Closing the Gap: Cognitively Adequate, Fast Broad-Coverage Grammatical Role Parsing

Gerold Schneider, Fabio Rinaldi, Kaarel Kaljurand and Michael Hess

Institute of Computational Linguistics, University of Zurich, Switzerland

Abstract. We present Pro3Gres, a fast robust broad-coverage and deep-linguistic parser that has been applied to and evaluated on unrestricted amounts of text from unrestricted domains. We show that it is largely cognitively adequate. We argue that Pro3Gres contributes to closing the gap between psycholinguistics and language engineering, between probabilistic parsing and formal grammar-based parsing, between shallow parsing and full parsing, and between deterministic parsing and non-deterministic parsing. We also describe a successful application of Pro3Gres for parsing research texts from the BioMedical domain.

1 Introduction

Computational psycholinguistics and language engineering are often seen as distinct research activities. Engineering aims at practical, fast solutions. Computational psycholinguistics often concentrates on the detailed modeling of few well-understood constructions. But recent approaches such as [1–3] have shown that human sentence processing and statistical parsing technology share major objectives: to rapidly and accurately understand the text and utterances they encounter.

[1] shows that true probabilities are essential for a cognitive model of sentence processing of access and disambiguation. *Access* – retrieving linguistic structure from some mental grammar – directly corresponds to a low probability threshold that cuts possible but improbable readings. *Disambiguation* – choosing among combinations of structures – happens as structures are put together and corresponds to a summing of parsing decisions in which only the few most probable analyses are kept due to memory limitations. In garden path sentences the fact that the human reader has cut low probability structures leads to reanalysis, indicated by longer human reading times and by re-reading in eye tracking experiments.

[2] show that statistical modeling not only allows us to model which sentences humans find difficult to analyze, but equally why we parse the majority of sentences with little effort and mental load but high accuracy. [1] offers a first probabilistic psycholinguistic parsing model. As with other psycholinguistic models, however, its syntactic coverage, scalability and performance remain unclear and unproven. [4, 3] is the first broad-coverage psycholinguistic parsing model, aiming at closing the gap between language engineering and psycholinguistics. Their approach is psycholinguistically appealing as it is left-to-right incremental and as it calculates local decision probabilities, emphasizing parsing as a decision process. While broad-coverage and performant, the

Schneider G., Rinaldi F., Kaljurand K. and Hess M. (2005).
Closing the Gap: Cognitively Adequate, Fast Broad-Coverage Grammatical Role Parsing.
In *Proceedings of the 2nd International Workshop on Natural Language Understanding and Cognitive Science*, pages 178-184
DOI: 10.5220/0002574901780184

layered approach of cascaded Markov models is not a full parsing approach. We present a full parsing approach similar in spirit to [3]. The paper is structured as follows: we first take a psycholinguistic (section 2) and then a language engineering perspective (section 3) on our parser. Then we discuss evaluations and applications (section 4).

2 Psycholinguistic Adequacy

The Probability Model We explain Pro3Gres' probability model by comparing it to Collins' parser Model 1 [5]. [6] discusses this relation in more detail. Both Collins' Model 1 and Pro3Gres are mainly dependency-based maximum-likelihood (MLE) statistical parsers parsing over heads of chunks. Collins' MLE and the Pro3Gres MLE can be juxtaposed as follows[1]. Collins' MLE estimation: $P(R|\langle a, atag\rangle, \langle b, btag\rangle, dist) \cong$

$$\frac{\#(R, \langle a, atag\rangle, \langle b, btag\rangle, dist)}{\#(\langle a, atag\rangle, \langle b, btag\rangle, dist)} \tag{1}$$

Main Pro3Gres MLE estimation [7]: $P(R, dist|a, b) \cong p(R|a, b) \cdot p(dist|R) \cong$

$$\frac{\#(R, a, b)}{\sum_{i=1}^{n} \#(R_i, a, b)} \cdot \frac{\#(R, dist)}{\#R} \tag{2}$$

The psycholinguistically most relevant differences are: (1) The co-occurrence count in the MLE denominator is not the sentence-context, but the sum of competing relations. For example, the $object$ and the $adjunct$ relation are in competition, as they are licensed by the same tag configuration. Pro3Gres models attachment probabilities, local decision probabilities like [1, 3], in accordance with the psycholinguistic view of parsing as a decision process. (2) Relations (R) have a Functional Dependency Grammar definition [8], including long-distance dependencies (see section 3). Grammatical Relations are psycholinguistically intuitive and easily mappable to predicate-argument structures.

Incrementality and Parallelism Some authors (e.g. [9]) argue that left-to-right incrementality needs to be strictly imposed, others favour a deferring approach in which structures are not fully connected at all times [10]. [11] has shown that dependency parsing cannot be fully incremental, but is so in the majority of cases, and delays are typically very short.

An issue related to incrementality is the question whether alternative analyses should be considered at the same time (parallel) or defered to a later stage (serial), e.g. by means of backtracking. Computational parsing models are usually based on some form of dynamic programming algorithm, and hence are parallel in nature. Traditionally, psycholinguistic parsers are serial like e.g. the shift-reduce algorithm, equating psycholinguistic re-analysis and backtracking of the algorithm in garden path situations. The CKY algorithm, which we use, can be seen as parallelized shift-reduce. [1] discusses that garden-path phenomena can also be modeled with a parallel architecture, and that some psycholinguistic results argue for parallel parsing [12, 13].

[1] R= Gram. Relation; a, b = head lemmas, $atag, btag$ = their POS tags, $dist$ = distance (chunks)

Modularity The issue of modularity is a fiercely debated topic in the sentence processing literature. [14] argue for a pre-syntactic module that is responsible for lexical category decisions. They show that a unique categorial label is initially preferred, which leads to augmented reading times in garden path situations. If we follow their argument then a tagging preprocessing step and a chunking preprocessing step [15] is psycholinguistically adequate in addition to increasing the parsing efficiency.

Traces In the Mind? [1] argues in favour of a traceless theory, which has the advantage of being representationally more minimal and not positing empty categories in the mind. We follow his argument (see section 3). A traceless theory also allows for a modular analysis of e.g. passives, in which no increased memory load or reading times for constructing a trace is reported. The issue whether traceless theories are psycholinguistically preferable remains contested, however [16].

Re-analysis Sample Experiment A sample experiment focusses on garden path situations. In them, a locally most plausible interpretation needs to be revised due to subsequent text data so that a globally possible interpretation can be found. In a statistical parser this is conveyed by a locally relatively unlikely interpretation becoming the most likely one at a later stage. If we parse using short beam lengths, locally least probable analyses get lost. Without a repair mechanism such as backtracking the globally correct interpretation cannot be reached when using the short beam[2]. In a widely used 500 sentence evaluation corpus ([17], see 4), although true garden path sentences are rare, 13 sentences get less correct analyses in the short beam scenario, e.g.

> Mitchell$_1$ said$_2$ [the Meiner administration]$_3$ and$_4$ [the Republican]$_5$ controlled$_6$ [State Senate]$_7$ share$_8$ [the blame]$_9$

Comparing the parse chart entries reveals that an $object$ relation between *controlled* (at position 6) and *Senate* (at position 7) is about 20 times more likely than an $adjective$ relation. The chart spans from *Republican* (position 5) to *Senate* (position 7) leading to the correct global span additionally include the rare $nchunk$ relation – a relation that corrects chunking shortcomings. The chart entry containing a $subject$ relation to *Republican* and an $object$ relation to *Senate* is 210 times more likely than the $nchunk$ plus $adjective$ reading that leads to the globally correct span. If aggressive pruning such as short beam is used at this stage, no global span can be found by the parser: the parse fails, corresponding to a situation that triggers a human parser to re-analyse.

3 Fast and Robust Grammatical Role Parsing

From a language engineering perspective, Pro3Gres has been designed to keep search spaces and parsing complexity low while only taking minimal linguistic compromises [18] and to be robust for broad-coverage parsing [19]. In order to keep parsing complexity as low as possible, aggressive use of pruning, shallow techniques and context-free parsing is made.

[2] We used 6 alternatives per span as a normal beam and 2 alternatives per span as a small beam

Pruning As predicted by [1, 2] parsing speed increases tremendously while the performance is hardly affected until extreme pruning parameters are used (see section 4 for evaluation details). Exploring aggressive pruning strategies bridges the gap between full parsing and deterministic parsing by e.g. [20].

Tagging and Chunking Low-level linguistic tasks that can be reliably solved by finite-state techniques, tagging and chunking, are handed over to them [15, 18]. Such an approach implements the modular hypothesis introduced by [14]. Parsing takes place only between the heads of chunks, and only using the best tag suggested by the tagger. In a test with a toy NP and verb-group grammar parsing was about 4 times slower.

Hand-written Grammar Combining a hand-written grammar with statistical disambiguation reflects our view of grammar as rule-based *competence* and disambiguation as statistical *performance*. Writing grammar rules is an easy task for a linguist, particularly when using a framework that is close to traditional school grammar assumptions, such as Dependency Grammar (DG) [8].

Long-Distance Dependencies Treating long-distance dependencies is very costly as they are context-sensitive. Classical statistical Treebank trained parsers thus fully or largely ignore them. [21] presents a pattern-matching algorithm for post-processing the Treebank output of such parsers to add empty nodes expressing long-distance dependencies to their parse trees. Encouraging results are reported for perfect parses, but performance drops considerably when using parser output trees. We have applied structural patterns to the Treebank, patterns similar to [21]'s but relying on functional labels and empty nodes thus reaching near-full precision. We use the extracted lexical counts as training material. Every dependency relation has a group of structural extraction patterns associated with it.

Movements are generally supposed to be of arbitrary length, but a closer investigation reveals that many types of movement are fixed and can thus be replaced by a single, local dependency. This is most obvious for *passive* and *control*, but [7] explains how most long-distance dependencies except for complex WH-movement can be modeled locally in DG[3]. The resulting DG trees are flatter which has the advantages that less and less sparse decisions are needed at parse-time, and that the costly overhead for dealing with unbounded dependencies can be largely avoided.

4 Evaluations and Applications

Pro3Gres has been evaluated widely, using dependency-oriented evaluation [17]. We report results on the 500 sentence Carroll corpus [17] and on 100 random sentences from the BioMedical GENIA corpus [22] in 1. Results on the GENIA corpus are particularly good because near-perfect terminology and thus improved chunking information is available, other evaluations are affected by remaining mapping problems, e.g. differing grammar assumptions, imperfect tagging and chunking.

[3] We use a mildly-context-sensitve TAG approach for complex WH-movement

Table 1. Percentage results of evaluating on Carroll's test corpus and on GENIA on subject, object and PP-attachment relations

Percentages on CARROLL	Subject	Object	noun-PP	verb-PP
Precision	91.5	90.3	70.5	72.5
Recall	80.6	83.4	64.0	86.4
Percentages on GENIA	Subject	Object	noun-PP	verb-PP
Precision	90	93	85	82
Recall	87	91	82	84

Pro3Gres has been used in a number of Text Mining applications over Biomedical literature. In [23] we describe experiments over the GENIA corpus, aimed at detecting domain-relevant semantic relations. GENIA [22][4] is a corpus of 2000 MEDLINE abstracts. Pro3Gres is the core component of the applications targeted within the OntoGene project (`http://www.ontogene.org/`), which aims at evaluating the hypotheses that high-precision hybrid parsing technologies have reached a sufficient level of maturity to become usable in practical large-scale Text Mining efforts. A significant early result has been the release of **DepGENIA**[5]: a corpus of Dependency Annotations, which is an enriched version of the GENIA corpus, built using Pro3Gres. Applications of Pro3Gres for Question Answering (QA) and Knowledge Management are discussed in [24, 25].

5 Conclusions

[26] compare speed and accuracy of [5] to a robust LFG system based on [27]. They show that the gap between probabilistic context-free parsing and deep-linguistic parsing can be closed. A conclusion that can be drawn from their and our work is that research in simplifying, restricting and limiting Formal Grammar expressiveness is increasingly bridging the gap between probabilistic and formal grammar-based parsing, between shallow and full parsing, and between deterministic and non-deterministic parsing.

We have presented a widely applied, fast robust broad-coverage and deep-linguistic parser that contributes to bridging these gaps. We have argued that it is largely cognitively adequate and thus additionally contributes to bridging the gap between psycholinguistics and language engineering.

References

1. Jurafsky, D.: A probabilistic model of lexical and syntactic access and disambiguation. Cognitive Science (1996)
2. Brants, T., Crocker, M.: Probabilistic parsing and psychological plausibility. In: Proceedings of 18th International Conference on Computational Linguistics COLING-2000, Saarbrücken/Luxembourg/Nancy (2000)

[4] `http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/`
[5] `www.ifi.unizh.ch/CL/kalju/download/depgenia/`

3. Crocker, M., Brants, T.: Wide coverage probabilistic sentence processing. Journal of Psycholinguistic Research **29(6)** (2000) 647–669
4. Brants, T.: Cascaded markov models. In: Proceedings of the Ninth Conference of the European Chapter of the Association for Computational Linguistics (EACL'99), Bergen, Norway, University of Bergen (1999) 118–125
5. Collins, M.: Head-Driven Statistical Models for Natural Language Parsing. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA (1999)
6. Schneider, G., Rinaldi, F., Dowdall, J.: Fast, deep-linguistic statistical dependency parsing. In: Coling 2004 Workshop on Recent Advances in Dependency Grammar, Geneva, Switzerland, August 2004 (2004)
7. Schneider, G.: Extracting and using trace-free Functional Dependencies from the Penn Treebank to reduce parsing complexity. In: Proceedings of Treebanks and Linguistic Theories (TLT) 2003, Växjö, Sweden (2003)
8. Tesnière, L.: Eléments de Syntaxe Structurale. Librairie Klincksieck, Paris (1959)
9. Sturt, P.: Incrementality in syntactic processing: Computational models and experimental evidence. In Keller, F., Clark, S., Crocker, M., Steedman, M., eds.: Proceedings of the ACL Workshop Incremental Parsing: Bringing Engineering and Cognition Together, Barcelona, Spain, Association for Computational Linguistics (2004) 66–66
10. Mulders, I.: Transparent parsing: Head-driven processing of verb-final structures. PhD thesis, Netherland's Graduate School of Linguistics (2002)
11. Nivre, J.: Incrementality in deterministic dependency parsing. In Keller, F., Clark, S., Crocker, M., Steedman, M., eds.: Proceedings of the ACL Workshop Incremental Parsing: Bringing Engineering and Cognition Together, Barcelona, Spain, Association for Computational Linguistics (2004) 50–57
12. Boland, J.E.: The Use of Lexical Knowledge in Sentence Processing. PhD thesis, University of Rochester, Rochester (1991)
13. MacDonald, M.: The interaction of lexical and syntactic ambiguity. Journal of Memory and Language **32** (1993) 692–715
14. Crocker, M., Corley, S.: Modular architectures and statistical mechanisms: The case from lexical category disambiguation. In Merlo, P., Stevenson, S., eds.: The Lexical Basis of Sentence Processing: Formal, Computational and Experimental Issues. John Benjamins, Amsterdam (2002)
15. Abney, S.: Chunks and dependencies: Bringing processing evidence to bear on syntax. In Cole, J., Green, G., Morgan, J., eds.: Computational Linguistics and the Foundations of Linguistic Theory, CSLI (1995) 145–164
16. Lee, M.W.: Another look at the role of empty categories in sentence processing (and grammar). Journal of Psycholinguistic Research **33(1)** (2004) 51 – 73
17. Carroll, J., Minnen, G., Briscoe, T.: Corpus annotation for parser evaluation. In: Proceedings of the EACL-99 Post-Conference Workshop on Linguistically Interpreted Corpora, Bergen, Norway (1999)
18. Schneider, G.: Combining shallow and deep processing for a robust, fast, deep-linguistic dependency parser. In: ESSLLI 2004 Workshop on Combining Shallow and Deep Processing for NLP (ComShaDeP 2004), Nancy, France, August 2004 (2004)
19. Schneider, G., Dowdall, J., Rinaldi, F.: A robust and deep-linguistic theory applied to large-scale parsing. In: Coling 2004 Workshop on Robust Methods in the Analysis of Natural Language Data (ROMAND 2004), Geneva, Switzerland, August 2004 (2004)
20. Nivre, J.: Inductive dependency parsing. In: Proceedings of Promote IT, Karlstad University (2004)
21. Johnson, M.: A simple pattern-matching algorithm for recovering empty nodes and their antecedents. In: Proceedings of the 40th Meeting of the ACL, University of Pennsylvania, Philadelphia (2002)

22. Kim, J., Ohta, T., Tateisi, Y., Tsujii, J.: GENIA corpus - a semantically annotated corpus for bio-textmining. Bioinformatics **19** (2003) 180–182
23. Rinaldi, F., Schneider, G., Kaljurand, K., Dowdall, J., Andronis, C., Persidis, A., Konstanti, O.: Mining relations in the genia corpus. In Scheffer, T., ed.: Accepted for publication in: Second European Workshop on Data Mining and Text Mining for Bioinformatics, ECML/PKDD (2004)
24. Rinaldi, F., Dowdall, J., Schneider, G., Persidis, A.: Answering Questions in the Genomics Domain. In: ACL 2004 Workshop on Question Answering in restricted domains, Barcelona, Spain (2004)
25. Rinaldi, F., Schneider, G., Kaljurand, K., Hess, M., Roussel, D.: Exploiting technical terminology for Knowledge Management. In Buitelaar, P., Cimiano, P., Magnini, B., eds.: Ontology Learning and Population. IOS Press (2005)
26. Kaplan, R., Riezler, S., King, T.H., III, J.T.M., Vasserman, A., Crouch, R.: Speed and accuracy in shallow and deep stochastic parsing. In: Proceedings of HLT/NAACL 2004, Boston, MA (2004)
27. Riezler, S., King, T.H., Kaplan, R.M., Crouch, R., Maxwell, J.T., Johnson, M.: 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 (2002)