



Task-Oriented Evaluation of Dependency Parsing with Open Information Extraction

Pablo Gamallo¹ and Marcos Garcia²

¹ Centro Singular de Investigación en Tecnoloxías da Información (CíTIUS),
Universidade de Santiago de Compostela, Galiza, Spain

`pablo.gamallo@usc.es`

² LyS Group, Universidade da Coruña, Galiza, Spain

`marcos.garcia.gonzalez@udc.gal`

Abstract. This paper presents a comparative evaluation of several Portuguese parsers. Our objective is to use dependency parsers in a specific information extraction task, namely Open Information Extraction (OIE), and measure the impact of each parser in this task. The experiments show that the scores obtained by the evaluated parsers are quite similar even though they allow to extract different (and then complementary) items of information.

Keywords: Dependency parsing · Open Information Extraction
Task-oriented evaluation

1 Introduction

The most popular method for dependency parser comparison involves the direct measurement of the parser output accuracy in terms of metrics such as labeled attachment score (LAS) and unlabeled attachment score (UAS). This assumes the existence of a gold-standard test corpus developed with the use of a specific tagset and a list of dependency names by following some specific syntactic criteria. Such an evaluation procedure makes it difficult to evaluate parsing systems developed with different syntactic criteria from those used in the gold-standard test. Direct evaluation has been thought to compare strategies based on different algorithms but trained on the same treebanks and using the same tokenization. In fact, the strict requirements derived from direct evaluation prevents us from making fair comparisons among systems based on very different frameworks.

In this paper, we present a task-oriented evaluation of different dependency syntactic analyzers for Portuguese using the specific task of Open Information Extraction (OIE). This evaluation allows us to compare under the same conditions very different systems, more precisely, parsers trained on treebanks with different linguistic criteria, or even data-driven and rule-based parsers. Other task-oriented evaluation work focused on measure parsing accuracy through its

influence in the performance of different types of NLP systems, such as sentiment analysis [11].

OIE is an information extraction task that consists of extracting basic propositions from sentences [2]. There are many OIE systems for English language, including those based on shallow syntactic information, e.g. TextRunner [2] and ReVerb [6], and those using syntactic dependencies: e.g. OLLIE [14] or ClauseIE [4]. There are also some proposals for Portuguese language: DepOE [10], Report [17], ArgOE [8], DependentIE [13], and the extractor of open relations between named entities reported in [3]. In order to use OIE systems to evaluate dependency parsers for Portuguese, we need an OIE system for Portuguese taking as input dependency trees. For the purpose of our indirect evaluation, we will use the open source system described in [8], which takes as input dependency trees in CoNLL-X format.

2 The Role of Dependency Parsing in OIE

We consider that it is possible to indirectly evaluate a parser by measuring the performance of the OIE system in which the parser is integrated as many errors made by the OIE system come from the parsing step. Let us take for example one of the sentences of our evaluation dataset (and described in the next section):

A regulação desses processos depende de várias interações de indivíduos com os seus ambientes

The regulation of these processes depends on several interactions of individuals with their environments

One of the evaluated systems extracts the following two basic propositions (to simplify we show just the English translation):

(“the regulation of these processes”, “depends on”, “several interactions of individuals”),
 *(“the regulation of these processes”, “depends with”, “their environments”)

The second proposition is not correct since it has been extracted from an odd dependency, such as shown in Fig. 1. The dependency between “environments” and “depends” (red arc below the sentence) is incorrect since “environments” is actually dependent on the noun “interactions”.¹ In sum, any odd dependency given by the parser makes the OIE system incorrectly extract, at least, one odd triple.

Furthermore, the resulting triples extracted by an OIE system are also an excellent way of visualizing the type of errors made by the dependency parser and, thereby, dependency-based OIE systems can be seen as useful linguistic tools to carry out error analysis on the parsing step.

¹ In this analysis, we use labels and syntactic criteria based on Universal Dependencies, e.g. prepositions are case-marking elements that are dependents of the noun or clause they attach to or introduce.

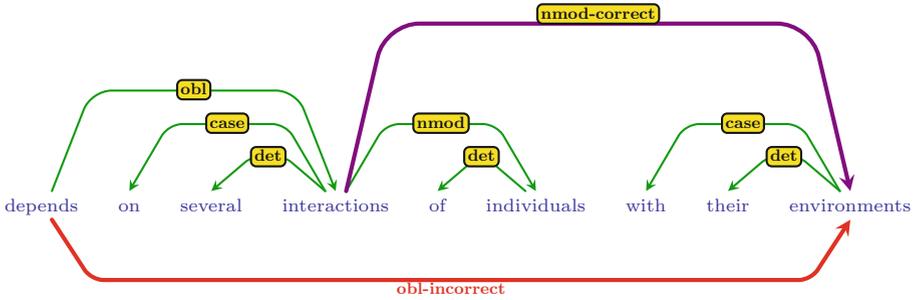


Fig. 1. Dependency analysis with Universal Dependencies. The head of “environments” is the noun “interactions” via *nmod* dependency, and not the verb “depends”. (Color figure online)

3 Experiments

Our objective is to evaluate and compare different Portuguese dependency parsers which can be easily integrated into an open-source OIE system. For this purpose, we use the OIE module of *LinguaKit*, described in [8], which takes as input any text parsed in CoNLL-X format. We were able to integrate five Portuguese parsers into the OIE module: two rule-based parsers and three data-driven parsers. The rule-based systems are two different versions of *DepPattern* [7,9]:

The parser used by *ArgOE* [8], and that available in *LinguaKit*.² The three data-driven parsers were trained using *MaltParser* 1.7.1³ and two different algorithms: *Nivre eager* [15], based on arc-eager algorithm, and *2-planar* [12]. They were trained with two versions of *Floresta Sintá(c)tica* treebank: Portuguese treebank *Bosque* 8.0 [1] and *Universal Dependencies Portuguese* treebank (*UD_Portuguese*) [16], which aims at full compatibility with CoNLL UD specifications.

In order to adapt the parsers to be used by the OIE system, we implemented some shallow conversion rules to align the tagset and dependency names of *Bosque* 8.0 and *UD_Portuguese* to the PoS tags and dependency names used by the OIE system. This is not a full and deep conversion since the OIE system only uses a small list of PoS tags and dependencies. So, before training a parser on the Portuguese treebank, first we must identify the specific PoS tags and dependencies used by the extraction module, and second, we have to change them by the corresponding labels. For *UD_Portuguese*, we also have to change the syntactic criteria on preposition dependencies. Concerning the rule-based parsers, no adaptation is required since the OIE system is based on the dependency labels of *DepPattern*. A priori, this could benefit systems that did not have to be adapted, but we have no way of measuring it.

² <http://github.org/citiususc/linguakit>.

³ <http://www.maltparser.org/>.

To evaluate the results of the OIE system with the parsers defined above, five systems were configured, each one with a different parser. OIE evaluation is inspired by that reported in [4, 8]. The dataset consists of 103 sentences from a domain-specific corpus, called *CorpusEco* [18], containing texts on ecological issues. These sentences were processed by the 5 extractors, given rise to 862 triples. Then, each extracted triple was annotated as correct (1) or incorrect (0) according to some evaluation criteria: triples are not correct if they denote incoherent and uninformative propositions, or if they are constituted by over-specified relations, i.e., relations containing numbers, pronouns, or excessively long phrases. We follow similar criteria to those defined in previous OIE evaluations [4, 5]. Annotation was made on the whole set of extracted triples without identifying the system from which each triple had been generated.

The results are summarized in Table 1. *Precision* is defined as the number of correct extractions divided by the number of returned extractions. *Recall* is estimated by identifying a pool of relevant extractions which is the total number of different correct extractions made by all the systems (this pool is our gold-standard). So, *recall* is the number of correct extractions made by the system divided by the total number of correct expressions in the pool (346 correct triples in total).⁴

Table 1. Evaluation of five OIE systems configured with five dependency parsers

| Systems | Precision | Recall | Fscore |
|------------------------|-----------|--------|--------|
| deppattern-ArgOE | .440 | .265 | .330 |
| deppattern-Linguakit | .612 | .361 | .454 |
| maltparser-nivrearc | .581 | .248 | .347 |
| maltparser-2planar | .516 | .228 | .316 |
| maltparser-nivrearc-ud | .616 | .236 | .341 |

The results show that there is no clear difference among the evaluated systems except in the case of *deppattern-Linguakit*, which relies on a rule-based parser. However, a deeper analysis allows us to observe that rule-based and data-driven parsers might be complementary parsers as they merely share about 25% of the correct triples. More precisely, the number of correct extractions made by *deppattern-Linguakit* reaches 125 triples, but only 30 of them are also extracted by *maltparser-nivrearc*. This means that a voting OIE system consisting of the two best rule-based and data-driver parsers would improve recall in a very significant way without losing precision.

⁴ Labeled extractions along with the gold standard are available at https://gramatica.usc.es/~gamallo/datasets/OIE_Dataset-pt.tgz.

4 Conclusions

In this article, we showed that it is possible to use OIE systems to easily compare parsers developed with different strategies, by making use of a coarse-grained and shallow adaptation of tagsets and syntactic criteria. By contrast, comparing very different parsers by means of direct evaluation is a much harder task since it requires carrying out deep changes on the training corpus (golden treebank). These changes involve adapting tagsets before training, reconsidering syntactic criteria at all analysis level and yielding the same tokenization as the golden treebank. Moreover, the proposed task-oriented evaluation might help linguists make deep error analysis of the parsers since the extraction of basic propositions allows humans to visualize and interpret linguistic mistakes in an easier way than obscure syntactic outputs.

Acknowledgments. Pablo Gamallo has received financial support from a 2016 BBVA Foundation Grant for Researchers and Cultural Creators, TelePares (MINECO, ref:FFI2014-51978-C2-1-R), the Consellería de Cultura, Educación e Ordenación Universitaria (accreditation 2016–2019, ED431G/08) and the European Regional Development Fund (ERDF). Marcos Garcia has been funded by the Spanish Ministry of Economy, Industry and Competitiveness through the project with reference FFI2016-78299-P, by a Juan de la Cierva grant (IJCI-2016-29598), and by a 2017 Leonardo Grant for Researchers and Cultural Creators, BBVA Foundation.

References

1. Afonso, S., Bick, E., Haber, R., Santos, D.: Floresta sintá(c)tica: a treebank for Portuguese. In: The Third International Conference on Language Resources and Evaluation, LREC 2002, pp. 1698–1703, Las Palmas de Gran Canaria, Spain (2002)
2. Banko, M., Cafarella, M.J., Soderland, S., Broadhead, M., Etzioni, O.: Open information extraction from the web. In: International Joint Conference on Artificial Intelligence, pp. 2670–2676 (2007)
3. Collovini, S., Machado, G., Vieira, R.: Extracting and structuring open relations from Portuguese text. In: Silva, J., Ribeiro, R., Quaresma, P., Adami, A., Branco, A. (eds.) PROPOR 2016. LNCS (LNAI), vol. 9727, pp. 153–164. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41552-9_16
4. Del Corro, L., Gemulla, R.: ClausIE: clause-based open information extraction. In: Proceedings of the World Wide Web Conference (WWW-2013), pp. 355–366, Rio de Janeiro, Brazil (2013)
5. Etzioni, O., Fader, A., Christensen, J., Soderland, S., Mausam, M.: Open information extraction: the second generation. In: International Joint Conference on Artificial Intelligence, pp. 3–10. AAAI Press (2011)
6. Fader, A., Soderland, S., Etzioni, O.: Identifying relations for open information extraction. In: Conference on Empirical Methods in Natural Language Processing, pp. 1535–1545. ACL (2011)
7. Gamallo, P.: Dependency parsing with compression rules. In: Proceedings of the 14th International Workshop on Parsing Technology (IWPT 2015), Bilbao, Spain, pp. 107–117. Association for Computational Linguistics (2015)

8. Gamallo, P., Garcia, M.: Multilingual open information extraction. In: Pereira, F., Machado, P., Costa, E., Cardoso, A. (eds.) EPIA 2015. LNCS (LNAI), vol. 9273, pp. 711–722. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23485-4_72
9. Gamallo, P., Garcia, M.: Dependency parsing with finite state transducers and compression rules. *Inf. Process. Manag.* (2018). Accessed 5 June 2018
10. Gamallo, P., Garcia, M., Fernández-Lanza, S.: Dependency-based open information extraction. In: ROBUS-UNSUP 2012: Joint Workshop on Unsupervised and Semi-Supervised Learning in NLP at the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012), Avignon, France, pp. 10–18 (2012)
11. Gómez-Rodríguez, C., Alonso-Alonso, I., Vilares, D.: How important is syntactic parsing accuracy? An empirical evaluation on sentiment analysis. *Artif. Intell. Rev.* 1–17 (2017, forthcoming). <https://doi.org/10.1007/s10462-017-9584-0>
12. Gómez-Rodríguez, C., Nivre, J.: A transition-based parser for 2-planar dependency structures. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, ACL 2010, Stroudsburg, PA, USA, pp. 1492–1501 (2010)
13. Claro, D.B., de Oliveira, L.S., Glauber, R.: Dependente: an open information extraction system on Portuguese by a dependence analysis. In: Proceedings of XIV Encontro Nacional de Inteligência Artificial e Computacional, pp. 271–282 (2017)
14. Mausam, M., Schmitz, M., Soderland, S., Bart, R., Etzioni, O.: Open language learning for information extraction. In: Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 523–534 (2012)
15. Nivre, J., et al.: Maltparser: a language-independent system for data-driven dependency parsing. *Nat. Lang. Eng.* **13**(2), 115–135 (2007)
16. Rademaker, A., Chalub, F., Real, L., Freitas, C., Bick, E., de Paiva, V.: Universal dependencies for Portuguese. In: Proceedings of the Fourth International Conference on Dependency Linguistics (Depling), pp. 197–206, Pisa, Italy, September 2017
17. Santos, V., Pinheiro, V.: Report: um sistema de extração de informações aberta para a língua Portuguesa. In: Proceedings of the X Brazilian Symposium in Information and Human Language Technology (STIL), Natal, RN, Brazil, pp. 191–200 (2015)
18. Zavaglia, C.: O papel do léxico na elaboração de ontologias computacionais: do seu resgate à sua disponibilização. In: Martins, E.S., Cano, W.M., Filho, W.B.M. (eds.) *Linguística IN FOCUS - Léxico e morfofonologia: perspectivas e análises*, pages 233–274. EDUFU, Uberlândia (2006)