Comparison of XTAG and LEXSYS Grammars

Martine Smets
Cognitive and Computing Sciences
University of Sussex, Brighton BN1 9HQ, UK
martines@cogs.susx.ac.uk

1 Introduction

This paper presents work that forms part of the ongoing LEXSYS project on wide-coverage parsing,[1] and more precisely, some differences between our D-Tree grammar and XTAG 1995.

2 Grammar Formalism

We use the Lexicalised D-Tree Grammar (LDTG) formalism (Rambow et al. 95), which is based on the Lexicalized Tree Adjoining Grammar (LTAG) formalism. In LDTG, there are two types of edges between nodes: *d-edges*, represented with a broken line, and *p-edges*, represented by a solid line. Trees are combined by two substitution-like operations, both of which involve combining two descriptions, by equating exactly one node from each description. One of the operations is always used to add complements, and involves equating a frontier node (in the d-tree that is getting the complement) with the root of some component (in the d-tree that is providing the complement), such that the two nodes being equated are compatible . An example of substitution is shown in Figure 1.

The d-tree for *to adore* is composed with the d-tree for *seems* by equating the two nodes labelled VP[fin: –]. The top component of the *to adore* tree can then be fitted into the resulting d-tree by equating the root of the *seems* tree with the lower S of the *to adore* tree.

A second operation is used to add modifiers, but we are not going to discuss it in this paper.

[1]This work is supported by UK EPSRC project GR/K97400 'Analysis of Naturally-occurring English Text with Stochastic Lexicalized Grammars' (http://www.cogs.susx.ac.uk/lab/nlp/dtg/).

3 Differences between XTAG and LEXSYS Grammars

3.1 Trees Are Syntactic Representations

A first difference between our DTG and TAG is that we do not claim that elementary trees express in all cases the predicate-argument structure of their anchor; instead, they represent the syntactic requirements of their anchor. To illustrate, because raising verbs subcategorize for a syntactic subject, they anchor a standard verb tree with a subject, and not a tree rooted in VP without a subject, as in TAG. On the other hand, there are trees rooted in VP which represent VP complements and can be anchored by any verb. In those trees, there is no subject (because VP complements do not have syntactic subjects), and a semantic argument of the verb is thus missing.

This choice allows us to adopt other linguistic analyses than the ones supported by XTAG, as will be shown in the next sections.

3.1.1 Complementation and Long Distance Dependency

A main difference between the two grammars is that there are VP complements in our grammar, when there are only S complements in XTAG (except for auxiliaries and raising verbs). To the sentence in (1), our grammar gives the analysis in (1a), while XTAG gives the analysis in (1b).

```
(1a) [S He wants [VP to [VP come]]]
(1b) [S He wants
        [S [NP PRO] [VP to [VP come]]]]
```

In (1a), the complement of *want* is a VP; in (1b), it is an S, and the subject of the sentence is *PRO* (an empty pronominal). The analysis in (1a) is the one proposed in lexical theories

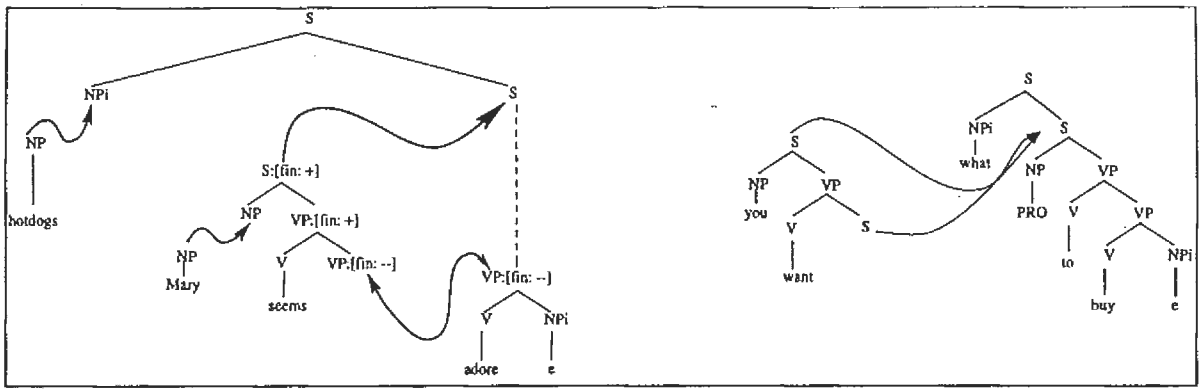

Figure 1: Example of unbounded dependency in DTG (left) and in TAG (right)

such as Head-Driven Phrase Structure Grammar (HPSG, Pollard and Sag 1994) and Lexical Functional Grammar (LFG, Bresnan 1982) while the analysis in (1b) is the analysis of Government and Binding (GB, Haegeman 1991).

Arguments given in the XTAG report for the representation in (1b) include a uniform treatment for indicative, infinitive and gerund embedded clauses (XTAG report 1995, 1998). This implies that both infinitive and gerunds are analyzed as having an empty subject, which is questionable, because there is no evidence for the existence of *PRO*[2] ; it is even more questionable for gerunds, which have the same distribution as NPs (this is true even for verbal gerunds), and can hardly be characterized as clauses (Malouf 1997 *inter alia*).

An important reason for XTAG to adopt the analysis in (1b) is that it seems to be the only type of analysis possible in that formalism (except if *equi* verbs like *want* anchor the elementary tree for raising verbs). This comes from the fact that unbounded dependencies which extend across more than one clause boundary are achieved through the use of auxiliary trees in XTAG: to derive the sentence in (2), an initial tree for *buy* is combined with an auxiliary tree for *want* (Figure 1).

(2) What do you want to buy?

[2] *PRO*, besides being unmotivated, creates theory-internal problems: XTAG has to define two different infinitive auxiliaries *to*, one which assigns the case *no case* (when the subject is *PRO*) and the other one which does not assign any case (when the complementizer *for* assigns accusative case to the subject). This distinction between two *to* is of course ad-hoc.

The auxiliary tree for *want* is grafted onto the lower S of the *buy* tree, and the recursivity of the process creates unbounded dependency. And because in auxiliary trees the root node and the foot node must be of the same category, verbs such as *want* cannot take a VP complement (assuming *want* anchors an S-tree).

In our grammar, on the other hand, there is no such restriction, and verbs can take S complements as well as VP complements. This decision to introduce VP complements in the grammar has a number of consequences (some of which are related to what was discussed in section 1):

- auxiliaries and raising verbs anchor the same tree family as other verbs which take VP complements;
- passive trees are rooted in VP;
- because trees for auxiliaries and raising verbs are rooted in S as any other verb tree, there are no predicative trees;
- the grammar has at least twice as many verb trees as XTAG 95 (each tree rooted in S has a counterpart rooted in VP), and in fact, more than that as we use multiple instances of the same tree to represent disjunctive feature values.

Each of these points will be addressed in the next sections.

3.2 Verbs of *Considering*

Another type of construction for which we assume the existence of a VP complement is the

Subject-to-Object Raising (SOR) structure illustrated in (3).

(3) We believe [Kim] [to be very smart]

In that analysis[3], which we adopt, raising verbs such as *believe* have two complements, an NP and a VP. In the XTAG analysis, SOR verbs have only one complement, a clause.

We assign the same kind of analysis to another type of verbs, referred to as *verbs of considering* in Pollard and Sag (1994): *consider* in (4) and *regard* in (5) have two complements, an NP and respectively an AP and a PP.

(4) I consider Jack quite intelligent
(5) We regard him as a nuisance

This analysis has been debated since the early seventies, and supported by a number of researchers, Pollard and Sag (1994) among others.

XTAG, on the other hand, adopt the GB analysis, which considers that verbs of *considering* and the like have only one complement, a small clause. Small clauses are *Ss* headed by an empty verb, and anchored by the complement of that verb (NP, PP or AP). This account is not without problems. First, it has to postulate an unmotivated empty verb position: there is no evidence that such a position should exist. Its purpose is to allow adjunction of raising and auxiliary verbs, but this is a purely technical device which is not supported by linguistic evidence.

A more important problem is the fact that verbs which take small clause complements must be able to constrain the small clause predicate: *consider* allows PPs, NPs and APs (6) while *prefer* allows PPs only (7).

(6)
We consider Kim a good teacher
We consider Kim quite good
We consider Kim out of his mind

(7)
*We prefer Kim a good teacher
*We prefer Kim quite good
We prefer Kim out of here

[3]The SOR analysis has been advocated with compelling arguments by Bresnan (1982), Postal and Pullum (1988) and Pollard and Sag (1994) *inter alia*.

Verbs who subcategorize for clausal complements cannot specify the subcategorization requirements of the verb in the complement clause; for example, there is no example of a verb like *say* which would stipulate what kind of complement the verb in its clausal complement should have. Accordingly, in the XTAG account, the clausal complement is not expanded, whether it is a standard clause or a small clause. But the data in (6) and (7) show that verbs of *considering* and the like do select the type of phrases which follow the NP; the solution adopted in XTAG is to use the feature *mode* (whose values are usually *indicative, imperative, subjunctive*, etc.) and to add to the range of features *nom* and *prep* (for NP and AP, and PP respectively). The verb *consider* selects an S which has a feature mode with value *nom/prep*, while *prefer* selects a small clause with *prep* as value for the feature mode. Of course, the decision to add these values to the range of values of the feature mode is ad-hoc, as they have nothing to do with verb mode, and are only a technical device to match the subcategorization requirements of the verb of *considering* with the actual category of the complement in the embedded small clause. Our solution, on the other hand, is straightforward: if the verb *consider* constrains the type of phrase that follows the NP it is because this phrase is also one of its complements.

Our choice of analysis, besides being straightforward and motivated by the data, also allows for a more uniform account of passive: the passive of verbs of *considering* and the like is handled by the same lexical rules as for other transitive verbs.

3.3 Auxiliaries and Raising Verbs

In XTAG, raising verbs and auxiliary verbs anchor the same auxiliary tree rooted in VP. In our grammar, on the other hand, those verbs anchor trees rooted in S, and belong to different families.

There have been debates in the literature about the status of auxiliary verbs, and several authors have argued that auxiliaries and modals should be considered as main verbs (Pullum and Wilson (1977), Gazdar et al. (1982)). Arguments include the fact that some auxiliaries behave also like main verbs (*be* and *have*, *ought*, *is* in *is to*), and the existence of semi-auxiliaries

(*need, used, dare* and *have to*[4]) which behave like main verbs in certain environments and like auxiliaries in other environments. So, the distinction between auxiliaries and main verbs is not clear-cut, and either the tree family for auxiliary verbs will include verbs which do not always behave like auxiliaries, or verbs classified as main verbs will share characteristics with auxiliary verbs. In both cases, the obvious solution is to abandon the distinction between main verbs and auxiliaries in terms of drastically different types of tree, and adopt instead a unified representation for both kinds of verbs.

A second issue is the fact that in the tree for auxiliaries and raising verbs, the complement of the anchor is a VP. This implies that all subject raising verbs subcategorize for VP, which is clearly not the case (*become* subcategorizes for AP or NP, *turn out* for AP, NP or VP). Thus, in order to get the right distribution of subcategorization, constraints on the complement of the raising verbs have to be expressed through percolation of the *mode* feature, which use has already been shown to be ad-hoc in similar instances.

3.3.1 Predicative Trees

There are no predicative trees in our grammar: this is a consequence of our decision to adopt a tree rooted in S for both raising verbs and auxiliaries. Also, we want a uniform treatment of predicative complements, and this would not be the case if we adopted different trees for predicative complements of verbs of *considering* and predicative complements of other types of verbs. So, predicative complements just substitute in the tree of their governing verb, like other types of complements.

A main criticism of our approach will be that the basic trees do not express all semantic relations: a predicative complement places semantic restrictions on the subject, and this cannot be captured in the basic trees, because predicative complements are substituted in the tree for the auxiliary/raising verb; similarly, for the VP complement trees, which do not have a subject[5]; finally, in the case of passive, the passive participle anchors a VP tree too, and the subject is not expressed either in the elementary tree.

We agree with this, but we do not claim that we can express every type of relation between constituents in basic trees; instead, we believe that it is impossible to capture all relevant information, syntactic and semantic, in the basic trees. We therefore adopt a modular representation, with the basic trees expressing mainly syntactic information, and the derivation tree most of the semantic information. We hope that this division of labour will allow us to express motivated syntactic analyses in the grammar, without having to compromise in order to also express at the same level semantic relations.

3.4 Conclusion

This paper has presented some differences between XTAG and the grammar we are developing in the LEXSYS project. It has shown that the DTG formalism gives us the possibility to adopt linguistic analyses which have proven to be more motivated than the GB ones (which can also be expressed with the same formalism).

The fact that we will have much more trees than TAGs might seem like a drawback to our approach. But Evans and Weir (1998) are exploring ways to allow a compact representation of the grammar for parsing purposes.

References

Bresnan, Joan. 1982. Control and complementation. In J. Bresnan, eds., *The Mental Representation of Grammatical Relations*.

Evans, Roger and David Weir. 1998. A structure-sharing parser for lexicalized grammars. In *COLING/ACL'98*.

Haegeman, Liliane. 1991. *Introduction to Government and Binding Theory*. Blackwell Publishers, Oxford, UK.

Malouf, Rob. 1996. A constructional approach to English verbal gerunds. In *Proceedings of the Twenty-second Meeting of the Berkeley Linguistics Society*, 1996, pages 255-266.

Pollard, C., and I. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago University Press, Chicago, IL.

[4]Actually, *have to* behaves like a main verb in all environments, but has a meaning very similar to *must*. This shows that which verbs are auxiliaries cannot be predicted from semantic information alone, as was noted by Pullum and Wilson (1977).

[5]I do not see any advantage of having *PRO* instead of no subject, though.

Pullum, Geoffrey and Paul Postal. 1988. Expletive noun phrases in subcategorized positions. In *Linguistic Inquiry*, 19, pages 635-670.

Pullum, Geoffrey and Deirdre Wilson. 1977. Autonomous syntax and the analysis of auxiliaires. In *Language*, v.53,4, pages 741-788.

Rambow, Owen, K. Vijay-Shanker and David Weir. 1995. Parsing D-Tree Grammars. In *Proceedings of the 33rd Meeting of the Association for Computational Linguistics (ACL '95)*, pages 151-158.

XTAG research group. 1995. A lexicalized Tree Adjoining Grammar of English, Technical Report IRCS 95-03, University of Pennsylvania.