and another left-pointing subject.

5 Secondary dependencies

Dependency relations (between content words) are the backbone of semantic disambiguation, be it frame annotation, semantic roles or the transfer stage of a rule-base MT system. Thus, choosing one translation of a verb over another often depends on the semantic class of its subject or object. For instance, *ride* needs 2-3 different translations in many languages, depending on its object dependent, i.e. whether you ride a horse, bicycle or train. But what do you do, if the necessary dependent either is not there, semantically empty or too far away in the syntactic tree? This is the case in Danish infinitive clauses (missing subject) and can be the case in relative clauses (missing object):

Den forsikring, han tegnede, var meget dyr. ('The insurance he took out, was very expensive')

Unlike in section 2.3, in this example there is no relative pronoun, a secondary tag could be added to. Therefore, a more structural solution is in order: Secondary dependencies. These can of course be added by the application program, MT or otherwise, that take the CG annotation as input - by following dependency paths and duplicating them where necessary. However, it is also possible to address the problem CG-internally, using the ADDRELATION(S) operator introduced in CG3 (Bick & Didriksen 2015). It allows a two-way relation, here named 'c-acc' (accusative object child), when seen from the dependent (*, the antecedent of the relative clause's "invisible" object), and 'p-acc' (parent-of-accusative), when seen from the relative clause verb (@FS-N<).

```
ADDRELATIONS (c-acc) (p-acc) TARGET (*)
    TO (c @FS-N<)
    (*-1 @SUBJ>
        BARRIER <rel> OR _TARGET_)
    (NEGATE *1S <mv>
        LINK c PRP LINK NONE c @P<);
```

In order to make sure that there is indeed an elliptic object, the rule asks for a surface subject in the relative clause and the absence of a stranded preposition, i.e. a preposition without it's argument child (c @P<), that could also be elliptic in Danish relative clauses.

Other candidates for secondary dependencies are

- subject relation between the object of a sensory or controlling verb and a dependent infinitive *(see/let **someone buy** a ticket)*
- attributive relation between subject complement and subject, or between object complement and object
- coordination, linking conjuncts both to each other and to their joint head

Expanding our example sentence to cover all these cases, automatic (DanGram) annotation[9] will look like this:

Konsulenten lod ham vide, at den forsikring, han havde tegnet, var både dyr og dårlig. ('The consultant let him know that the insurance he had taken out was both expensive and bad.)

```
Konsulenten [konsulent] <Hprof> N UTR S DEF
    NOM @SUBJ> #1->2 (The consultant)
lod  [lade-1] V IMPF AKT @FS-STA #2->0 (let)
ham [han] <aci-subj> PERS UTR 3S ACC @<ACC
    #3->2 R:c-subj:4 (him)
vide [vide] <mv> V INF AKT @ICL-<OA #4->2
    R:p-subj:3 (know)
$, [,] PU @PU #5->0
at [at] <clb> KS @SUB #6->14 (that)
den [den] <dem> DET UTR S @>N #7->8 (the)
forsikring [forsikring] <f-right> N UTR S IDF NOM
    @SUBJ> #8->14 R:c-acc:11 R:p-attr:16 R:p-
    attr:18 (insurance)
$, [,] PU @PU #9->0
```

[9] The annotation was somewhat simplified by omitting valency and certain other secondary tags.

han [han] PERS UTR 3S NOM @SUBJ> #10->11 *(he)*
havde [have] <aux> V IMPF AKT @FS-N< #11->8 **R:p-acc:8** *(had)*
tegnet tegne] V PCP2 AKT @ICL-AUX< #12->11
$, [,] PU @PU #13->0 *(taken out)*
var [være] <mv> V IMPF AKT @FS-<ACC #14->4 *(was)*
både [både] ADV @FOC> #15->17 *(both)*
dyr [dyr] <cjt-head> <jval> ADJ UTR S IDF NOM @<SC #16->14 **R:p-cjt:18** **R:c-attr:8** *(expensive)*
og [og] <co-sc> KC @CO #17->16 *(and)*
dårlig [dårlig] <cjt> <jqual> ADJ UTR S IDF NOM @<SC #18->14 **R:c-cjt:16** **R:c-attr:8** *(bad)*
$. [.] PU @PU #19->0

(wordform [lemma] <secondary tags> POS INFLECTION @SYNTACTIC_FUNTION #id[dep]->id[head])

Secondary relations are appended as R: tags on both tokens involved in a (binary) relation, and contain a relation name followed by the id of the other token. *'R:c-attr:8'*, for instance, means the child end of an attributive relation, where 8 is the id of the (attributed) parent token. The latter here gets the same[10] relation name, but with a 'p-' (parent) prefix. Similarly, conjunction is tagged as *c/p-cjt*, and the subject and object relations as *subj* and *acc*, respectively.

It should be noted that all of the above are meant as primarily *syntactic* dependencies and that they are *secondary* in the sense that the child tokens in question unorthodoxically are allowed *two* dependency heads, where an ordinary syntactic tree would allow them either one or the other, but never both.

This is different from systematically adding a completely new, non-syntactic layer of dependency, as is the case when DanGram assigns semantic dependencies for frame- and role-carrying tokens (Bick 2011). In this case, a second, semantic tree is constructed, and the individual relations may or may not coincide with primary or secondary syntactic relations:

Den[den] <dem> DET UTR S @>N #1->2
forsikring [forsikring] <f-right> N UTR S IDF NOM @SUBJ> **R:sd-TH:5** **R:sd-TH:7** **§TH** #2->7
$, [,] PU @PU #3->0
han [han] PERS UTR 3S NOM @SUBJ> **R:sd-AG:5** **§AG** #4->5
tegnede [tegne] <fn:buy> <mv> V IMPF AKT @FS-N< **R:sd-ATR:2** **§ATR** #5->2
$, [,] PU @PU #6->0
var [være] <fn:be_copula> <mv> V IMPF AKT @FS-STA #7->0
meget [meget] <aquant> ADV @>A #8->9
dyr [dyr] <jval> <Deco> ADJ UTR S IDF NOM @<SC **R:sd-ATR:7** **§ATR** #9->7
$. [.] PU @PU #10->0
(sd=semantic dependency, fn: = framenet class, §AG=agent, §TH=theme, §ATR=attribute

6 Evaluation and statistics

Some preliminary, inspection-based[11], evaluation was carried out for four categories: (1) number propagation and disambiguation, (2) person-number tagging for finite verbs (from scratch), (3) future tense marking (<fut>) and (4) iterative marking (<iter>). In order to provide well-mixed attributes for these features, we used a section from Korpus 2010[12] containing blog/internet data.

	R	P	F
number	95.8%	100%	97.9
v pers/num	97.1%	97.5%	97.3
<fut>	83.3%	88.2%	85.7
<iter>	92.9%	78.8%	85.3

Table 2: Category tagging accuracy

Results indicate that the propagation and specification of morphological features (such as number and person) works best (F scores above 97%), most likely because they are mostly already inflection-marked on some other word in the sentence tree. Truly implicit features, that are never marked morphologically in Danish, are much harder to determine (F scores around 85%). Interestingly, <fut> suffered more from false negatives (low recall), while <iter> had more problems with false positives (low precision).

[10] CG3 allows arbitrary relation names, for both ends of a relation, so using prefixes and a common relation name is just a convention chosen here.

[11] Inspection is a fairly safe method for morphological categories, because there are few clear categories and clear morphological clues elsewhere in the sentence. The <fut> and <iter> categories are more likely to cause controversy in a multi-annotator scenario. As a "hard" criterion we plan to use the Greenlandic translation, that must make these categories explicit.

[12] Korpus 2010 was compiled by the Danish Society of Language and Literature (DSL) as part of the DK-CLARIN project.

7 On-the-fly corpus search markers

The focus of this paper has been the enrichment of an ordinary annotation run, feeding into an application (like MT) that needs implicit information made explicit (as tags) for specific purposes, or simply providing complete annotation of a category that does exist in the language in question, but is often left under-specified.

However, the same method can be put to a rather different use - on-the-fly marking of "corpus searchables". In this scenario, it is up to the (expert) user of an annotated corpus to formulate a search as a CG mapping rule, rather than an ordinary tag field query. In other words, the search engine interprets a 1-rule mini-CG at run time. Obviously, such a rule can exploit any existing annotation in a fully context-capable way, handling complexities far beyond any ordinary search. The example below presupposes a dependency tree and semantic role annotation, and also exploits two of the secondary tags introduced above (<fut> and <if>), but in principle, rules could even be written for raw text, using CG3's regular expression format.

"Find verbs with an experiencer subject, and check which ones are modified by conditional clauses in the future tense."

MAP (£mark) TARGET <mv> (*-1S VFIN LINK c @SUBJ + §EXP) (c @FS-<ADVL + <fut> LINK (c ("fordi")) OR (0 <if>)) ;

For smaller corpora, this is possible in real time, but for larger corpora, live processing and the ensuing impossibility of an optimized search structure (such as a database) means that search results cannot be piped to a GUI, but need to be written to a file for later inspection.

8 Conclusions and outlook

We have shown, how an existing CG annotation can be enriched without changing the original grammar, in a modular and application-driven fashion. Obviously, for both scenarios discussed here, corpus linguistics and machine translation, the choice of categories is task-dependent. For instance, propbank-style ARG0, ARG1, ... annotation could be added for a corpus user, and a different target language would require different categories. Thus, "non-factuality" is an inflection category in Romance languages (subjunctive), but not explicitly marked in Danish.

Future work should explore and evaluate which categories can be inferred from a standard (Danish) CG annotation with a reasonable level of accuracy, and which would need alterations in the original grammar or lexica.

References

Bick, Eckhard & Tino Didriksen. 2015. CG-3 - Beyond Classical Constraint Grammar. In: Beáta Megyesi: Proceedings of NODALIDA 2015, May 11-13, 2015, Vilnius, Lithuania. pp. 31-39. Linköping: LiU Electronic Press. ISBN 978-91-7519-098-3

Bick, Eckhard (2011). A FrameNet for Danish. In: Proceedings of NODALIDA 2011, May 11-13, Riga, Latvia. NEALT Proceedings Series, Vol 11, pp.34-41. Tartu: Tartu University Library. (ISSN 1736-6305)

Bick, Eckhard. 2001. En Constraint Grammar Parser for Dansk, in Peter Widell & Mette Kunøe (eds.), 8. Møde om Udforskningen af Dansk Sprog, 12.-13. oktober 2000, pp. 40-50, Århus University

Freitas, Cláudia & Rocha, Paulo & Bick, Eckhard. 2008. "Floresta Sintá(c)tica: Bigger, Thicker and Easier", in: António Teixeira et al. (eds.) Computational Processing of the Portuguese Language (Proceedings of PROPOR 2008, Aveiro, Sept. 8th-10th, 2008), pp.216-219. Springer