A Second-Order Joint Eisner Model for Syntactic and Semantic Dependency Parsing

Xavier Lluís **Stefan Bott** **Lluís Màrquez**
TALP Research Center – Software Department (LSI)
Technical University of Catalonia (UPC)
{xlluis,sbott,lluism}@lsi.upc.edu

Abstract

We present a system developed for the CoNLL-2009 Shared Task (Hajič et al., 2009). We extend the Carreras (2007) parser to jointly annotate syntactic and semantic dependencies. This state-of-the-art parser factorizes the built tree in second-order factors. We include semantic dependencies in the factors and extend their score function to combine syntactic and semantic scores. The parser is coupled with an on-line averaged perceptron (Collins, 2002) as the learning method. Our averaged results for all seven languages are 71.49 macro F_1, 79.11 LAS and 63.06 semantic F_1.

1 Introduction

Systems that jointly annotate syntactic and semantic dependencies were introduced in the past CoNLL-2008 Shared Task (Surdeanu et al., 2008). These systems showed promising results and proved the feasibility of a joint syntactic and semantic parsing (Henderson et al., 2008; Lluís and Màrquez, 2008).

The Eisner (1996) algorithm and its variants are commonly used in data-driven dependency parsing. Improvements of this algorithm presented by McDonald et al. (2006) and Carreras (2007) achieved state-of-the-art performance for English in the CoNLL-2007 Shared Task (Nivre et al., 2007). Johansson and Nugues (2008) presented a system based on the Carreras' extension of the Eisner algorithm that ranked first in the past CoNLL-2008 Shared Task. We decided to extend the Carreras (2007) parser to jointly annotate syntactic and semantic dependencies.

The present year Shared Task has the incentive of being multilingual with each language presenting their own particularities. An interesting particularity is the direct correspondence between syntactic and semantic dependencies provided in Catalan, Spanish and Chinese. We believe that these correspondences can be captured by a joint system. We specially look at the syntactic-semantic alignment of the Catalan and Spanish datasets.

Our system is an extension of the Lluís and Màrquez (2008) CoNLL-2008 Shared Task system. We introduce these two following novelties:

- An extension of the second-order Carreras (2007) algorithm to annotate semantic dependencies.
- A combined syntactic-semantic scoring for Catalan and Spanish to exploit the syntactic-semantic mappings.

The following section outlines the system architecture. The next sections present in more detail the system novelties.

2 Architecture

The architecture consists on four main components: 1) Preprocessing and feature extraction. 2) Syntactic preparsing. 3) Joint syntactic-semantic parsing. 4) Predicate classification.

The preprocessing and feature extraction is intended to ease and improve the performance of the parser precomputing a binary representation of

each sentence features. These features are borrowed from existing and widely-known systems (Xue and Palmer, 2004; McDonald et al., 2005; Carreras et al., 2006; Surdeanu et al., 2007).

The following step is a syntactic pre-parse. It is only required to pre-compute additional features (e.g., syntactic path, syntactic frame) from the syntax. These new features will be used for the semantic role component of the following joint parser.

The joint parser is the core of the system. This single algorithm computes the complete parse that optimizes a score according to a function that depends on both syntax and semantics. Some of the required features that could be unavailable or expensive to compute at that time are provided by the previous syntactic pre-parse.

The predicate sense classification is performed as the last step. Therefore no features representing the predicate sense are employed during the training. The predicates are labeled with the most frequent sense extracted from the training corpus.

No further postprocessing is applied.

3 Second-order Eisner model

The Carreras' extension of the Eisner inference algorithm is an expensive $O(n^4)$ parser. The number of assignable labels for each dependency is a hidden multiplying constant in this asymptotic cost.

We begin describing a first-order dependency parser. It receives a sentence x and outputs a dependency tree y. A dependency, or first-order factor, is defined as $f_1 = \langle h, m, l \rangle$. Where h is the head token, m the modifier and l the syntactic label. The score for this factor f_1 is computed as:

$$\text{score}_1(f_1, x, \mathbf{w}) = \phi(h, m, x) \cdot \mathbf{w}^{(l)}$$

Where $\mathbf{w}^{(l)}$ is the weight vector for the syntactic label l and ϕ a feature extraction function.

The parser outputs the best tree y^* from the set $\mathcal{T}(x)$ of all projective dependency trees.

$$y^*(x) = \underset{y \in \mathcal{T}(x)}{\text{argmax}} \sum_{f_1 \in y} score(f_1, x, \mathbf{w})$$

The second-order extension decomposes the dependency tree in factors that include some children of the head and modifier. A second-order factor is:

$$f_2 = \langle h, m, l, c_h, c_{mo}, c_{mi} \rangle$$

where c_h is the daughter of h closest to m within the tokens $[h, \ldots, m]$; c_{mo} is the outermost daughter of m outside $[h, \ldots, m]$; and $c_m i$ is the furthest daughter of m inside $[h, \ldots, m]$.

The score for these new factors is computed by

$$\begin{aligned}\text{score}_2(f_2, x, \mathbf{w}) = \phi(h, m, x) \cdot \mathbf{w}^{(l)} + \\ \phi(h, m, c_h, x) \cdot \mathbf{w}^{(l)}_{c_h} + \\ \phi(h, m, c_{mi}, x) \cdot \mathbf{w}^{(l)}_{c_{mi}} + \\ \phi(h, m, c_{mo}, x) \cdot \mathbf{w}^{(l)}_{c_{mo}}\end{aligned}$$

The parser builds the best-scoring projective tree factorized in second-order factors. The score of the tree is also defined as the sum of the score of its factors.

3.1 Joint second-order model

We proceeded in an analogous way in which the Lluís and Màrquez (2008) extended the first-order parser. That previous work extended a first-order model by including semantic labels in first-order dependencies.

Now we define a second-order joint factor as:

$$\begin{aligned}f_{2\text{syn-sem}} = \\ \left\langle h, m, l, c_h, c_{mo}, c_{mi}, l_{semp_1}, \ldots, l_{semp_q} \right\rangle\end{aligned}$$

Note that we only added a set of semantic labels $l_{semp_1}, \ldots, l_{semp_q}$ to the second-order factor. Each one of these semantic labels represent, if any, one semantic relation between the argument m and the predicate p_i. There are q predicates in the sentence, labeled $p_1, \ldots, p_q$.

The corresponding joint score to a given joint factor is computed by adding a semantic score to the previously defined *score*$_2$ second-order score function:

$$\begin{aligned}\text{score}_{2\text{syn-sem}}(f_{2\text{syn-sem}}, x, \mathbf{w}) = \\ \text{score}_2(f_2, x, \mathbf{w}) + \\ \sum_{p_i} \frac{\text{score}_{\text{sem}}(h, m, p_i, l_{semp_i}, x, \mathbf{w})}{q}\end{aligned}$$

where,

$$\begin{aligned}\text{score}_{\text{sem}}(h, m, p_i, l_{sem}, x, \mathbf{w}) = \\ \phi_{sem}(h, m, p_i, x) \cdot \mathbf{w}^{(l_{sem})}\end{aligned}$$

We normalize the semantic score by the number of predicates q. The semantic score is computed as a score between m and each sentence predicate p_i. No second-order relations are considered in these score functions. The search of the best c_h, c_{mo} and c_{mi} is independent of the semantic components of the factor. The computational cost of the algorithm is increased by one semantic score function call for every m, h, and p_i combination. The asymptotic cost of this operation is $O(q \cdot n^2)$ and it is sequentially performed among other $O(n^2)$ operations in the main loop of the algorithm.

Algorithm 1 Extension of the Carreras (2007) algorithm

$C[s][t][d][m] \leftarrow 0, \forall s, t, d, m$
$O[s][t][d][l] \leftarrow 0, \forall s, t, d, l$
for $k = 1, \ldots, n$ **do**
  for $s = 0, \ldots, n - k$ **do**
    $t \leftarrow s + k$
    $\forall l \quad O[s][t][\leftarrow][l] = \max_{r,cmi,ch}$
      $C[s][r][\rightarrow][cmi] + C[r+1][t][\leftarrow][ch]$
      $+\text{score}(t, s, l) + \text{score}_{cmi}(t, s, c_{mi}, l) +$
      $\text{score}_{ch}(t, s, l, c_h) +$
      $\sum_{p_i} \max_{l_{sem}} \text{scoresem}(t, s, p_i, l_{sem})/q$

    $\forall l \quad O[s][t][\rightarrow][l] = \max_{r,cmi,ch}$
      $C[s][r][\rightarrow][ch] + C[r+1][t][\leftarrow][cmi] +$
      $\text{score}(s, t, l) + \text{score}_{cmi}(s, t, c_{mi}, l) +$
      $\text{score}_{ch}(s, t, l, c_h) +$
      $\sum_{p_i} \max_{l_{sem}} \text{scoresem}(t, s, p_i, l_{sem})/q$

    $\forall m \quad C[s][t][\leftarrow][m] = \max_{l,cmo}$
      $C[s][m][\leftarrow][cmo] + O[m][t][\leftarrow][l] +$
      $\text{score}_{cmo}(s, m, l, c_{mo})$

    $\forall m \quad C[s][t][\rightarrow][m] = \max_{l,cmo}$
      $O[s][m][\rightarrow][l] + C[m][t][\rightarrow][cmo] +$
      $\text{score}_{cmo}(m, t, l, c_{mo})$
  end for
end for

Our implementation slightly differs from the original Carreras algorithm description. The main difference is that no specific features are extracted for the second-order factors. This allows us to reuse the feature extraction mechanism of a first-order parser.

Algorithm 1 shows the Carreras' extension of the Eisner algorithm including our proposed joint semantic scoring.

The tokens s and t represent the start and end tokens of the current substring, also called span. The direction $d \in \{\leftarrow, \rightarrow\}$ defines whether t or s is the head of the last dependency built inside the span. The score functions *score*$_{ch}$,*score*$_{cmi}$ and *score*$_{cmo}$ are the linear functions that build up the previously defined second-order global score, e.g., *score*$_{ch}$= $\phi(h, m, c_h, x) \cdot \mathbf{w}_{c_h}^{(l)}$. The two tables C and O maintain the dynamic programming structures.

Note that the first steps of the inner loop are applied for all l, the syntactic label, but the semantic score function does not depend on l. Therefore the best semantic label can be chosen independently.

For simplicity, we omitted the weight vectors required in each score function and the backpointers tables to save the local decisions. We also omitted the definition of the domain of some variables. Moreover, the filter of the set of assignable labels is not shown. A basic filter regards the POS of the head and modifier to filter out the set of possible arguments for each predicate. Another filter extract the set of allowed arguments for each predicate from the frames files. These last filters were applied to the English, German and Chinese.

3.2 Catalan and Spanish joint model

The Catalan and Spanish datasets (Taulé et al., 2008) present two interesting properties. The first property, as previously said, is a direct correspondence between syntactic and semantic labels. The second interesting property is that all semantic dependencies exactly overlap with the syntactic tree. Thus the semantic dependency between a predicate and an argument always has a matching syntactic dependency between a head and a modifier. The Chinese data also contains direct syntactic-semantic mappings. But due to the Shared Task time constraints we did not implemented a specific parsing method for this language.

The complete overlap between syntax and semantics can simplify the definition of a second-order joint factor. In this case, a second-order factor will only have, if any, one semantic dependency. We only allow at most one semantic relation l_{sem} between the head token h and the modifier m. Note that h must be a sentence predicate and m its argument if

l_{sem} is not null. We extend the second-order factors with a single and possibly null semantic label, i.e., $f_{2\text{syn-sem}} = \langle h, m, l, c_h, c_{mo}, c_{mi}, l_{sem} \rangle$. This slightly simplifies the scoring function:

$$\begin{aligned} \text{score}_{2\text{syn-sem}}(f_{2\text{syn-sem}}, x, \mathbf{w}) = \\ \text{score}_2(f_2, x, \mathbf{w}) + \\ \alpha \cdot \text{score}_{\text{sem}}(h, m, x, \mathbf{w}) \end{aligned}$$

where α is an adjustable parameter of the model and,

$$\text{score}_{\text{sem}}(h, m, x, \mathbf{w}) = \phi_{sem}(h, m, x) \cdot \mathbf{w}^{(l_{sem})}$$

The next property that we are intended to exploit is the syntactic-semantic mappings. These mappings define the allowed combinations of syntactic and semantic labels. The label combinations can only be exploited when there is semantic dependency between the head h and the modifier m of a factor. An argument identification classifier determines the presence of a semantic relation, given h is a predicate. In these cases we only generate factors that are compliant with the mappings. If a syntactic label has many corresponding semantic labels we will score all of them and select the combination with the highest score.

The computational cost is not significantly increased as there is a bounded number of syntactic and semantic combinations to score. In addition, the only one-argument-per-factor constraint reduces the complexity of the algorithm with respect to the previous joint extension.

We found some inconsistencies in the frames files provided by the organizers containing the correspondences between syntax and semantics. For this reason we extracted them directly from the corpus. The extracted mappings discard the 7.9% of the correct combinations in the Catalan development corpus that represent a 1.7% of its correct syntactic dependencies. The discarded semantic labels are the 5.14% for Spanish representing the 1.3% of the syntactic dependencies.

4 Results and discussion

Table 1 shows the official results for all seven languages, including out-of-domain data labeled as *ood*. The high computational cost of the second-order models prevented us from carefully tuning the system parameters. After the shared task evaluation deadline, some bug were corrected, improving the system performance. The last results are shown in parenthesis.

The combined filters for Catalan and Spanish hurt the parsing due to the discarded correct labels but we believe that this effect is compensated by an improved precision in the cases where the correct labels are not discarded. For example, in Spanish these filters improved the syntactic LAS from 85.34 to 86.77 on the development corpus using the gold syntactic tree as the pre-parse tree.

Figure 1 shows the learning curve for the English and Czech language. The results are computed in the development corpus. The semantic score is computed using gold syntax and gold predicate sense classification. We restricted the learning curve to the first epoch. Although the this first epoch is very close to the best score, some languages showed improvements until the fourth epoch. In the figure we can see better syntactic results for the joint system with respect to the syntactic-only parser. We should not consider this improvement completely realistic as the semantic component of the joint system uses gold features (i.e., a gold pre-parse). Nonetheless, it points that a highly accurate semantic component could improve the syntax.

Table 2 shows the training time for a second-order syntactic and joint configurations of the parser. Note that the time per instance is an average and some sentences could require a significantly higher time. Recall that our parser is $O(n^4)$ dependant on the sentence length. We discarded large sentences during training for efficiency reasons. We discarded sentences with more than 70 words for all languages except for Catalan and Spanish where the threshold was set to 100 words in the syntactic parser. This larger number of sentences is aimed to improve the syntactic performance of these languages. The shorter sentences used in the joint parsing and the pruning of the previously described filters reduced the training time for Catalan and Spanish. The amount of main memory consumed by the system is 0.5–1GB. The machine used to perform the computations is an *AMD64 Athlon 5000+*.

	avg	cat	chi	cze	eng	ger	jap	spa
macro F_1	71.49 (74.90)	56.64 (73.21)	66.18 (70.91)	75.95	81.69	72.31	81.76	65.91 (68.46)
syn LAS	79.11 (82.22)	64.21(84.20)	70.53 (70.90)	75.00	87.48	81.94	91.55	83.09 (84.48)
semantic F_1	63.06 (67.41)	46.79 (61.68)	59.72 (70.88)	76.90	75.86	62.66	71.60	47.88 (52.30)
ood macro F_1	71.92	-	-	74.56	73.91	67.30	-	-
ood syn LAS	75.09	-	-	72.11	80.92	72.25	-	-
ood sem F_1	68.74	-	-	77.01	66.88	62.34	-	-

Table 1: Overall results. In parenthesis post-evaluation results.

	cat	chi	cze	eng	ger	jap	spa
syntax only (s/sentence)	18.39	8.07	3.18	2.56	1.30	1.07	15.31
joint system (s/sentence)	10.91	9.49	3.99	3.13	2.36	1.25	12.29

Table 2: Parsing time per sentence.

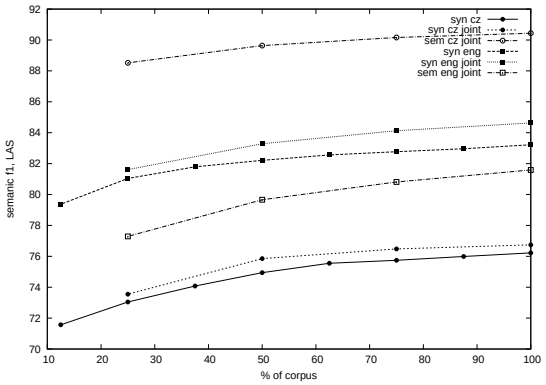

Figure 1: Learning curves for the syntactic-only and joint parsers in Czech and English.

5 Conclusion

We have shown that a joint syntactic-semantic parsing can be based on the state-of-the-art Carreras (2007) parser at an expense of a reasonable cost. Our second-order parser still does not reproduce the state-of-the art results presented by similar systems (Nivre et al., 2007). Although we achieved mild results we believe that a competitive system based in our model can be built. Further tuning is required and a complete set of new second-order features should be implemented to improve our parser.

The multilingual condition of the task allows us to evaluate our approach in seven different languages. A detailed language-dependent evaluation can give us some insights about the strengths and weaknesses of our approach across different languages. Unfortunately we believe that this objective was possibly not accomplished due to the time constraints.

The Catalan and Spanish datasets presented interesting properties that could be exploited. The mapping between syntax and semantics should be specially useful for a joint system. In addition the semantic dependencies for these languages are aligned with the projective syntactic dependencies, i.e., the predicate-argument pairs exactly match syntactic dependencies. This is a useful property to simultaneously build joint dependencies.

6 Future and ongoing work

Our syntactic and semantic parsers, as many others, is not exempt of bugs. Furthermore, very few tuning and experimentation was done during the development of our parser due to the Shared Task time constraints. We believe that we still did not have enough data to fully evaluate our approach. Further experimentation is required to asses the improvement of a joint architecture vs. a pipeline architecture. Also a careful analysis of the system across the different languages is to be performed.

Acknowledgments

We thank the corpus providers (Taulé et al., 2008; Palmer and Xue, 2009; Hajič et al., 2006; Surdeanu et al., 2008; Burchardt et al., 2006; Kawahara et al., 2002) for their effort in the annotation and conversion of the seven languages datasets.

References

Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó, and Manfred Pinkal. 2006. The SALSA corpus: a German corpus resource for lexical semantics. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC-2006)*, Genoa, Italy.

Xavier Carreras, Mihai Surdeanu, and Lluís Màrquez. 2006. Projective dependency parsing with perceptron. In *Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL-2006)*.

Xavier Carreras. 2007. Experiments with a higher-order projective dependency parser. In *Proceedings of the 11th Conference on Computational Natural Language Learning (CoNLL-2007)*.

Michael Collins. 2002. Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*.

Jason M. Eisner. 1996. Three new probabilistic models for dependency parsing: An exploration. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING-96)*.

Jan Hajič, Jarmila Panevová, Eva Hajičová, Petr Sgall, Petr Pajas, Jan Štěpánek, Jiří Havelka, Marie Mikulová, and Zdeněk Žabokrtský. 2006. Prague Dependency Treebank 2.0.

Jan Hajič, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. 2009. The CoNLL-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL-2009), June 4-5*, Boulder, Colorado, USA.

James Henderson, Paola Merlo, Gabriele Musillo, and Ivan Titov. 2008. A latent variable model of synchronous parsing for syntactic and semantic dependencies. In *Proceedings of the 12th Conference on Computational Natural Language Learning (CoNLL-2008)*, Manchester, UK.

Richard Johansson and Pierre Nugues. 2008. Dependency-based syntactic–semantic analysis with propbank and nombank. In *Proceedings of the 12th Conference on Computational Natural Language Learning (CoNLL-2008)*, Manchester, UK.

Daisuke Kawahara, Sadao Kurohashi, and Kôiti Hasida. 2002. Construction of a Japanese relevance-tagged corpus. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, pages 2008–2013, Las Palmas, Canary Islands.

Xavier Lluís and Lluís Màrquez. 2008. A joint model for parsing syntactic and semantic dependencies. In *Proceedings of the 12th Conference on Computational Natural Language Learning (CoNLL-2008)*, Manchester, UK.

Ryan McDonald and Fernando Pereira. 2006. Online learning of approximate dependency parsing algorithms. In *11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2006)*.

Ryan McDonald, Koby Crammer, and Fernando Pereira. 2005. Online large-margin training of dependency parsers. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL-2005)*.

J. Nivre, J. Hall, S. Kübler, R. McDonald, J. Nilsson, S. Riedel, and D. Yuret. 2007. The CoNLL 2007 shared task on dependency parsing.

Martha Palmer and Nianwen Xue. 2009. Adding semantic roles to the Chinese Treebank. *Natural Language Engineering*, 15(1):143–172.

Mihai Surdeanu, Lluís Màrquez, Xavier Carreras, and Pere R. Comas. 2007. Combination strategies for semantic role labeling. *Journal of Artificial Intelligence Research*.

Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez, and Joakim Nivre. 2008. The CoNLL-2008 shared task on joint parsing of syntactic and semantic dependencies. In *Proceedings of the 12th Conference on Computational Natural Language Learning (CoNLL-2008)*.

Mariona Taulé, Maria Antònia Martí, and Marta Recasens. 2008. AnCora: Multilevel Annotated Corpora for Catalan and Spanish. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC-2008)*, Marrakesh, Morroco.

Nianwen Xue and Martha Palmer. 2004. Calibrating features for semantic role labeling. In *Proceedings of the Empirical Methods in Natural Language Processing (EMNLP-2004)*.