Report on Ph.D. Thesis

Alessandro Mazzei: Formal and Empirical Issues of Applying Dynamics to Tree Adjoining Grammars

The thesis by Alessandro Mazzei presents a variant of Lexicalized Tree Adjoining Grammar (LTAG) called a "Dynamic Version of TAG" of DVTAG. DVTAG is motivated from psycholinguistic evidence, with additional motivation from theoretical syntax and natural language processing applications (including dialog systems). The bulk of the thesis is taken up by three chapters dealing with formal, linguistic, and empirical issues. In the formal chapter, Mazzei presents a formal definition of DVTAG and proves theorems relating to its formal powers in relation to CFG and TAG. He then discusses several different linguistic constructions, including the famous Dutch cross-serial dependencies, which lead him to propose an extension to DVTAG. Finally, he reports on some experiments deriving actual DVTAG grammars from other grammars and from corpora.

The thesis is very interesting and represents an important contribution to several fields of study. The most important contribution to the field is the exploration of a link between the TAG literature and the literature on incremental grammars and in particular "dynamic grammars" (for example Phillips, Milward, but also Categorial Grammarians such as Steedman). Both TAG and the literature on incremental parsing have contributed important insights into natural language syntax; this work lays the foundation for combining these insights. For example, it will be interesting to see whether other psycholinguistic experimental results can be explained using the DVTAG framework (for example, garden path effects). In addition, the definition of a new formalism in the TAG family may have a direct impact on natural language processing applications, as it may turn out to be a good formalism for broad-coverage parsers (though the thesis does not investigate this option in detail).

The author is to be specially commended for not just presenting a formal definitions and proofs, but also addressing specific linguistic issues, including the rather complex Dutch cross-serial dependencies. In addition, he has also addressed the issue of actually constructing large-scale DVTAG grammars, in different ways. Both of these aspects of his work (linguistic aspects and large-scale grammar development) are particularly important: a formalism intended for natural language syntax is useless if it has not been used to describe real natural language syntactic phenomena, including very tricky ones; and with the availability of annotated corpora, it is also important to get a sense of the adequacy of the formalism for the range of phenomena actually encountered in text. Thus, the reader of the thesis comes away with the firm impression that DVTAG is well thought out and worth taking seriously.

One slight regret is that Chapter 1, the introductory and motivating chapter, isn't longer. Many interesting issues are brought up (many of which I am not that familiar with), and more time could be spent on summarizing relevant experiments, discussing issues in theoretical syntax, and motivating the three fundamental assumptions. For example, can the Incrementality-in-Competence Hypothesis actually be tested empirically in some way? Does it reduce to the assumption of off-line left-association (see footnote 2 on page 149)? Questions of on-line vs. off-line are often considered issues of efficiency rather than fundamental issues -- why is it a different matter here? Or consider the Strong Connectivity Hypothesis. The claim that semantic interpretation is only possible on single syntactic constituents can seem somewhat stipulative. For example, why does the interpretation of two initial NPs in a head-final language require them to form a syntactic unit of any sort? Could the expectation of an object being transferred after hearing a nominative and a dative NP not come from the conceptual association from the individually interpreted NPs (with a dative NP always being interpreted as a recipient)? After all, waitresses typically do transfer hamburgers to customers. I do not raise these issues as serious objections, it would just have been nice to see a bit more discussion, and perhaps a summary of several other relevant experiments.

The literature is discussed thoroughly; the only slightly odd omission is Joshi 1990. While the paper is cited, it is only discussed in the context of the competence grammar for Dutch. However, this paper represents one of the few previous attempts to define an incremental processing model for TAG and to use it to make specific predictions which are verified by psycholinguistic experiments. This whole aspect is not discussed in Mazzei's thesis. While Joshi's approach is quite different from Mazzei's (in particular, no Strong Connectivity Hypothesis), it would be interesting to see a more detailed comparison. Can DVTAG make the right predictions? Also, while Categorial Grammar, and in particular Steedman's CCG, is discussed here and there in some detail, it would have been useful to have a comprehensive comparison of DVTAG and CCG.

Overall, the thesis is very well organized. The English is good (and I am grateful Mazzei chose to write it in English), but the thesis could profit from some general proofreading by the author.

In the following, I list some minor points.

P.11. In the shadowing experiment, it is actually unclear exactly what information is being used.

P.12 and elsewhere. The two NPs are "involved in a syntactic dependency" -- this is not the meaning of "dependency" used in "dependency tree". In fact, the exact nature of this "dependency" is fairly important, since it justifies (in part) the Strong Connectivity Hypothesis. This should be formulated more carefully and discussed in greater depth.

P.14. The support from theoretical syntax could be argued much more crisply. Is non-constituent coordination the only example? There have been solutions of non-constituent coordination which do not rely on incrementality -- what is wrong with them?

P.27. ATNs have recursion! Also, for CFGs as well you can say that there is no difference between derivable and parsable, as we have parsing algorithms which have been proven to be correct (i.e., they assign a derivation to those and only those strings which can be derived by the grammar). So it is not clear what claim is being made here.

P.31. The operations in TAG are not defined over pairs of elementary trees (if yes, we would never have derived trees composed of more than two elementary trees!).

p.31. Seven or eight operations?

P.34. Isn't the strong generative capacity usually defined as the set of derived trees, not derivation trees?

P.50. The notion of "grafting" is really too vague to be used in a definition without previous formal definition.

P.55. In 1, it is not clear which term is being defined. Is it really "left-context" (boldfaced)?

P.57. What is the parameter E in G(E)? How does it relate to the quadruple which G(E) is defined as being? Related to this, I suspect that in Theorem 2.17 (p.66) it is important that the "E" is the same in $G_1(E)$

and $G_2(E)$, but since this usage of E has not been defined, it is hard to tell what is going on.

P.66. In the proof of this theorem it seems that "if" and "only if" have been reversed.

P.77. Definition 2.19. D is not bound in the definition. Just above, I think one instance if $\gamma_1 < \gamma_2$ should be the reverse.

P.78. In the proof of Theorem 2.20, the arrows for the "if part" etc have been reversed.

P.83. The thesis never addresses the issue of what the definition of adjunction is -- the original definition, or the definition of Schabes and Shieber (1994), which is now widely used. Presumably the original definition.

P.93. The grammar given just generates {wew | w in $\{a,b\}^*$ }.

P.117. These constructions in Dutch are embedded clauses, NOT relative clauses! Many occurrences of this.

P.132. Xia and Palmer 2001 actually is an algo to go from dependency to phrase structure -- you need a different citation for Xia. Also, when citing work on extracting a TAG from the Treebank, please also cite John Chen (thesis, University of Delaware). He and Xia worked independently on the same problem at the same time.

P.143. It would be nice to actually see an explicit (and compact) reformulation of the CETM for DVTAG. (The discussion is clear, however.)

P.150, Fig 4.3 What kind of tree is the S-rooted tree on the last row of the figure?

P.161. The DVTAG grammar is still large (even when based on the extracted grammar), though much smaller. It would be easy enough to test coverage by continuing the procedure on the development corpus (Section 00) and seeing how many new trees are generated. For the LTAG extraction, about 0.5% of trees are unseen in Section 00 after training on Sections 02-21. Is this number increased for DVTAG?

In conclusion, based on reading his thesis, I can without hesitation recommend that Alessandro Mazzei be awarded the Ph.D. The thesis represents an important piece of scholarship, and an engaging contribution to the literature that will be of interest to several different communities, and that will impact several different domains (formal systems, theoretical syntax, natural language processing, psycholinguistics).

Owen Rambow

Research Scientist Center for Computational Learning Systems Columbia University 212 870 1282 www.cs.columbia.edu/~rambow rambow@cs.columbia.edu