

Accurate Parsing of the Proposition Bank

Gabriele Musillo

Depts of Linguistics and Computer Science
University of Geneva
2 Rue de Candolle
1211 Geneva 4 Switzerland
musillo4@etu.unige.ch

Paola Merlo

Department of Linguistics
University of Geneva
2 Rue de Candolle
1211 Geneva 4 Switzerland
merlo@lettres.unige.ch

Abstract

We integrate PropBank semantic role labels to an existing statistical parsing model producing richer output. We show conclusive results on joint learning and inference of syntactic and semantic representations.

1 Introduction

Recent successes in statistical syntactic parsing based on supervised techniques trained on a large corpus of syntactic trees (Collins, 1999; Charniak, 2000; Henderson, 2003) have brought the hope that the same approach could be applied to the more ambitious goal of recovering the propositional content and the frame semantics of a sentence. Moving towards a shallow semantic level of representation has immediate applications in question-answering and information extraction. For example, an automatic flight reservation system processing the sentence *I want to book a flight from Geneva to New York* will need to know that *from Geneva* indicates the origin of the flight and *to New York* the destination.

(Gildea and Jurafsky, 2002) define this shallow semantic task as a classification problem where the semantic role to be assigned to each constituent is inferred on the basis of probability distributions of syntactic features extracted from parse trees. They use learning features such as phrase type, position, voice, and parse tree path. Consider, for example, a sentence such as *The authority dropped at midnight Tuesday to \$ 2.80 trillion* (taken from section 00 of PropBank (Palmer et al., 2005)). The fact that *to \$ 2.80 trillion* receives a direction semantic label

is highly correlated to the fact that it is a Prepositional Phrase (PP), that it follows the verb *dropped*, a verb of change of state requiring an end point, that the verb is in the active voice, and that the PP is in a certain tree configuration with the governing verb. All the recent systems proposed for semantic role labelling (SRL) follow this same assumption (CoNLL, 2005).

The assumption that syntactic distributions will be predictive of semantic role assignments is based on linking theory. Linking theory assumes the existence of a hierarchy of semantic roles which are mapped by default on a hierarchy of syntactic positions. It also shows that regular mappings from the semantic to the syntactic level can be posited even for those verbs whose arguments can take several syntactic positions, such as psychological verbs, locatives, or datives, requiring a more complex theory. (See (Hale and Keyser, 1993; Levin and Rappaport Hovav, 1995) among many others.) If the internal semantics of a predicate determines the syntactic expressions of constituents bearing a semantic role, it is then reasonable to expect that knowledge about semantic roles in a sentence will be informative of its syntactic structure, and that learning semantic role labels at the same time as parsing will be beneficial to parsing accuracy.

We present work to test the hypothesis that a current statistical parser (Henderson, 2003) can output rich information comprising both a parse tree and semantic role labels robustly, that is without any significant degradation of the parser's accuracy on the original parsing task. We achieve promising results both on the simple parsing task, where the accuracy of the parser is measured on the standard Parseval measures, and also on the parsing task where more

complex labels comprising both syntactic labels and semantic roles are taken into account.

These results have several consequences. First, we show that it is possible to build a single integrated system successfully. This is a meaningful achievement, as a task combining semantic role labelling and parsing is more complex than simple syntactic parsing. While the shallow semantics of a constituent and its structural position are often correlated, they sometimes diverge. For example, some nominal temporal modifiers occupy an object position without being objects, like *Tuesday* in the Penn Treebank representation of the sentence above. The indirectness of the relation is also confirmed by the difficulty in exploiting semantic information for parsing. Previous attempts have not been successful. (Klein and Manning, 2003) report a reduction in parsing accuracy of an unlexicalised PCFG from 77.8% to 72.9% in using Penn Treebank function labels in training. The two existing systems that use function labels successfully, either inherit Collins’ modelling of the notion of complement (Gabbard, Kulick and Marcus, 2006) or model function labels directly (Musillo and Merlo, 2005). Furthermore, our results indicate that the proposed models are robust. To model our task accurately, additional parameters must be estimated. However, given the current limited availability of annotated treebanks, this more complex task will have to be solved with the same overall amount of data, aggravating the difficulty of estimating the model’s parameters due to sparse data.

2 The Data and the Extended Parser

In this section we describe the augmentations to our base parsing models necessary to tackle the joint learning of parse tree and semantic role labels.

PropBank encodes propositional information by adding a layer of argument structure annotation to the syntactic structures of the Penn Treebank (Marcus et al., 1993). Verbal predicates in the Penn Treebank (PTB) receive a label REL and their arguments are annotated with abstract semantic role labels A0-A5 or AA for those complements of the predicative verb that are considered arguments while those complements of the verb labelled with a semantic functional label in the original PTB receive the com-

posite semantic role label AM- X , where X stands for labels such as LOC, TMP or ADV, for locative, temporal and adverbial modifiers respectively. PropBank uses two levels of granularity in its annotation, at least conceptually. Arguments receiving labels A0-A5 or AA do not express consistent semantic roles and are specific to a verb, while arguments receiving an AM- X label are supposed to be adjuncts, and the roles they express are consistent across all verbs.

To achieve the complex task of assigning semantic role labels while parsing, we use a family of state-of-the-art history-based statistical parsers, the Simple Synchrony Network (SSN) parsers (Henderson, 2003), which use a form of left-corner parse strategy to map parse trees to sequences of derivation steps. These parsers do not impose any a priori independence assumptions, but instead smooth their parameters by means of the novel SSN neural network architecture. This architecture is capable of inducing a finite history representation of an unbounded sequence of derivation steps, which we denote $h(d_1, \dots, d_{i-1})$. The representation $h(d_1, \dots, d_{i-1})$ is computed from a set f of hand-crafted features of the derivation move d_{i-1} , and from a finite set D of recent history representations $h(d_1, \dots, d_j)$, where $j < i - 1$. Because the history representation computed for the move $i - 1$ is included in the inputs to the computation of the representation for the next move i , virtually any information about the derivation history could flow from history representation to history representation and be used to estimate the probability of a derivation move. In our experiments, the set D of earlier history representations is modified to yield a model that is sensitive to regularities in structurally defined sequences of nodes bearing semantic role labels, within and across constituents. For more information on this technique to capture structural domains, see (Musillo and Merlo, 2005) where the technique was applied to function parsing. Given the hidden history representation $h(d_1, \dots, d_{i-1})$ of a derivation, a normalized exponential output function is computed by the SSNs to estimate a probability distribution over the possible next derivation moves d_i .

To exploit the intuition that semantic role labels are predictive of syntactic structure, we must pro-

vide semantic role information as early as possible to the parser. Extending a technique presented in (Klein and Manning, 2003) and adopted in (Merlo and Musillo, 2005) for function labels with state-of-the-art results, we split some part-of-speech tags into tags marked with AM- X semantic role labels. As a result, 240 new POS tags were introduced to partition the original tag set which consisted of 45 tags. Our augmented model has a total of 613 non-terminals to represent both the PTB and PropBank labels, instead of the 33 of the original SSN parser. The 580 newly introduced labels consist of a standard PTB label followed by one or more PropBank semantic roles, such as PP-AM-TMP or NP-A0-A1. These augmented tags and the new non-terminals are included in the set f , and will influence bottom-up projection of structure directly.

These newly introduced fine-grained labels fragment our PropBank data. To alleviate this problem, we enlarge the set f with two additional binary features. One feature decides whether a given preterminal or nonterminal label is a semantic role label belonging to the set comprising the labels A0-A5 and AA. The other feature indicates if a given label is a semantic role label of type AM- X , or otherwise. These features allow the SSN to generalise in several ways. All the constituents bearing an A0-A5 and AA labels will have a common feature. The same will be true for all nodes bearing an AM- X label. Thus, the SSN can generalise across these two types of labels. Finally, all constituents that do not bear any label will now constitute a class, the class of the nodes for which these two features are false.

3 Experiments and Discussion

Our extended semantic role SSN parser was trained on sections 2-21 and validated on section 24 from the PropBank. Testing data are section 23 from the CoNLL-2005 shared task (Carreras and Marquez, 2005).

We perform two different evaluations on our model trained on PropBank data. We distinguish between two parsing tasks: the PropBank parsing task and the PTB parsing task. To evaluate the former parsing task, we compute the standard Parseval measures of labelled recall and precision of constituents, taking into account not only the 33 original labels,

but also the newly introduced PropBank labels. This evaluation gives us an indication of how accurately and exhaustively we can recover this richer set of non-terminal labels. The results, computed on the testing data set from the PropBank, are shown in the PropBank column of Table 1, first line. To evaluate the PTB task, we ignore the set of PropBank semantic role labels that our model assigns to constituents (PTB column of Table 1, first line to be compared to the third line of the same column).

To our knowledge, no results have yet been published on parsing the PropBank.¹ Accordingly, it is not possible to draw a straightforward quantitative comparison between our PropBank SSN parser and other PropBank parsers. However, state-of-the-art semantic role labelling systems (CoNLL, 2005) use parse trees output by state-of-the-art parsers (Collins, 1999; Charniak, 2000), both for training and testing, and return partial trees annotated with semantic role labels. An indirect way of comparing our parser with semantic role labellers suggests itself.² We merge the partial trees output by a semantic role labeller with the output of the parser on which it was trained, and compute PropBank parsing performance measures on the resulting parse trees. The third line, PropBank column of Table 1 reports such measures summarised for the five best semantic role labelling systems (Punyakank et al., 2005b; Haghghi et al., 2005; Pradhan et al., 2005; Marquez et al., 2005; Surdeanu and Turmo, 2005) in the CoNLL 2005 shared task. These systems all use (Charniak, 2000)’s parse trees both for training and testing, as well as various other information sources including sets of n -best parse trees, chunks, or named entities. Thus, the partial trees output by these systems were merged with the parse trees returned by Charniak’s parser (second line, PropBank column).³

These results jointly confirm our initial hypothe-

¹(Shen and Joshi, 2005) use PropBank labels to extract LTAG spinal trees to train an incremental LTAG parser, but they do not parse PropBank. Their results on the PTB are not directly comparable to ours as calculated on dependency relations and obtained using gold POS.

²Current work aims at extending our parser to recovering the argument structure for each verb, supporting a direct comparison to semantic role labellers.

³Because of differences in tokenisations, we retain only 2280 sentences out of the original 2416.

	PTB	PropBank
SSN+Roles model	89.0	82.8
CoNLL five best	-	83.3–84.1
Henderson 03 SSN	89.1	-

Table 1: Percentage F-measure of our SSN parser on PTB and PropBank parsing, compared to the original SSN parser and to the best CoNLL 2005 SR labellers.

sis. The performance on the parsing task (PTB column) does not appreciably deteriorate compared to a current state-of-the-art parser, even if our learner can output a much richer set of labels, and therefore solves a considerably more complex problem, suggesting that the relationship between syntactic PTB parsing and semantic PropBank parsing is strict enough that an integrated approach to the problem of semantic role labelling is beneficial. Moreover, the results indicate that we can perform the more complex PropBank parsing task at levels of accuracy comparable to those achieved by the best semantic role labellers (PropBank column). This indicates that the model is robust, as it has been extended to a richer set of labels successfully, without increase in training data. In fact, the limited availability of data is increased further by the high variability of the argumental labels A0-A5 whose semantics is specific to a given verb or a given verb sense.

Methodologically, these initial results on a joint solution to parsing and semantic role labelling provide the first direct test of whether parsing is necessary for semantic role labelling (Gildea and Palmer, 2002; Punyakanok et al., 2005a). Comparing semantic role labelling based on chunked input to the better semantic role labels retrieved based on parsed trees, (Gildea and Palmer, 2002) conclude that parsing is necessary. In an extensive experimental investigation of the different learning stages usually involved in semantic role labelling, (Punyakanok et al., 2005a) find instead that sophisticated chunking can achieve state-of-the-art results. Neither of these pieces of work actually used a parser to do SRL. Their investigation was therefore limited to establishing the usefulness of syntactic features for the SRL task. Our results do not yet indicate that parsing is beneficial to SRL, but they show that the joint task can be performed successfully.

Acknowledgements We thank the Swiss NSF for supporting this research under grant number 101411-105286/1, James Henderson and Ivan Titov for sharing their SSN software, and Xavier Carreras for providing the CoNLL-2005 data.

References

- X. Carreras and L. Marquez. 2005. Introduction to the CoNLL-2005 shared task: Semantic role labeling. *Procs of CoNLL-2005*.
- E. Charniak. 2000. A maximum-entropy-inspired parser. *Procs of NAACL'00*, pages 132–139, Seattle, WA.
- M. Collins. 1999. *Head-Driven Statistical Models for Natural Language Parsing*. Ph.D. thesis, Pennsylvania.
- CoNLL. 2005. *Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, Ann Arbor, MI.
- R. Gabbard, S. Kulick and M. Marcus. 2006. Fully parsing the Penn Treebank. *Procs of NAACL'06*, New York, NY.
- D. Gildea and D. Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.
- D. Gildea and M. Palmer. 2002. The necessity of parsing for predicate argument recognition. *Procs of ACL 2002*, 239–246, Philadelphia, PA.
- A. Haghighi, K. Toutanova, and C. Manning. 2005. A joint model for semantic role labeling. *Procs of CoNLL-2005*, Ann Arbor, MI.
- K. Hale and J. Keyser. 1993. On argument structure and the lexical representation of syntactic relations. In K. Hale and J. Keyser, editors, *The View from Building 20*, 53–110. MIT Press.
- J. Henderson. 2003. Inducing history representations for broad-coverage statistical parsing. *Procs of NAACL-HLT'03*, 103–110, Edmonton, Canada.
- D. Klein and C. Manning. 2003. Accurate unlexicalized parsing. *Procs of ACL'03*, 423–430, Sapporo, Japan.
- B. Levin and M. Rappaport Hovav. 1995. *Unaccusativity*. MIT Press, Cambridge, MA.
- M. Marcus, B. Santorini, and M.A. Marcinkiewicz. 1993. Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics*, 19:313–330.
- L. Marquez, P. Comas, J. Gimenez, and N. Catala. 2005. Semantic role labeling as sequential tagging. *Procs of CoNLL-2005*.
- P. Merlo and G. Musillo. 2005. Accurate function parsing. *Procs of HLT/EMNLP 2005*, 620–627, Vancouver, Canada.
- G. Musillo and P. Merlo. 2005. Lexical and structural biases for function parsing. *Procs of IWPT'05*, 83–92, Vancouver, Canada.
- M. Palmer, D. Gildea, and P. Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31:71–105.
- S. Pradhan, K. Hacioglu, W. Ward, J. Martin, and D. Jurafsky. 2005. Semantic role chunking combining complementary syntactic views. *Procs of CoNLL-2005*.
- V. Punyakanok, D. Roth, and W. Yih. 2005a. The necessity of syntactic parsing for semantic role labeling. *Procs of IJ-CAI'05*, Edinburgh, UK.
- V. Punyakanok, P. Koomen, D. Roth, and W. Yih. 2005b. Generalized inference with multiple semantic role labeling systems. *Procs of CoNLL-2005*.
- L. Shen and A. Joshi. 2005. Incremental LTAG parsing. *Procs of HLT/EMNLP 2005*, Vancouver, Canada.
- M. Surdeanu and J. Turmo. 2005. Semantic role labeling using complete syntactic analysis. *Procs of CoNLL-2005*.