

PassPort: A Dependency Parsing Model for Portuguese

Leonardo Zilio(✉), Rodrigo Wilkens, and Cédrick Fairon

Centre de traitement automatique du langage – CENTAL, Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium
{leonardo.zilio,rodrigo.wilkens,cedrick.fairon}@uclouvain.be

Abstract. Parsers are essential tools for several NLP applications. Here we introduce PassPort, a model for the dependency parsing of Portuguese trained with the Stanford Parser. For developing PassPort, we observed which approach performed best in several setups using different existing parsing algorithms and combinations of linguistic information. PassPort achieved an UAS of 87.55 and a LAS of 85.21 in the Universal Dependencies corpus. We also evaluated the model's performance in relation to another model and different corpora containing three genres. For that, we annotated random sentences from these corpora using PassPort and the PALAVRAS parsing system. We then carried out a manual evaluation and comparison of both models. They achieved very similar results for dependency parsing, with a LAS of 85.02 for PassPort against 84.36 for PALAVRAS. In addition, the results from the analysis showed us that better performance in the part-of-speech tagging could improve our LAS.

Keywords: Dependency parsing · Parsing performance
Universal Dependencies · Parsing for Portuguese

1 Introduction

The processing of Portuguese has evolved much in the past years. We saw new corpora being created, and new tools emerge that came to cover the lack of resources that we formerly had in different areas of language processing. Still, there is some ground to cover, and one of the tools required for processing a natural language, the dependency parsing, has not fared that well compared, for instance, to the state of the art for English (e.g. neural parsers, such as [5,24]).

At the same time, the introduction of the Universal Dependencies (UD) [14], a project that developed freely available, dependency-annotated corpora for multiple languages, presents new corpora for Portuguese, and, coinciding with that, other studies present a series of new state-of-the-art parsing algorithms with a relatively simple training interface.

Supported by the Walloon Region (Projects BEWARE 1510637 and 1610378) and Altissia International.

© Springer Nature Switzerland AG 2018
A. Villavicencio et al. (Eds.): PROPOR 2018, LNAI 11122, pp. 479–489, 2018.
https://doi.org/10.1007/978-3-319-99722-3_48

In this paper, we will be focusing on dependency parsing for the Portuguese language, but we do not aim at conceiving a new parsing algorithm. We took our inspiration from the work of Silva et al. [18] for developing a battery of tests, this time having dependency parsing as main focus and using the Universal Dependency (UD) corpus for Portuguese. Our objective here is thus to test several setups and evaluate their performances with different algorithms. Among the tested algorithms, we selected the one with best performance and compared it with a widely used parsing system for Portuguese. To achieve that, we first directly compared the results of different parsing algorithms in the context of the UD for Portuguese, and, later, we compared the performances across different dependency formalisms. Our hypothesis is that the recent development in dependency parsing task allows for training a model for Portuguese using a black-box approach that outperforms a parser that was deeply customized for a specific language.[1]

This paper is organized as follows: in Sect. 2, we present existing parsing systems and briefly describe their algorithms; Sect. 3 then describes the Universal Dependency corpus for Portuguese that we use as basis for developing our model; in Sect. 4, we present the methodology and results for different models that were trained; in Sect. 5, we compare the best model with the PALAVRAS parsing system by means of a manual evaluation of dependency parsing accuracy; then, in Sect. 6, we make some considerations about the tag sets employed by the different formalisms; lastly, we present our final remarks in Sect. 7.

2 Related Work

Since we are interested in dependency parsing, this section will revolve around the state of the art of dependency parsing. We especially focus on the results for Portuguese of the CONLL-X shared task on Multilingual Dependency Parsing [4]. First, we briefly present parsing algorithms, focusing on those that were used for training a model for Portuguese. We then explore existing dependency parsers for Portuguese.

The approaches presented in CONLL-X may be organized in two categories [9]: graph-based (e.g., the MaltParser [12]) and transition-based (e.g., the MST Parser [8,10] and the Stanford Parser [5]). In terms of algorithms for choosing dependency pairs, the MST Parser uses an online, large-margin learning algorithm [7], MaltParser employs Support Vector Machine, and the Stanford Parser takes advantage of neural network learning [5]. By comparing those three parsing algorithms, the results of Chen and Manning [5] for Chinese and English point to a better performance of the Stanford Parser, followed by the MST Parser. The *CONLL-X 2006* [4] used the Bosque corpus [1] as basis for the Portuguese language, and the LAS of the systems were all above 70. The best results were 87.60 (MaltParser [13]), followed by 86.8 (MST Parser [10]).

[1] The parser model, along with the material that was used in this paper can be found in https://cental.uclouvain.be/resources/smalla_smille/passport/.

Apart from the CONLL shared task, among the existing systems that cover dependency parsing for Portuguese, probably the most well known is the PALAVRAS Parsing System [3]. This system provides full parsing stack, while also annotating semantic information and several other features that can be applied to both the Brazilian and the European variants. The system is based on a Constraint Grammar and reports a performance of 96.9 in terms of LAS in a five-thousand-word sample [3].

Another system that provides dependency parsing for Portuguese is the LX-DepParser[2], which was trained using the MST Parser [8,10] on the CINTIL corpus [2] and reports an unlabeled attachment score (UAS) of 94.42 and a labeled attachment score (LAS) of 91.23.

Finally, Gamallo [6] presented the DepPattern, a dependency parsing system that uses a rule-based finite-state parsing strategy [6,15,16]. Its algorithm minimizes the complexity of rules by using a technique driven by the "single-head" constraint of Dependency Grammar. It was compared with MaltParser using Bosque (version 8). MaltParser achieved an UAS of 88.2 and DepPattern, 84.1.

3 Resources

For training the parser models, we used the Portuguese Universal Dependency (PT-UD) corpus [17][3]. The PT-UD corpus has 227,792 tokens and 9,368 sentences. It was automatically converted from the Bosque corpus [1], which was originally annotated with the PALAVRAS parser [3], and then revised. This corpus contains samples from Brazilian and European Portuguese, and is available in three separate sets: training, test and development.

For testing different setups of dependency parsing for Portuguese, we used different linguistic information and three off-the-shelf parsing systems, which were already introduced in Sect. 2: Stanford Parser 3.8.0 [5], MST Parser 0.5.0 [8], and MaltParser 1.9.1 [12].

4 Dependency Parsing

In this section, we use the resources presented so far in a series of experiments. First, we describe how we organized the setups for the experiments and then we compare the systems among themselves. In the comparison subsection, we first test how much each individual feature contributes to dependency parsing, and then we apply different combinations of these features to train and compare the performance of existing parsing algorithms for Portuguese.

[2] lxcenter.di.fc.ul.pt/services/pt/LXServicesParserDepPT.html.

[3] By the time of the execution of the experiments in this paper, the available PT-UD corpus was in its version 2.1.

4.1 Setup Organization

The first step was to establish different setups that could be used to test the different linguistic information that was available in the corpus. There are four main categories of information available in the PT-UD corpus: surface form, lemma form, short part of speech (short POS), and long part of speech (long POS). The difference between short and long POS reflects the richness of the Portuguese morphology, so that the short POS presents only the word class, while the long POS displays more detailed morphosyntactic information on top of the word class (e.g., person, number, tense). The short POS can normally be automatically derived from the long POS, but there are some ambiguous cases in the corpus[4].

Before going further into the setups, it is important to highlight that we cleaned the long POS field in the corpus, so that all tags that were between angular brackets in the long POS information were deleted, since these represent various types of information that are not always morphosyntactic[5].

From the three systems that were employed for training, all use extensively, per default, the surface and long POS information from the training file, and the Stanford Parser and the MST Parser have an influence of the lemma information[6]. To assure that the parser would receive only the information that we wanted, all information that was not relevant was set to "_" (i.e., underline) in the training, test and development sets. Since the Stanford Parser also uses embedding information during training, we used a model with 300 dimensions[7] that was trained on the brWaC corpus [20–22] using word2vec [11].

4.2 System Comparison

At first, we wanted to observe which of the four main linguistic features contributed the most for the dependency parser accuracy. As such, we tested four setups that contained only one feature (surface, lemma, short POS, or long POS), aiming to evaluate, as a secondary hypothesis, if the addition of morphology has an impact on the dependency parsing task (long versus short POS). Results have shown that the Stanford Parser model was superior in all four individual features, and they ranked from long POS (LAS = 82.74) to short POS (LAS = 79.82), to lemma (LAS = 77.54), and, finally, to surface (LAS = 74.28).

We then followed up with various setups using two features. This time, as we can see in Table 1, it was made clear that, on the morphosyntactic aspect, the long POS is superior to short POS in all setups; however, on the lexical side, the differences in the setups with lemma and surface were not significant

[4] For instance, the tag *DET* in the short POS appears as *DET* or *ART* in the long POS, while the tag *DET* in the long POS appears as *DET* or *PRON* in the short POS.

[5] This modified version of the corpus is available along with the parser model at the PassPort website https://cental.uclouvain.be/resources/smalla_smille/passport/.

[6] We detected some fluctuation in the scores during preliminary testing.

[7] Zeman et al. [23] argue that larger dimensions may yield better results for parsing.

(95% confidence)[8]. We can also see that the Stanford Parser outranks the other two in performance, achieving consistently better scores.

Table 1. Setups using two features as basis (UAS: unlabeled attachment score; LA: label accuracy; LAS: labeled attachment score)

System	Score	Lemma + POS_{short}	Lemma + POS_{long}	POS_{short} + POS_{long}	Surface + POS_{short}	Surface + POS_{long}
Stanford	UAS	85.92	**87.17**	86.28	86.32	**86.90**
	LA	90.80	**92.42**	91.70	90.98	**91.98**
	LAS	83.01	**84.88**	83.53	83.20	**84.29**
MST	UAS	84.57	85.45	85.00	85.19	85.60
	LA	88.54	89.64	89.18	88.96	89.61
	LAS	80.22	81.61	80.85	80.89	81.67
Malt	UAS	84.96	85.29	84.73	84.47	85.09
	LA	88.43	89.39	89.51	88.25	88.95
	LAS	81.59	82.73	81.83	81.15	82.42

Lastly, since the Stanford Parser and the MST Parser do present some fluctuations in the score when lemma information is added to the mix, we created two further setups for these two parsers, both using surface and lemma, but one using only short POS and the other, only long POS. The results have shown that there was no significant difference (with 95% confidence) in any of the measures (UAS, LA, and LAS).

By looking at these results, we can conclude that, in terms of dependency parsing, it is possible to choose one type of lexical information (either surface or lemma) and one morphosyntactic information and it is enough to have good results, but the richer the morphosyntactic information, the better (long POS proved to be significantly better than short POS)[9]. It is also clear that the Stanford Parser yielded the best results for the task, outperforming the other two in all setups that were trained.

After testing this battery of setups, we focused on improving the parser output quality and, for that, we trained a new embeddings model. Up until now, we have been using a model with 300 dimensions, but Chen and Manning [5] suggest using a model of 50 dimensions. So we trained a new embeddings model, by applying word2vec [11] on the raw-text brWaC corpus [20–22], and the results did improve significantly (95% confidence). In Table 2, we present our two previous best setups trained using the new embeddings model, and, in fact, the use of less dimensions proved to be better.

[8] The best system was run five times with randomized train and test sets.

[9] Using the most recent PT-UD corpus (version 2.2) in similar setups, we also had a better performance using long POS information over short POS.

Table 2. Stanford Parser: Two best models using embeddings of 50 dimensions (UAS: unlabeled attachment score; LA: label accuracy; LAS: labeled attachment score.)

System	Score	Lemma + POS_{long}	Surface + POS_{long}
Stanford	UAS	87.48	87.55
	LA	92.12	92.41
	LAS	85.00	85.21

Since the UD presents two corpora for Portuguese (one with only Brazilian Portuguese and the one that we used with both European and Brazilian variants), we also tested the performance of the Stanford Parser on the Brazilian UD corpus (BR-UD)[10]. The BR-UD corpus features only surface and short POS, so we used only these features, and the LAS of the model was 87.30. This corpus yields a better score, but it also has fewer information, and it is dedicated to only one variant of the Portuguese language. For the remainder of this paper, we will refer to our best model that uses surface and long POS from the PT-UD (with LAS of 85.21) as *PassPort* (Parsing System for Portuguese). PassPort is the model that we compare with PALAVRAS in the next section.

5 Parsing: Manual Evaluation

After comparing several parsing models, we wanted to compare the results of PassPort with those from one of the most well-known and customized parsers for Portuguese: the PALAVRAS parsing system [3]. Since both parsers employ different tag sets and formalisms, a direct evaluation of both systems using a single gold standard is not possible. To bridge these two different tag sets and organization of dependency parsing, we designed a manual evaluation using as basis a single corpus of 90 randomly selected sentences from three different genres[11].

The selected genres were literature[12], newspaper articles (from the Diário Gaúcho corpus[13]) and subtitles (from the Portuguese corpus of subtitles compiled by [19]). Thirty sentences were randomly extracted from each of these corpora and all of them were then parsed using PassPort and PALAVRAS. The genres present very different sentence sizes, so here we present the evaluated token account for the three samples: 471 tokens for newspaper, 182 tokens for subtitles, and 642 tokens for literature.

[10] Available at: https://github.com/UniversalDependencies/UD_Portuguese-GSD/tree/master.

[11] Although there are 30 sentences selected from each genre, in the results, it is possible to observe that both parsing systems (PassPort and PALAVRAS) use their own sentence splitters, so that the final sentence numbers are different (for instance, PALAVRAS splits sentences when there is a colon).

[12] Selected romances from www.dominiopublico.gov.br.

[13] This corpus was compiled in the scope of the project PorPopular (www.ufrgs.br/textecc/porlexbras/porpopular/index.php).

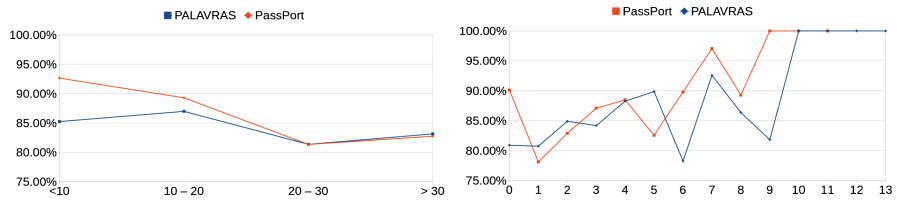
The annotation of both parsers was manually evaluated by one linguist in terms of accuracy (UAS, LA, and LAS), respecting the individual assumptions of each parser (tags, tag order, attachment patterns etc.). The results of the evaluation are shown in Table 3. In the table the results are shown in terms of evaluated tokens[14] and full sentences (sentences in which all tokens were correct for the given measure). The results show that both parsers are very similar in the tested corpus: in terms of tokens, PALAVRAS gets better dependency parsing in the newspaper subcorpus, but PassPort has superior dependency parsing for subtitles and literature and also in the full corpus; in terms of full sentences, PALAVRAS has better results for literature, but PassPort fares better in the full corpus and individually for newspaper articles and subtitles. The differences, however, are small for both sides, and both systems perform very similarly in terms of LAS. In terms of part of speech, PassPort is worse, achieving 94.59% of accuracy against PALAVRAS' 97.53% in the full corpus.

Table 3. Accuracy evaluation of PassPort and the PALAVRAS parsing system (UAS: unlabeled attachment score; LA: label accuracy; LAS: labeled attachment score)

		Newspaper		Subtitles		Literature		Total	
		PassPort	PAL	PassPort	PAL	PassPort	PAL	PassPort	PAL
Tokens	UAS	88.75	**89.56**	**96.70**	90.75	**89.41**	89.36	**90.19**	89.63
	LA	88.32	**91.42**	**92.86**	89.02	**88.79**	87.23	**89.19**	88.97
	LAS	84.93	**87.70**	**92.86**	86.71	**82.87**	81.34	**85.02**	84.36
Sentences	UAS	**43.33**	40.00	**90.32**	68.75	30.00	**43.90**	**54.95**	50.49
	LA	**36.67**	30.00	**70.97**	65.63	26.67	**34.15**	**45.05**	42.72
	LAS	**30.00**	26.67	**70.97**	62.50	16.67	**29.27**	**39.56**	38.83

Following the work of McDonald and Nivre [9], we further investigated the parsing results of the manually evaluated corpus. We start by looking at the labeled attachment score (LAS) in function of the length of the sentences. After dividing the sentences in ranges of evaluated tokens (10, 20, 30+ tokens), we analyzed their mean LAS. The results are shown in Fig. 1a. As we can see, PassPort performed better at lower sentence lengths and was a bit worse in longer sentences (more than 30 words); however, a t-test ($p < 0.05$) reveals that these results are not significantly different. We also evaluated how the deepness of the dependency (i.e., the distance of the token in relation to the root) affects the LAS. The results in Fig. 1b indicate that both parsers perform well even in deeper dependencies.

[14] We did not evaluate punctuation tokens, since PALAVRAS does not provide dependency label for them and, in both parsing models, they are simply attached to the root or the closest dependency to the root.



(a) LAS versus sentence length (b) LAS versus deepness

Fig. 1. Analysis of sentence length and deepness in relation to LAS

6 Discussion

As we could see in Sect. 5, PassPort performs well and is on par with PALAVRAS. Even so, there are some considerations to be made in terms of the dependency tags for both parsers.

Regarding the Universal Dependencies (UD), which were used in PassPort, at least in the PT-UD corpus that was used for training, the tag *obl* is not very informative, since it applies both to adjuncts and to indirect objects introduced by preposition (dative pronouns are tagged as *iobj*)[15]. The UD also present no tag for predicative relations, since the copula verbs are always attached to the predicative (which receives a *root* or a clausal tag). This is much more richly done by PALAVRAS, which presents different tags both for predicatives and for distinguishing indirect objects and adjuncts (but the one for adjuncts doesn't have a good label accuracy – LA – in our corpus: 77.9).

In the case of the tags presented by the PALAVRAS parsing system, the two most frequent tags in our evaluation corpus are *@N* and *@P*[16]. Both of these tags, have a LA higher than 95.4, but they do not describe a dependency relation, they only indicate that the token is attached to a token with a certain part of speech (noun or preposition, respectively). As such, these labels are redundant in the annotation. This is also true for some less frequent tags, such as *@A*, which indicates attachment to an adjective. These cases are better represented in the UD, which presents a label for the relations, and not only the attachment. In addition, PALAVRAS does not consider parataxis, which could pose a problem for annotating oral texts and more freely written language.

7 Final Remarks

In this paper, we trained a new dependency parsing model for Portuguese based on the Universal Dependencies. We used the PT-UD corpus and trained several

[15] This is not in line with the UD guidelines (universaldependencies.org/u/dep/iobj.html), which indicate that the indirect objects should be marked as *obj* (if they are the sole object of the verb) or as *iobj* (if there is another *obj* in the clause). According to the guidelines, *obl* should only be used for adjuncts, but that is not the case in the PT-UD corpus.

[16] The tags present also a < or > symbol, which indicates the attachment direction.

different parsing models based on different lexical and morphological information before selecting the best setup. During the testing phase, we compared three parsing systems (MST, MaltParser, and Stanford Parser) in terms of their performance. Stanford Parser presented the best results in all setups.

After the testing phase, we used our best setup and trained a new parsing model, which we called PassPort. Aiming at observing how PassPort compare to another dependency parser for Portuguese, we compiled a corpus of sentences from different genres, and we then used this common corpus to manually evaluate the accuracy of PassPort against the PALAVRAS parsing system. This evaluation showed that both parsers performed very similarly in terms of the standard parsing scores (unlabeled attachment score, label accuracy, and labeled attachment score). We then ran some further analysis to evaluate the performance of the labeled attachment score in relation to sentence length and deepness of the dependency (distance to the root), and we saw that, here too, both models perform very similarly.

Regarding our hypothesis that the recent development in the dependency parsing task allows for training a model for Portuguese using a black-box approach that outperforms a highly customized parser, we could see that PassPort competes toe to toe with PALAVRAS, having a slight edge on the scores[17].

Overall, PassPort had a performance that is compatible to the state of the art in Portuguese and also in other languages (according to the results of Chen and Manning [5] for English and Chinese using the Stanford Parser). This performance could perhaps be improved if we had delved deeper into the tuning of the parser model, and possibly also if we had dedicated the same attention to the part-of-speech tagging as we dedicated to the dependency parsing model. This remains, however, as a future development of PassPort.

References

1. Afonso, S., Bick, E., Santos, D., Haber, R.: Floresta sintá (c) tica: um "treebank" para o português. quot. In: Gonçalves, A., Correia, C.N., (eds.) Actas do XVII Encontro Nacional da Associação Portuguesa de Linguística (APL 2001), Lisboa 2–4 de Outubro de 2001, Lisboa Portugal: APL (2001)
2. António, B., Castro, S., Silva, J., Costa, F.: Cintil depbank handbook: design options for the representation of grammatical dependencies. Department of Informatics, University of Lisbon, Technical reports nb. di-fcul-tr-11-03, pp. 86–89 (2011)
3. Bick, E.: The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. Aarhus Universitetsforlag (2000)
4. Buchholz, S., Marsi, E.: CoNLL-X shared task on multilingual dependency parsing. In: Proceedings of the Tenth Conference on Computational Natural Language Learning, pp. 149–164. Association for Computational Linguistics (2006)
5. Chen, D., Manning, C.: A fast and accurate dependency parser using neural networks. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 740–750 (2014)

[17] The model, training datasets and evaluation files will be made available with the final version.

6. Gamallo, P.: Dependency parsing with compression rules. In: Proceedings of the 14th International Conference on Parsing Technologies, pp. 107–117 (2015)
7. McDonald, R., Crammer, K., Pereira, F.: Online large-margin training of dependency parsers. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pp. 91–98. Association for Computational Linguistics (2005)
8. McDonald, R., Lerman, K., Pereira, F.: Multilingual dependency analysis with a two-stage discriminative parser. In: Proceedings of the Tenth Conference on Computational Natural Language Learning, pp. 216–220. Association for Computational Linguistics (2006)
9. McDonald, R., Nivre, J.: Analyzing and integrating dependency parsers. Comput. Linguist. **37**(1), 197–230 (2011)
10. McDonald, R., Pereira, F.: Online learning of approximate dependency parsing algorithms. In: 11th Conference of the European Chapter of the Association for Computational Linguistics (2006)
11. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
12. Nivre, J., Hall, J., Nilsson, J.: MaltParser: a data-driven parser-generator for dependency parsing. In: International Conference on Language Resources and Evaluation, vol. 6, pp. 2216–2219 (2006)
13. Nivre, J., Hall, J., Nilsson, J., Eryiğit, G., Marinov, S.: Labeled pseudo-projective dependency parsing with support vector machines. In: Proceedings of the Tenth Conference on Computational Natural Language Learning, pp. 221–225. Association for Computational Linguistics (2006)
14. Nivre, J., et al.: Universal dependencies v1: a multilingual treebank collection. In: International Conference on Language Resources and Evaluation (2016)
15. Otero, P.G., González, I.: DepPattern: a multilingual dependency parser. In: International Conference on Computational Processing of the Portuguese Language (PROPOR 2012), Coimbra, Portugal, pp. 659–670. Citeseer (2012)
16. Otero, P.G., López, I.G.: A grammatical formalism based on patterns of part of speech tags. Int. J. Corpus Linguist. **16**(1), 45–71 (2011)
17. 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), Pisa, Italy, pp. 197–206, September 2017. http://aclweb.org/anthology/W17-6523
18. Silva, J., Branco, A., Castro, S., Reis, R.: Out-of-the-box robust parsing of Portuguese. In: Pardo, T.A.S., Branco, A., Klautau, A., Vieira, R., de Lima, V.L.S. (eds.) PROPOR 2010. LNCS (LNAI), vol. 6001, pp. 75–85. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12320-7_10
19. Tiedemann, J.: Finding alternative translations in a large corpus of movie subtitle. In: International Conference on Language Resources and Evaluation (2016)
20. Filho, J.A.W., Wilkens, R., Zilio, L., Idiart, M., Villavicencio, A.: Crawling by readability level. In: Silva, J., Ribeiro, R., Quaresma, P., Adami, A., Branco, A. (eds.) PROPOR 2016. LNCS (LNAI), vol. 9727, pp. 306–318. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41552-9_31
21. Wagner Filho, J., Wilkens, R., Idiart, M., Villavicencio, A.: The brWaC corpus: a new open resource to aid in the processing of Brazilian Portuguese. In: 11th edition of the Language Resources and Evaluation Conference (LREC) (2018)
22. Wagner Filho, J.A., Wilkens, R., Villavicencio, A.: Automatic construction of large readability corpora. In: Workshop on Computational Linguistics for Linguistic Complexity (CL4LC), p. 164 (2016)

23. Zeman, D., et al.: CoNLL 2017 shared task: multilingual parsing from raw text to universal dependencies. In: Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, pp. 1–19 (2017)
24. Zhou, H., Zhang, Y., Huang, S., Chen, J.: A neural probabilistic structured-prediction model for transition-based dependency parsing. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, vol. 1: Long Papers, pp. 1213–1222 (2015)