

LX-Parser and LX-DepParser: Online Services for Constituency and Dependency Parsing

João Silva, Ruben Reis, Patricia Gonçalves, António Branco

University of Lisbon

{jsilva, ruben.reis, patricia.nunes, antonio.branco}@di.fc.ul.pt

Abstract This paper presents two freely-available online services for parsing natural language: a constituency parser and a dependency parser.

Key words: constituency parsing, dependency parsing

1 Constituency Parser

LX-Parser is a freely available on-line service for constituency parsing of Portuguese sentences.

The service is supported by the Stanford factored parser [1].

The parser was trained and evaluated over CINTIL-Treebank, a treebank produced from the output of a deep processing grammar by manually selecting the correct parse for a sentence from among all the possible parses produced by the grammar.

Under the Parseval metric, LX-Parser achieves an accuracy of 88% (value obtained through 10-fold cross-evaluation).

Figure 1 (a) shows a screenshot of the LX-Parser website after a sentence has been parsed.

The service can be found at <http://lxparser.di.fc.ul.pt>.

2 Dependency Parser

LX-DepParser is a freely available online service that allows visualization of grammatical relations between the words in a sentence entered by the user.

The sentence is processed by LX-Suite [2], a set of shallow processing tools that include a state-of-the-art POS tagger.

The resulting annotated sentence is then processed by MSTParser (Minimum-Spanning Tree Parser) [3], which outputs the grammatical dependencies between the words of the sentence.

Like LX-Parser, LX-DepParser was trained over the CINTIL treebank after the dependencies were extracted from the X-bar syntactic structure.

Using a 10-fold cross-evaluation methodology, LX-DepParser achieves scores of 91.4% for unlabeled accuracy and 86.8% for labeled accuracy.

Figure 1 (b) shows a screenshot of the LX-DepParser website after a sentence has been parsed.

The service can be found at <http://lxdepparser.di.fc.ul.pt>.

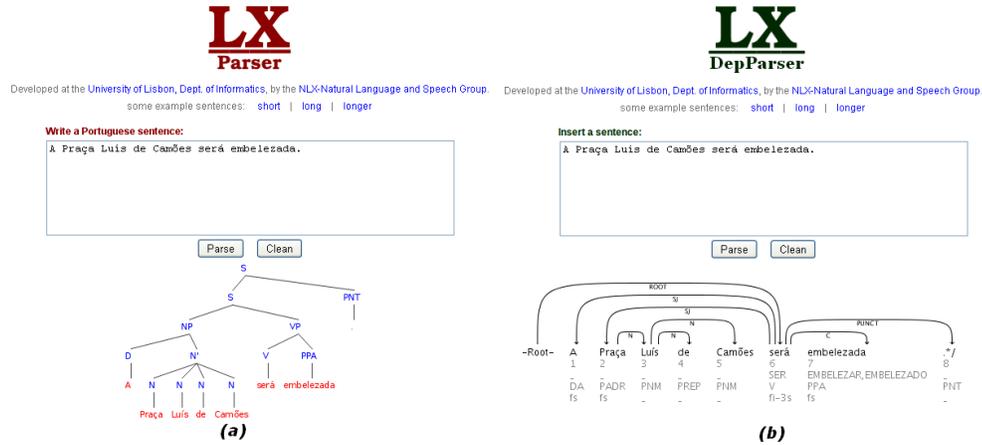


Figure 1. (a) LX-Parser— (b) LX-DepParser

3 Conclusions

In this paper we present two freely-available online services for parsing Portuguese sentences, LX-Parser for constituency parsing and LX-DepParser for dependency parsing, that present state of the art performance.

These online services were developed and being maintained and extended by the NLX-Natural Language and Speech Group (<http://nlx.di.fc.ul.pt>), of the University of Lisbon, Department of Informatics.

Acknowledgements. This work was partly supported by FCT-Fundation of Science and Technology under the grant FCT/PTDC/PLP/81157/2006 for the project SemanticShare.

References

1. Klein, D., Manning, C.: Fast exact inference with a factored model for NLP. *Advances in Neural Language Processing Systems* **15** (2003) 3–10
2. Branco, A., Silva, J.: Evaluating solutions for the rapid development of state-of-the-art POS taggers for Portuguese. In: *Proceedings of the 4th Language Resources and Evaluation Conference (LREC)*. (2004) 507–510
3. McDonald, R., Lerman, K., Pereira, F.: Multilingual dependency analysis with a two-stage discriminative parser. In: *Proceedings of the 10th Conference on Natural Language Learning*. (2006) 216–220