Cross-Framework Evaluation for Portuguese POS Taggers and Parsers

Sandra Collovini[1], Henrique D.P. Santos[1], Thiago Lima[1], Evandro Fonseca[1], Bolivar Pereira[1], Marlo Souza[2], Sílvia Moraes[1] and Renata Vieira[1]

{sandra.abreu, henrique.santos.003, thiago.lima.001, evandro.fonseca, bolivar.pereira}@acad.pucrs.br, msouza1@ufba.br, {silvia.moraes, renata.vieira}@pucrs.br

[1]Pontifícia Universidade Católica do Rio Grande do Sul
Porto Alegre – Rio Grande do Sul – Brazil

[2]Universidade Federal da Bahia
Salvador – Bahia – Brazil

Abstract. This work compares POS and parsing systems for the Portuguese language. We analyse available features, tagsets, and compare the results of POS tagging, and syntactic structure identification by means of both intrinsic and extrinsic evaluation methods. For such, we use in this work well-known metric for parser evaluation such as bracket cross, leaf ancestor for intrinsic evaluation, as well as the application of such parsers to the task of noun phrase identification, for extrinsic evaluation. The comparison proposed in this work takes into account the different linguistic theories and frameworks each parser subscribes to, but it is not dependent of any particular one.

Keywords: Parser evaluation, Portuguese parsers, Portuguese POS

1 Introduction

Parsing technology has increased greatly in the last decades, giving rise to a number of robust automatic parsers available in the field. Particularly, the rise of statistical parsers allied with machine learning methods for syntactic structure prediction, allowed the easy construction of automatic parsers based on annotated treebanks. With the proliferation of such tools, however, the problem of comparing their results have become apparent, a task commonly known as parser evaluation.

Such an evaluation is, in fact, not a simple task. The reason for this is that the parsers (and the treebanks used to construct them) are usually based on competing linguistic theories and frameworks, and the syntactic structure of a same sentence can diverge significantly according to each framework. As such, while one can easily apply well-established methodologies and metrics to compare parsers constructed over the same grammar/treebank, it is still an open problem how to compare two parser based on different linguistic theories or domains.

Recent research has been conducted on this problem, known as parser evaluation across linguistic domains and frameworks [7, 19, 20, 4, 27]. It is not clear yet in the literature, however, which methodology is more suitable for such a comparison.

Authors such as Carroll et al. [7], among others, have proposed that the use of representations such as Grammatical Relations (GR) are more adequate for parser evaluation than the use of tree structure. On the other hand, authors such as Mollá and Hutchinson [15] or Yuret *et al.* [27] propose that an extrinsic evaluation can provide a better way to evaluate parsing systems than intrinsic evaluations such as Grammatical Relations or syntactic tree comparison. This work acknowledges such challenges and brings some alternatives to compare parsing systems for the Portuguese language, independently of the linguistic framework to which each parser subscribes to.

This work is organized as follows: In Section 2, we present the linguistic Portuguese tools that will be compared. In Section 3, we describe the evaluation methods used, The Bosque Treebank, used as reference for the evaluation, and results are presented. Finally, Section 4 presents conclusions and future work.

2 Linguistic Annotation Tools

Linguistic annotation tools that provide information at the morpho-syntactic levels are extensively used as constituents of larger systems for many NLP tasks, and the quality of these annotations directly impacts the results of the bigger tasks. Currently, most syntactic parsers and POS taggers have been developed for the English language, with only a few available of them which can process texts in other languages, such as Portuguese. In the following, we present an overview of the morphosyntactic annotators available for the Portuguese language.

2.1 Part-Of-Speech Taggers

Part-Of-Speech (POS) tagging is an important preprocessing stage in NLP applications, and it is almost indispensable for any corpus research [10]. In order to analyze the sentence structure, for example, it is necessary to first recognize the grammatical categories of the words. Automatic POS tagger is a system responsible for identifying the grammatical category for each of the lexical items in a sentence.

In this section, we present the main, available, POS taggers for Portuguese. We analyze and compare their set of tags, since depending on the application for which the tagged text will be used, the number of tags can vary, and generally the quality of labeling directly impacts on applications performance. The POS tagger studied are presented below.

TreeTagger was developed by Helmut Schmid [22]. Besides Portuguese, it has been successfully used to tag many other languages. It can be used on any language if a lexicon and a manually tagged training corpus are available. It

is freely available for research, education and evaluation tasks. The Portuguese parameter file was provided by Pablo Gamallo.

NLTK the Natural Language Toolkit [3] is a suite of Python program modules, data sets and tutorials supporting research and teaching in computational linguistics and natural language processing. NLTK is written in Python and distributed under the GPL open source license. Over the past year the toolkit has been rewritten, simplifying many linguistic data structures and taking advantage of recent enhancements in the Python language.

NLPnet [9] is a tagger that was trained over a revision of Mac-Morpho [1], the biggest corpus of Portuguese text containing manually annotated POS tags. Many errors were corrected, yielding a much more reliable resource. We also trained a neural network based classifier for the POS tagging task, following an architecture that achieves state-of-the-art results in English.

UDPipe [26] performs tokenization, morphological analysis, part-of-speech tagging, lemmatization and dependency parsing for nearly all treebanks of Universal Dependencies,the latter is not available yet for the Portuguese language. In addition, the pipeline is easily trainable with training data in CoNLL-U format (and in some cases also with additional raw corpora) and requires minimal linguistic knowledge on the users' part. The training code is also released.

2.2 Syntactic Parsers

Syntactic parsers perform the structural analysis of phrases and their constituents. In this work we have analyzed the following parsers for the Portuguese language: Palavras, Freeling, CoGrOO, LX-Parser and MaltParser.

Palavras is an automatic tagger and parser for Portuguese that was developed by Eckhard Bick [2]. The formalism used follows the Constraint Grammar tradition (CG), introduced by Fred Karlsson [12]. PALAVRAS served as a model for the analysis of other languages in the VISL project (`http://visl.sdu.dk`), which is a core of tools and linguistic databases available for online use. Palavras provides the following information levels: morphological, POS, syntactic and dependency. This parser also has semantic prototype tags for nouns, proper nouns, verbs and same adjectives. An example output of Palavras is presented in Figure 1, corresponding to sentence (1).

(1) Fisher teria conhecido o ginecologista.
(Fisher would have known the gynecologist.)

Freeling is an open-source multilingual language processing library providing a wide range of analysis (morphological, named entity detection, PoS-tagging, parsing, Word Sense Disambiguation, Semantic Role Labelling, etc.) [18]. For Portuguese language, Freeling PT has the following functionalities available: morphological analysis, POS tagging, shallow parsing, named entity detection and classification. Figure 2 illustrates sentence (1) output provided by the Freeling PT. Freeling project was undertaken at the TALk research center to provide advances towards general availability of basic NLP tool and resources (`http://nlp.lsi.upc.edu/freeling/demo/demo.php`).

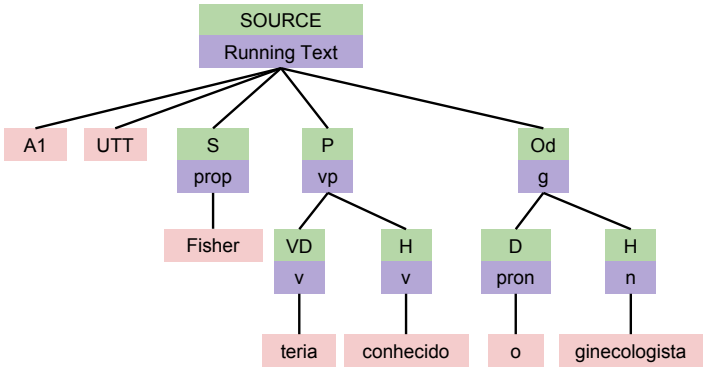

Fig. 1. Palavras Synctactic Tree (adapted)

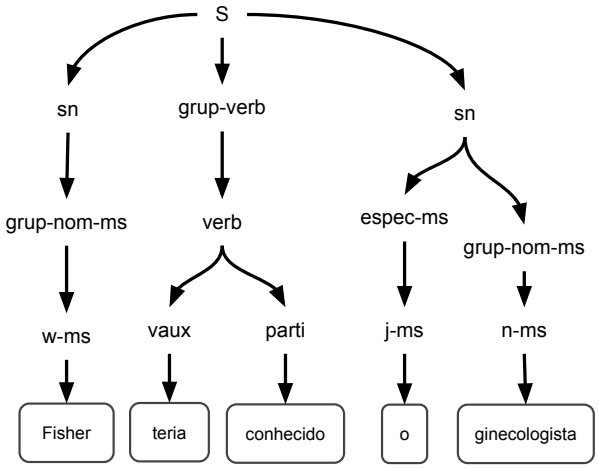

Fig. 2. Freeling Synctactic Tree (adapted)

CoGrOO [25] is an open-source grammar checker widely used for Portuguese. It is capable of identifying Portuguese grammatical errors such as pronoun placement, noun agreement, subject-verb agreement, usage of the accent stress marker, subject-verb agreement, and other common errors of Portuguese writing. Besides its use as a grammar checker, CoGrOO provides a set of linguistic annotation tools which can be used to process texts in the Portuguese language, such as a POS-taggers, chunkers and morphosyntactic annotators. In Figure 3, we can see Cogroo's output.

```
Sentence: Fisher teria conhecido o ginecologista
  Tokens:
      Fisher        []               prop    M=S
      Teria         [ter]            v-fin   COND=3S
      Conhecido     [conhecer]       v-pcp   M=S
      O             [o]              art     M=S
      ginecologista [ginecologista]  n       F=S
  Chunks: [NP: Fisher ]
          [VP: teria conhecido ]
          [NP: o ginecologista ]
  Shallow Structure: [SUBJ: Fisher ]
                     [P: teria conhecido ]
                     [ACC: o ginecologista ]
```

Fig. 3. CoGrOO– Syntactic Strucutures (adapted)

LX-Parser [24] is a robust parser for the Portuguese language freely available both as a web service and for download at `http://lxparser.di.fc.ul.pt/`. The current version of the LX-parser was built training a model for the version 1.6.5 of Standford parser [13] using the CINTIL treebank [5]. The LX-Parser syntactic tree for the running example is depicted in Figure 4.

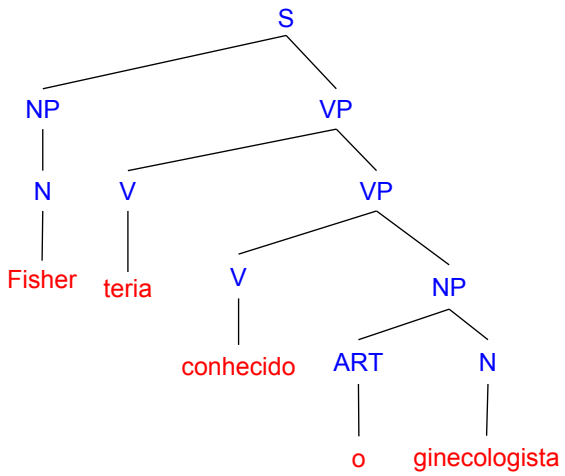

Fig. 4. LX-Parser Synctactic Tree (adapted)

MaltParser [17] is a language-independent system for dependency parser-generation which learns a deterministic dependency parser from a treebank. The

parser-generator implements the arc-eager deterministic projective parser [16], which was trained in our work using a linear Support Vector Machine [11].

In this work, we trained the parser using the Universal Dependencies treebank for the Brazilian variant of the Portuguese language [14]. In order to use this parser to annotate unseen text, we implemented a simple rule-based tokenizer for Portuguese and trained a POS tagger using the NLTK's [3] implementation of the Brill Tagger [6].

3 Tools Evaluation

In this section, we present the Bosque Treebank, from which we extracted sentences used as reference for the evaluation. The evaluation methods used and the results achieved also are described.

3.1 Bosque Treebank

Bosque is a subset of treebank for Portuguese, "Floresta Sintá(c)tica", composed of newspaper articles written in Brazilian and European Portuguese. Bosque has been automatically annotated by the Palavras parser and fully manually revised, with a current size of 9,368 sentences and 190,513 lexical units. In this work, we used version 7.4 (Brazilian Portuguese) in Constrain Grammar (CG) format [12] revised manually[1]. Figures 5 and 6 present examples of sentence (1) in CG and PennTreebank format, respectively.

We extracted 10 reference sentences from Bosque for the evaluation of the taggers and parsers, and the selection criterion was sentence size equal to or greater than the average size of Bosque sentences.

```
Em              [em]            <sam-> PRP          @ADVL>     #1->7
essa            [esse]          <dem>  DET    F S  @>N        #2->3
época           [época]                N      F S  @P<        #3->1
,                                                              #4->0
Fishel          [Fishel]               PROP   M S  @SUBJ>     #5->6
teria           [ter]           <aux>  V COND 3S   @STA       #6->0
conhecido       [conhecer]      <mv>   V PCP       @ICL-AUX<  #7->6
o               [o]             <artd> ART    M S  @>N        #8->9
ginecologista   [ginecologista]        N      F S  @<ACC      #9->7
britânico       [britânico]            ADJ    M S  @N<        #10->9
Robert=Winston  [Robert_Winston]       PROP   M S  @N<        #11->9
```

Fig. 5. Sample in CG format

Since the Bosque treebank was constructed by manually correction of the outuput of the Palavras parser, we believe that we cannot evaluate the Palavras parser in a comparative manner while using this treebank. The reason for this is

[1] http://www.linguateca.pt/floresta/corpus.html

```
(STA:fcl (ADVL:pp (H:prp:em:: Em)
                  (P<:np (>N:pron-det:esse:F_S::dem: essa)
                         (H:n:época:F_S::np-def: época)))
         (,)
         (SUBJ:np (H:prop:Fishel:M_S::: Fishel))
         (P:vp (AUX:v-fin:ter:PR_3S_COND::: teria)
               (MV:v-pcp:conhecer:::: conhecido))
         (ACC:np (>N:art:o:M_S::artd: o)
                 (H:n:ginecologista:F_S::np-def: ginecologista)
                 (N<:adjp
                   (H:adj:britânico:M_S::: britânico))
                 (N<:np (H:prop:Robert_Winston:M_S::: Robert_Winston)
                 ...
```

Fig. 6. Sample in PennTreebank format

that, undeniably, the Bosque treebank reflects several theoretical and implementation decisions of the parser that would benefit Palavras in such an evaluation. The adopted evaluation methods are presented in the next Section.

3.2 Evaluation Methods

A variety of parser evaluation methods appear in the literature. A survey about the state-of-the-art in parser evaluation methodologies and metrics is presented in [7]. According to authors, the methods can be corpus-based or based on intrinsic properties of the parsers. The corpus-based methods are divided into those using annotated corpora and those using unannotated corpora. In this work, we apply some of these methods to evaluate the POS tagger and syntactic parser, which are described below.

General Comparison: We first present a general overview of the tools regarding linguistic information provided, in order of complexity, starting with the morphological and morphosyntactic levels (tokenization, lemmatization, POS tagging, morphology: gender, number, degree, person etc.); syntactic (shallow, full parsing, dependency); semantic and Named Entity (NE).

Part-of-Speech Assignment Accuracy: This measure computes the accuracy of POS tagger assignments of grammatical categories. Accuracy is used to evaluate POS tagger, due to the fact that there is a great deal of manually-corrected POS tagged data to use as test corpora. In literature, there are many ways to calculate the accuracy, here we computed, for each POS tagger, the number of correct words tagged, compared with the reference, divided by total number of words tagged. For this, the first step was to perform a mapping between the different POS tags of each tool in comparison to the reference. The reference sentences (from the Bosque corpus) follow the tagset of the Palavras parser.

Structural Consistency: To evaluate the result of the parsers, we analyzed the parsing results using well-known metrics for constituency parsing, such as bracket crossing [23] and leaf ancestor [21] metrics. It is of importance to notice two things in our evaluation. First, the results of all parsers were converted to the Penn Treebank format to perform such evaluation. The second is that we decided to evaluate only the structure of the trees, not whether the tags were correctly labelled. Since these metrics are sensible to differences in tokenization, a step of manual tokenization correction was necessary to guarantee that the trees resulted by the parsers could be analyzed.

Four parsers were evaluated according to these metrics: CoGrOO, Freeling, LX-Parser and Malt Parser. From them, only the MaltParser - a dependency parser - returns an output not consistent with the Penn Treebank format which required transformation of the output. To perform this transformation, the following rules were applied: (i) the entire sentence is inside a constituent with the head as the root element of the syntax tree; (ii) for each token T, if there are tokens which depend on it in the syntax tree, then there is a constituent containing all the tokens which depend on T, for which the token T is the head.

Noun Phrases Assignment Accuracy: Current methods for evaluating the accuracy of syntactic parsers are based on measuring the degree to which parser output replicates the analyses assigned to sentences in a manually annotated corpus. In this work, we analyze the noun phrase (NP) contained in the 10 reference sentences compared the output parser: Cogroo, Freeling, LX-Parser, MaltParser. For this, we extract the noun phrases (the most external and also the internal ones) of both the reference and the outputs of each analyzed parser and we computed the accuracy. We consider the correct NP (when NP are equal at system output and at reference: accuracy = 1) and partially correct (when at least one NP token at system output (NPs) corresponds to a reference token (NPr): accuracy = 0.5 * (NPs / NPr) [8]).

3.3 Results

An overview of the features available for the Portuguese tools under analysis is illustrated in Table 1. We also present the language and terms/license that each tool was developed in. Most tools perform POS tagging; the annotation of lemmatization and morphological are performed by some of the tools: Palavras, Freeling, Cogroo, UDPipe and TreeTagger. There are two shallow parsers (Freeling and Cogroo), and Palavras is the only full parser. Palavras also provides semantic tags, dependency and Named Entity. It is noteworthy that the dependency information is provided also by MaltParser and Named Entities are annotated by Freeling.

Table 2 illustrates the mapping between the different POS taggers. We can see that Cogroo and NLTK use the same POS tags defined by the Palavras[2]; UDPipe and MaltParser use the Universal POS tags[3] POS tags; Freeling and

[2] https://visl.sdu.dk/visl/pt/info/portsymbol.html
[3] http://universaldependencies.org/u/pos/

TreeTagger use the EAGLES[4] POS tags; NLPnet use the tagset of Mac-Morpho; and LX-Parser use the tagset of CINTIL treebank[5]. One of the difficulties for the evaluation was the different attributes of each category considered by the POS taggers, for example, the category "verb" was considered by all the tools, however their attributes differed among them, such as type (auxiliary verb, main verb) and genre (feminine, masculine, neuter), among others.

Table 1. Tools Features Comparison

	Palavras	**Freeling**	**Cogroo**	**NLTK**	**NLPnet**	**UDPipe**	**MaltParser**	**LX-Parser**	**TreeTagger**
Language	C, Perl	C++	Java	Python	Python	Python	Java	Java	C++
Terms/License	Proprietary	Opensource	Opensource	Opensource	Opensource	Opensource	Opensource	Opensource	Opensource
Tokenization	√	√	√	√	√	√		√	√
Lemmatization	√	√	√			√			√
PoS tagging	√	√	√	√	√	√		√	√
Morphological	√	√	√			√			√
Shallow parsing		√	√						
Full parsing	√							√	
Dependency	√						√		
Semantic	√								
Named Entity	√	√							

Table 2. Portuguese Tools and Bosque Treebank Tagset

Tagset	**Bosque**	**Palavras**	**Freeling**	**Cogroo**	**NLTK**	**NLPnet**	**UDPipe**	**MaltParser**	**LX-Parser**	**TreeTagger**
definite article	ART	<artd> DET	DA	art	art	ART	DET	DET	ART	DA
indefinite article	ART	<arti> DET	DI	art	art	ART	DET	DET	ART	DI
preposition	PRP	PRP	SP	prp	prp	PREP	ADP	ADP	P	S
noun	N	N	NC	n	n	N	NOUN	NOUN	N	N
proper noun	PROP	PROP	NP	prop	prop	NPROP	PROPN	PROPN	N	NP
personal pronoun	PERS	PERS	PP	pron-pers	pron-pers	PROPESS	PRON	PRON	PRS	PP
determiner pronoun	DET	DET	D*	pron-det	pron-det	PROADJ	DET	DET	DEM	D
independent pronoun	INDP	SPEC	PR	pron-indp	pron-indp	PRO-KS	PRON	PRON	CL	PI
adjective	ADJ	ADJ	A	adj	adj	ADJ	ADJ	ADJ	A	A
adverb	ADV	ADV	R	adv	adv	ADV	ADV	ADV	ADV	R
auxiliary verb	<aux> V	<aux> V	VA*	v-fin	v-fin	VAUX	AUX	AUX	V	VA
finite verb present	V PR	V PR	V*	v-fin	v-fin	V	VERB	VERB	V	VM
finite verb past perfect	V PS	V PS	V*	v-fin	v-fin	V	VERB	VERB	V	VM
finite verb past imperfect	V IMPF	V IMPF	V*	v-fin	v-fin	V	VERB	VERB	V	VM
infinitive verb	V INF	V INF	V*N	v-inf	v-inf	V	VERB	VERB	V	VMN
participle verb in compoound tense	V PCP	V PCP	V*P	v-pcp	v-pcp	PCP	VERB	VERB	PPT	VM
participle verb not in compoound tense	V PCP	V PCP	V*P	v-pcp	v-pcp	PCP	VERB	VERB	PPA	VM
gerund verb	V GER	V GER	V*G	v-ger	v-ger	V	VERB	VERB	GER	VMG
gerund as auxiliary verb									GERAUX	VAG
numeral	NUM	NUM	Z	num	num	NUM	NUM	NUM	CARD, ORD	Z
subordinating conjunction	KS	KS	CS	conj-s	conj-s	KS	SCONJ	SCONJ	C	CS
coordinating conjunction	KC	KC	CC	conj-c	conj-c	KC	CCONJ	CCONJ	CONJ	CC
interjection	IN	IN	I	intj	intj	IN	INTJ	INTJ		I

[4] http://www.ilc.cnr.it/EAGLES/browse.html

Table 3. Parser Evaluation Results

Tools	POS ↑	Bracket Crossing	Leaf Ancestor	NPs
Cogroo	**98.81**	**50.68**	88.15	**68.08**
Palavras	96.93	NE	NE	NE
TreeTagger	92.73	–	–	–
Freeling	91.39	31.78	88.10	53.75
UDPipe	91.24	–	–	–
LX-Parser	87.91	44.27	77.71	47.25
MaltParser	87.20	35.33	**89.90**	36.63
NLPnet	81.17	–	–	–
NLTK HMM	77.95	–	–	–
NLTK Def	72.35	–	–	–

After, we calculated the average accuracy of the sentences for each POS tagger and parser, and illustrated the results in Table 3. The best results were achieved by the Cogroo parser, with the best POS tagging performance, Crossing-bracket and noun phrase accuracy. As a open-source software Cogroo also has a large set of features, missing just Named Entity identification, that Freeling does. Palavras was not evaluated due to its influence in the reference generation (NE). MaltParser presented better performance for leaf ancestor metric. We belive the reason for the discrepancy between the results for MaltParser from one metric to the other is due to the fact that there is an impact on the transformation of the dependency structure to the phrasal structure, where the constituents are less structured than a constituency parser, however this metric is more flexible and the conversion did not impact the results.

4 Conclusions and Future Work

In this work, we show the comparison of 9 annotation tools for (POS taggers and parsers) for the Portuguese language. A correlation Table presents the variety of tagsets used across the tools analysed in this work. To correlate different parsers is a difficult task, due to the variety of tags, different underlying assumptions and their structures. To overcome such difficulties, we employed several different evaluation methods and metrics in the literature and compared all the parser according to each of these metrics. As a result, considering the method adopted in this work, the Cogroo grammar-checker was the parser that presented a better overall performance.

As future work, we intend to perform the evaluation on the entirety of the Bosque treebank. This was not possible in this work since the comparison of the results required several steps of manual intervention. For example, in comparing the results of the POS Tagging task, it was required to map the possible equivalences of the tags in each tagsets and evaluate the resulting annotation of each sentence for all tools. Using the universal dependency tagset [14], however, for which such mappings have been provided, we believe we can automatize most of

such labour-intensive analysis and provide a more comprehensive comparison of the tools.

We would also like to combine the current evaluation methodology with that of grammatical relations, by mapping each syntactic construction in the constituency parsers' tagsets to a comprehensive set of grammatical relations. We believe applying different evaluation strategies, we can realistically compare the results of syntactic parsers for the Portuguese language, despite their different linguistic foundations.

References

1. S. Aluísio, J. Pelizzoni, A. R. Marchi, L. de Oliveira, R. Manenti, and V. Marquiafável. *An Account of the Challenge of Tagging a Reference Corpus for Brazilian Portuguese*, pages 110–117. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.
2. E. Bick. *The Parsing System Palavras. Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. University of Arhus, 2000.
3. S. Bird. Nltk: the natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions*, pages 69–72. Association for Computational Linguistics, 2006.
4. J. Bos, E. Briscoe, A. Cahill, J. Carroll, S. Clark, A. Copestake, D. Flickinger, J. van Genabith, J. Hockenmaier, A. Joshi, R. Kaplan, T. H. King, S. Kübler, D. Lin, J. T. Lonning, C. Manning, Y. Miyao, J. Nivre, S. Oepen, K. Sagae, N. Xue, , and Y. Zhang, editors. *Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation*. Association for Computational Linguistics, 2008.
5. A. Branco, J. Silva, F. Costa, and S. Castro. Cintil treebank handbook: Design options for the representation of syntactic constituency. Technical Report TR-2011-02, Universidade de Lisboa, 2011.
6. E. Brill. Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational linguistics*, 21(4):543–565, 1995.
7. J. Carroll, T. Briscoe, and A. Sanfilippo. Parser evaluation: a survey and a new proposal. In *Proceedings of the 1st International Conference on Language Resources and Evaluation*, pages 447–454, 1998.
8. N. C. Diana Santos and N. Seco. Avaliação no harem: Métodos e medidas. In Diana Santos and Nuno Cardoso (eds.), Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área. Linguateca. Departamento de Informática, Faculdade de Ciências da Universidade de Lisboa. DI-FCUL TR-06-17, Linguateca, november 2006 2007.
9. E. R. Fonseca and J. L. G. Rosa. Mac-morpho revisited: Towards robust part-of-speech tagging. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*, pages 98–107. sn, 2013.
10. E. Giesbrecht and S. Evert. Is part-of-speech tagging a solved task? an evaluation of pos taggers for the german web as corpus. In I. Alegria, I. Leturia, and S. Sharoff, editors, *Proceedings of the 5th Web as Corpus Workshop (WAC5)*, San Sebastian, Spain, 2009.
11. M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

12. F. Karlsson, A. Voutilainen, J. Heikkila, and A. Anttila, editors. *Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text*. Mouton de Gruyter, Berlin, 1995.
13. D. Klein and C. D. Manning. Fast exact inference with a factored model for natural language parsing. In *Advances in neural information processing systems*, pages 3–10, 2003.
14. R. T. McDonald, J. Nivre, Y. Quirmbach-Brundage, Y. Goldberg, D. Das, K. Ganchev, K. B. Hall, S. Petrov, H. Zhang, O. Täckström, et al. Universal dependency annotation for multilingual parsing. In *ACL (2)*, pages 92–97, 2013.
15. D. Mollá and B. Hutchinson. Intrinsic versus extrinsic evaluations of parsing systems. In *Proceedings of the EACL 2003 Workshop on Evaluation Initiatives in Natural Language Processing: are evaluation methods, metrics and resources reusable?*, pages 43–50. Association for Computational Linguistics, 2003.
16. J. Nivre. An efficient algorithm for projective dependency parsing. In *Proceedings of the 8th International Workshop on Parsing Technologies (IWPT*. Citeseer, 2003.
17. J. Nivre, J. Hall, J. Nilsson, A. Chanev, G. Eryigit, S. Kübler, S. Marinov, and E. Marsi. Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135, 2007.
18. L. Padró and E. Stanilovsky. Freeling 3.0: Towards wider multilinguality. In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, Istanbul, Turkey, 2012. ELRA.
19. J. Preiss. Using grammatical relations to compare parsers. In *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics-Volume 1*, pages 291–298. Association for Computational Linguistics, 2003.
20. K. Sagae, Y. Miyao, T. Matsuzaki, and J. Tsujii. Challenges in mapping of syntactic representations for framework-independent parser evaluation. In *the Workshop on Automated Syntatic Annotations for Interoperable Language Resources*, 2008.
21. G. Sampson and A. Babarczy. A test of the leaf-ancestor metric for parse accuracy. *Natural Language Engineering*, 9(4):365–380, 2003.
22. H. Schmid. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK, 1994.
23. S. Sekine and M. Collins. Evalb bracket scoring program. *URL: http://www. cs. nyu. edu/cs/projects/proteus/evalb*, 1997.
24. J. R. Silva, A. Branco, S. Castro, and R. Reis. Out-of-the-box robust parsing of portuguese. In *PROPOR*, volume 6001, pages 75–85. Springer, 2010.
25. W. D. C. Silva. Aprimorando o corretor gramatical cogroo. In *Master's Dissertation*. Universidade de São Paulo, 2013.
26. M. Straka, J. Hajic, and J. Straková. Udpipe: Trainable pipeline for processing conll-u files performing tokenization, morphological analysis, pos tagging and parsing. In *LREC*, 2016.
27. D. Yuret, A. Han, and Z. Turgut. Semeval-2010 task 12: Parser evaluation using textual entailments. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 51–56. Association for Computational Linguistics, 2010.