Pro3Gres Parser in the CoNLL Domain Adaptation Shared Task

Gerold Schneider and **Kaarel Kaljurand** and **Fabio Rinaldi** and **Tobias Kuhn**
Institute of Computational Linguistics, University of Zurich
Binzmühlestrasse 14
CH - 8050 Zurich, Switzerland
{gschneid,kalju,rinaldi,tkuhn}@ifi.uzh.ch

Abstract

We present Pro3Gres, a deep-syntactic, fast dependency parser that combines a hand-written competence grammar with probabilistic performance disambiguation and that has been used in the biomedical domain. We discuss its performance in the domain adaptation open submission. We achieve average results, which is partly due to difficulties in mapping to the dependency representation used for the shared task.

1 Introduction

The Pro3Gres parser is a dependency parser that combines a hand-written grammar with probabilistic disambiguation. It is described in detail in (Schneider, 2007). It uses tagger and chunker pre-processors – parsing proper happens only between heads of chunks – and a post-processor graph converter to capture long-distance dependencies. Pro3Gres is embedded in a flexible XML pipeline. It has been applied to many tasks, such as parsing biomedical literature (Rinaldi et al., 2006; Rinaldi et al., 2007) and the whole British National Corpus, and has been evaluated in several ways. We have achieved average results in the CoNLL domain adaptation track open submission (Marcus et al., 1993; Johansson and Nugues, 2007; Kulick et al., 2004; MacWhinney, 2000; Brown, 1973). The performance of the parser is seriously affected by mapping problems to the particular dependency representation used in the shared task.

The paper is structured as follows. We give a brief overview of the parser and its design policy in section 2, we describe the domain adaptations that we have used in section 3, comment on the results obtained in section 4 and conclude in section 5.

2 Pro3Gres and its Design Policy

There has been growing interest in exploring the space between Treebank-trained probabilistic grammars (e.g. (Collins, 1999; Nivre, 2006)) and formal grammar-based parsers integrating statistics (e.g. (Miyao et al., 2005; Riezler et al., 2002)). We have developed a parsing system that explores this space, in the vein of systems like (Kaplan et al., 2004), using a linguistic *competence* grammar and a probabilistic *performance* disambiguation allowing us to explore interactions between lexicon and grammar (Sinclair, 1996). The parser has been explicitly designed to be deep-syntactic like a formal grammar-based parser, by using a dependency representation that is close to LFG f-structure, but at the same time mostly context-free and integrating shallow approaches and aggressive pruning in order to keep search-spaces small, without permitting compromise on performance or linguistic adequacy. (Abney, 1995) establishes the chunks and dependencies model as a well-motivated linguistic theory. The non-local linguistic constraints that a hand-written grammar allows us to formulate, e.g. expressing X-bar principles or barring very marked constructions, further reduce parsing time by at least an order of magnitude. Since the grammar is on Penn tags (except for few closed classed words, e.g. allowing *including* to function as preposition) the effort for writing it manually is manageable. It has been developed from scratch in about a person month,

Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL 2007, pp. 1161–1165,
Prague, June 2007.

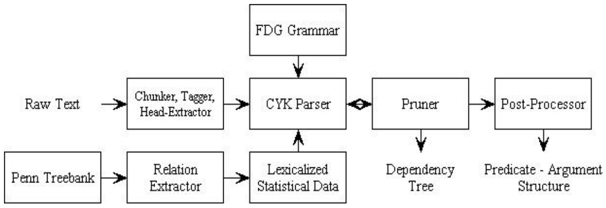

Figure 1: Pro3Gres parser flowchart

using traditional grammar engineering development cycles. It contains about 1000 rules, the number is largely so high due to tag combinatorics: for example, the various subject attachment rules combining a subject (*_NN, _NNS, _NNP, _NNPS*) and a verb (*_VBZ, _VBP, _VBG, _VBN, _VBD*) are all very similar.

The parser is fast enough for large-scale application to unrestricted texts, and it delivers dependency relations which are a suitable base for a range of applications. We have used it to parse the entire 100 million words British National Corpus (http://www.natcorp.ox.ac.uk) and similar amounts of biomedical texts. Its parsing speed is about 500,000 words per hour. The flowchart of the parser can be seen in figure 1.

Pro3Gres (PRObabilistic PROlog-implemented RObust Grammatical Role Extraction System) uses a dependency representation that is close to LFG f-structure, in order to give it an established linguistic background. It uses post-processing graph structure conversions and mild context-sensitivity to capture long-distance dependencies. We have argued in (Schneider, 2005) that LFG f-structures can be parsed for in a completely context-free fashion, except for embedded WH-questions, where a device such as functional uncertainty (Kaplan and Zaenen, 1989) or the equivalent Tree-Adjoining Grammar Adjoining operation (Joshi and Vijay-Shanker, 1989) is used. In Dependency Grammar, this device is also known as *lifting* (Kahane et al., 1998; Nivre and Nilsson, 2005).

We use a hand-written competence grammar, combined with performance-driven disambiguation obtained from the Penn Treebank (Marcus et al., 1993). The Maximum-Likelihood Estimation (MLE) probability of generating a dependency relation R given lexical heads (a and b) at distance (in chunks) δ is calculated as follows.

$$p(R,\delta|a,b) \cong p(R|a,b)\cdot p(\delta|R) = \frac{\#(R,a,b)}{\sum_{i=1}^{n}\#(R_i,a,b)}\cdot\frac{\#(R,\delta)}{\#R}$$

The counts are backed off (Collins, 1999; Merlo and Esteve Ferrer, 2006). The backoff levels include semantic classes from WordNet (Fellbaum, 1998): we back off to the lexicographer file ID of the most frequent word sense. An example output of the parser is shown in figure 2.

3 Domain Adaptation

Based on our experience with parsing texts form the biomedical domain, we have used the following two adaptations to the domain of chemistry.

(Hindle and Rooth, 1993) exploit the fact that in sentence-initial *NP PP* sequences the PP unambiguously attaches to the noun. We have observed that in sentence-initial *NP PP PP* sequences, also the second PP frequently attaches to the noun, the noun itself often being a relational noun. We have thus used such sequences to learn relational nouns from the unlabelled domain texts. Relational nouns are allowed to attach several argument PPs in the grammar, all other nouns are not.

Multi-word terms, adjective-preposition constructions and frequent PP-arguments have strong collocational force. We have thus used the collocation extraction tool XTRACT (Smadja, 2003) to discover collocations from large domain corpora. The probability of generating a dependency relation is augmented for collocations above a certain threshold. Since the tagging quality of the Chemistry testset is high, the impact of multi-word term recognition was lower than the biomedical domain when using a standard tagger, as we have shown in (Rinaldi et al., 2007).

For the CHILDES domain, we have not used any adaptation. The hand-written grammar fares quite well on most types of questions, which are very frequent in this domain. In the spirit of the shared task, we have not attempted to correct tagging errors, which were frequent in the CHILDES domain. We have restricted the use of external resources to the hand-written, domain-independent grammar, and to WordNet. Due to serious problems in mapping our

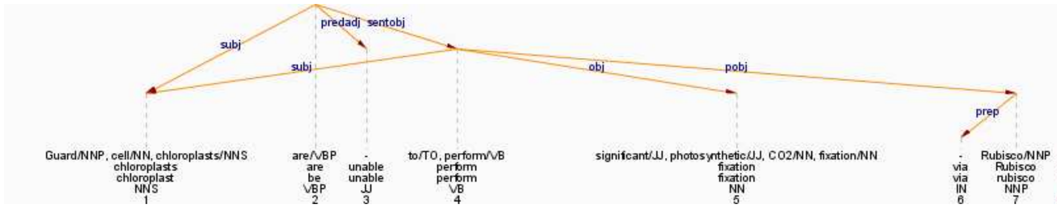

Figure 2: Example of original parser output

LFG f-structure based dependencies to the CoNLL representation, much less time than expected was available for the domain adaptation.

4 Our Results

We have achieved average results: Labeled attachment score: 3151 / 5001 * 100 = 63.01, unlabeled attachment score: 3327 / 5001 * 100 = 66.53, label accuracy score: 3832 / 5001 * 100 = 76.62. These results are about 10 % below what we typically obtain when using our own dependency representation or GREVAL (Carroll et al., 2003), a deep-syntactic annotation scheme that is close to ours. Detailed evaluations are reported in (Schneider, 2007). Our mapping was quite poor, especially when conjunctions are involved. Also punctuation is attached poorly. 5.7 % of all dependencies remained unmapped (*unknown* in the figure). We give an overview of the the relation-dependent results in figures 1 and 2.

Mapping problems include the following examples. First, headedness is handled very differently: while we assume auxiliaries, prepositions and coordinations to be dependents, the CoNNL representation assumes the opposite, which leads to incorrect mapping under complex interactions. Second, the semantics of parentheticals (*PRN*) partly remains unclear. In *Quinidine elimination was capacity limited with apparent Michaelis constant (appKM) of 2.6 microM (about 1.2 mg/L)* the gold standard annotates the second parenthesis as parenthetical, but the first as nominal modification, although both may be said to have appositional character. Third, we seem to have misinterpreted the roles of *ADV* and *AMOD*, as they are often mutually exchanged. Fourth, the logical subject (*LGS*) is sometimes marked on the by-PP (... *are strongly inhibited by-LGS carbon monoxide*) and sometimes on the participle (... *are increased-LGS by pretreatment*) in the gold standard. Relations between heads of chunks, which are central for predicate-argument structures which Pro3Gres aims to recover, such as *SBJ, NMOD, ROOT*, perform better than those for which Pro3Gres was not originally designed, particularly *ADV, AMOD, PRN, P*. Performance on *COORD* was particularly disappointing. Generally, mapping problems between different representations would be smaller if one used a dependency representation that maximally abstracts away from form to function, for example (Carroll et al., 2003).

deprel	gold	correct	system	recall (%)	prec. (%)
ADV	366	212	302	57.92	70.20
AMOD	87	8	87	9.20	9.20
CC	11	0	0	0.00	NaN
COORD	402	233	342	57.96	68.13
DEP	9	0	0	0.00	NaN
EXP	2	0	0	0.00	NaN
GAP	14	0	0	0.00	NaN
IOBJ	3	0	0	0.00	NaN
LGS	37	0	0	0.00	NaN
NMOD	1813	1576	1763	86.93	89.39
OBJ	185	146	208	78.92	70.19
P	587	524	525	89.27	99.81
PMOD	681	533	648	78.27	82.25
PRN	34	13	68	38.24	19.12
ROOT	195	138	190	70.77	72.63
SBJ	279	217	296	77.78	73.31
VC	129	116	136	89.92	85.29
VMOD	167	116	149	69.46	77.85
unknown	0	0	287	NaN	0.00

Table 1: Prec.&recall of DEPREL

We have obtained results slightly above average on the CHILDES domain, although we did not adapt the parser to this domain in any way (unlabeled attachment score: 3013 / 4999 * 100 = 60.27 %). The hand-written grammar, which includes rules for most types of questions, fares relatively well on this domain since questions are rare in the Penn Treebank (see (Hermjakob, 2001)). Pro3Gres has been employed for question parsing at a TREC conference (Burger and Bayer, 2005).

deprel	gold	correct	system	recall (%)	prec. (%)
ADV	366	161	302	43.99	53.31
AMOD	87	5	87	5.75	5.75
CC	11	0	0	0.00	NaN
COORD	402	170	342	42.29	49.71
DEP	9	0	0	0.00	NaN
EXP	2	0	0	0.00	NaN
GAP	14	0	0	0.00	NaN
IOBJ	3	0	0	0.00	NaN
LGS	37	0	0	0.00	NaN
NMOD	1813	1392	1763	76.78	78.96
OBJ	185	140	208	75.68	67.31
P	587	221	525	37.65	42.10
PMOD	681	521	648	76.51	80.40
PRN	34	12	68	35.29	17.65
ROOT	195	138	190	70.77	72.63
SBJ	279	190	296	68.10	64.19
VC	129	116	136	89.92	85.29
VMOD	167	85	149	50.90	57.05
unknown	0	0	287	NaN	0.00

Table 2: Prec.&recall of DEPREL+ATTACHMENT

5 Conclusion

We have described the Pro3Gres parser. We have achieved average results in the shared task with relatively little adaptation. Mapping to different representations is an often underestimated task. Our performance on the CHILDES task, where we did not adapt the parser, indicates that hand-written, carefully engineered *competence* grammars may be relatively domain-independent while *performance* disambiguation is more domain-dependent. We will adapt the parser to further domains and include more unsupervised learning methods.

References

Steven Abney. 1995. Chunks and dependencies: Bringing processing evidence to bear on syntax. In Jennifer Cole, Georgia Green, and Jerry Morgan, editors, *Computational Linguistics and the Foundations of Linguistic Theory*, pages 145–164. CSLI.

R. Brown. 1973. *A First Language: The Early Stages*. Harvard University Press.

John D. Burger and Sam Bayer. 2005. MITRE's Qanda at TREC-14. In E. M. Voorhees and Lori P. Buckland, editors, *The Fourteenth Text REtrieval Conference (TREC 2005) Notebook.*

John Carroll, Guido Minnen, and Edward Briscoe. 2003. Parser evaluation: using a grammatical relation annotation scheme. In Anne Abeillé, editor, *Treebanks: Building and Using Parsed Corpora*, pages 299–316. Kluwer, Dordrecht.

Michael Collins. 1999. *Head-Driven Statistical Models for Natural Language Parsing*. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Ulf Hermjakob. 2001. Parsing and question classification for question answering. In *Proceedings of the ACL 2001 Workshop on Open-Domain Question Answering*, Toulouse, France.

Donald Hindle and Mats Rooth. 1993. Structural ambiguity and lexical relations. *Computational Linguistics*, 19:103–120.

R. Johansson and P. Nugues. 2007. Extended constituent-to-dependency conversion for English. In *Proc. of the 16th Nordic Conference on Computational Linguistics (NODALIDA).*

Aravind K. Joshi and K. Vijay-Shanker. 1989. Treatment of long-distance dependencies in LFG and TAG: Functional uncertainty in LFG is a corollary in TAG. In *Proceedings of ACL '89.*

Sylvain Kahane, Alexis Nasr, and Owen Rambow. 1998. Pseudo-projectivity: A polynomially parsable non-projective dependency grammar. In *Proceedings of COLINGACL*, volume 1, pages 646–652, Montreal.

Ronald Kaplan and Annie Zaenen. 1989. Long-distance dependencies, constituent structure, and functional uncertainty. In Mark Baltin and Anthony Kroch, editors, *Alternative Concepts of Phrase Structrue*, pages 17 – 42. Chicago University Press.

Ron Kaplan, Stefan Riezler, Tracy H. King, John T. Maxwell III, Alex Vasserman, and Richard Crouch. 2004. Speed and accuracy in shallow and deep stochastic parsing. In *Proceedings of HLT/NAACL 2004*, Boston, MA.

S. Kulick, A. Bies, M. Liberman, M. Mandel, R. McDonald, M. Palmer, A. Schein, and L. Ungar. 2004. Integrated annotation for biomedical information extraction. In *Proc. of the Human Language Technology Conference and the Annual Meeting of the North American Chapter of the Association for Computational Linguistics (HLT/NAACL).*

B. MacWhinney. 2000. *The CHILDES Project: Tools for Analyzing Talk.* Lawrence Erlbaum.

M. Marcus, B. Santorini, and M. Marcinkiewicz. 1993. Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics*, 19(2):313–330.

Paola Merlo and Eva Esteve Ferrer. 2006. The notion of argument in PP attachment. *Computational Linguistics*, 32(2):341 – 378.

Yusuke Miyao, Takashi Ninomiya, and Jun'ichi Tsujii. 2005. Corpus-oriented grammar development for acquiring a Head-driven Phrase Structure Grammar from the Penn Treebank. In Keh-Yih Su, Jun'ichi Tsujii, Jong-Hyeok Lee, and Oi Yee Kwong, editors, *Natural Language Processing - IJCNLP 2004*, pages 684–693. Springer.

Joakim Nivre and Jens Nilsson. 2005. Pseudo-projective dependency parsing. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 99–106, Ann Arbor, Michigan, June. Association for Computational Linguistics.

Joakim Nivre. 2006. Constraints on non-projective dependency parsing. In *Proceedings of the European Chapter of the Association of Computational Linguistics (EACL) 2006*, pages 73 – 80, Trento, Italy. Association for Computational Linguistics.

Stefan Riezler, Tracy H. King, Ronald M. Kaplan, Richard Crouch, John T. Maxwell, and Mark Johnson. 2002. Parsing the Wall Street Journal using a Lexical-Functional Grammar and discriminative estimation techniques. In *Proc. of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*, Philadephia, PA.

Fabio Rinaldi, Gerold Schneider, Kaarel Kaljurand, Michael Hess, and Martin Romacker. 2006. . an environment for relation mining over richly annotated corpora: the case of GENIA. *BMC Bioinformatics*, 7(Suppl 3):S3.

Fabio Rinaldi, Gerold Schneider, Kaarel Kaljurand, Michael Hess, Christos Andronis, Ourania Konstanti, and Andreas Persidis. 2007. Mining of functional relations between genes and proteins over biomedical scientific literature using a deep-linguistic approach. *Journal of Artificial Intelligence in Medicine*, 39:127 – 136.

Gerold Schneider. 2005. A broad-coverage, representationally minimal LFG parser: chunks and F-structures are sufficient. In Mriram Butt and Traci Holloway King, editors, *The 10th international LFG Conference (LFG 2005)*, Bergen, Norway. CSLI.

Gerold Schneider. 2007. *Hybrid Long-Distance Functional Dependency Parsing*. Doctoral Thesis, Institute of Computational Linguistics, University of Zurich. accepted for publication.

John Sinclair. 1996. The empty lexicon. *International Journal of Corpus Linguistics*, 1, 1996.

Frank Smadja. 2003. Retrieving collocations from text: Xtract. *Computational Linguistics*, 19:1, Special issue on using large corpora:143–177.