

An NLP approach to impersonal *–se* in Brazilian Portuguese

Elvis A. de Souza¹, Magali S. Duran¹, Adriana S. Pagano²

¹Interinstitutional Center for Computational Linguistics (NILC)
University of São Paulo (USP)

²Federal University of Minas Gerais (UFMG)

{elvis.desouza99, magali.duran}@gmail.com, apagano@ufmg.br

Abstract. This paper introduces an annotation proposal for the reflexive pronoun *–se* in Brazilian Portuguese with a view to classifying different strategies for impersonalization through the use of one supercategory. We carried out experiments on a gold standard treebank for Portuguese in the Universal Dependencies project and verified that the implementation of our proposal results in the training of a morphosyntactic annotation model that annotates syntactic dependencies 1.27 percentage point better in accuracy. Moreover, a detailed evaluation showed an increase of up to 6.34 accuracy in the annotation of verb arguments, one of the most important classes for carrying out various Natural Language Processing tasks, highlighting the importance of informed linguistic modeling decisions in practical NLP results.

1. Introduction

In the field of Natural Language Processing (NLP), the annotation of reflexives presents a unique challenge. Traditional grammatical approaches often fail to support analyses that can yield consistent results from a cross-linguistic perspective. This, in turn, hampers the cross-comparability that projects such as Universal Dependencies (UD) [De Marneffe et al. 2021] aim to achieve, resulting in NLP models that can not perform well in the analysis of syntactic relations. Poor performance in parsing impacts several downstream tasks that would benefit from solid annotation.

In this paper, we approach annotation of reflexives in Brazilian Portuguese through an experiment which seeks to evaluate the impact of annotating two traditionally separate uses of the pronoun *–se*, namely subject indeterminacy and synthetic passive, with a single dependency relation tag, *expl:impers*, standing for “impersonal”, given their similar characteristics. This approach is meant to offer several advantages, such as recognizing impersonalization strategies and eliminating controversial grammatical agreement issues.

Our hypothesis is that this unification will enhance automatic learning of syntactic relations implicating verb arguments, which are vital to tasks such as information extraction and Q&A. To validate our approach, we performed a conversion of tags in PetroGold [de Souza and Freitas 2023], a Brazilian Portuguese Universal Dependencies treebank with well-defined guidelines for the annotation of the pronoun *–se*, and evaluated the impact of the conversion on the quality of an NLP parsing model [Straka et al. 2016].

2. The pronoun *–se* in Brazilian Portuguese

In Brazilian Portuguese, the pronoun *–se* can have an argumental and a non-argumental role in a verb’s argument structure. It has an argumental role when it functions as direct or

indirect object in a clause. In such cases, it can construe reflexive or reciprocal meanings and is co-referential with the subject of the clause, as illustrated in examples (1) and (2)¹.

(1) Ele não se perdoou pelo erro
he not REFL.3 forgive-PST.PERF.3SG for the mistake
He did not forgive himself for the mistake

(2) Eles se abraçaram depois do jogo
they RECP.3 hug-PST.PERF.3PL after of the match
They hugged each other after the match

A non-argumental role of *-se* is implicated in two particular uses. The first involves verbs selecting a reflexive pronoun to construe a specific meaning, derived or not from a transitive meaning. This use is relevant semantically, but not syntactically. In this case, there is agreement between subject person and reflexive pronoun person, as seen in examples (3), (4) and (5).

(3) Eu me abstive de votar
I REFL.1SG refrain- PST.PERF.1SG from voting
I refrained from voting

(4) O conflito se estendeu por meses
the conflict REFL.3 extend- PST.PERF.3SG for months
The conflict dragged on for months

(5) A plateia silenciou se
the audience silence- PST.PERF.3SG REFL.3
The audience fell silent

A second non-argumental use of *-se* is in impersonal constructions, where it can only operate as a third person pronoun for verbs in third person singular or plural forms. When the verb selecting *-se* is intransitive (6) or indirect transitive (7), *-se* is considered a marker of subject indeterminacy and the verb has a third person singular form. However, when the verb is direct transitive, *-se* is considered a marker for passive voice and its complement to the right classified as a passive subject, which requires it to agree in number with the verb, as seen in examples (8) and (9).

3. Reflexives in the Universal Dependencies framework

Within the UD framework, reflexives are annotated following their distinction in terms of whether they have an argumental role or not. Argumental reflexives are pronouns realizing objects (tagged *obj* or *iobj*) and in some languages obliques (*obl*). Non-argumental

¹ Examples include a gloss and a free translation to assist readers, particularly when *-se* does not translate into a corresponding pronoun in English.

(6) Caiu se numa armadilha
 fall-PST.PERF.3SG REFL.3 into a trap
We fell into a trap

(7) Pensou se em tudo
 think-PST.PERF.3SG REFL.3 in everything
We thought of all available possibilities

(8) Utilizou se uma única teoria
 use-PST.PERF.3SG REFL.3 a single theory
We used a single theory / A single theory was used

(9) Levantaram se várias hipóteses
 raise- PST.PERF.3PL REFL.3 several hypotheses
We raised several hypotheses / Several hypotheses were raised

reflexives are categorized as expletives. The latter include three types of expletives for reflexive markers: (1) markers of impersonal meanings (tagged *expl:impers*), (2) reflexive pronouns construing passive meanings (tagged *expl:pass*) and (3) reflexive pronouns selected by pronominal verbs (tagged *expl:pv*).

Given their extensive use and frequency in several language families, reflexives have been target of several studies, both as one among other expletive categories and as a category in itself. All studies point out inconsistencies in annotation with impact on UD's much desired cross-comparability among treebanks. [Bouma et al. 2018], for instance, discuss reflexives within the broad class of expletives and advocate current UD guidelines. [Marković and Zeman 2018] focus on reflexive markers and perform a cross-linguistic comparison of patterns in three language families: germanic, romance and slavic. They explore UD treebanks and find many inconsistencies in their annotation. They strongly advocate annotating reflexive markers by selecting “Yes” for the Reflex feature as a form of quickly querying treebanks for these markers. Regarding reflexive expletives, they advocate prioritizing the annotation of inherent reflexives (*expl:pv*), on the one hand, and impersonal and passive reflexives (*expl:impers* and *expl:pass*), on the other.

[Degraeuwe and Goethals 2020] also report problems in treebank annotation consistency, focusing on Spanish in particular. They advocate using existing UD tags for argumental (*obj*, *iobj*) and non-argumental (*expl:pv*, *expl:impers*, *expl:pass*) roles, but propose annotating fine-grained distinctions regarding reflexive and non-reflexive functions by means of feature properties.

4. A proposal for simplifying non-argumental and non-reflexive *-se* in Portuguese

Annotating the three subrelations proposed by UD guidelines in Brazilian Portuguese texts is actually a very hard and time-consuming task, as it involves lengthy discussions which sometimes do not achieve a consensus among annotators. This can seriously impact consistency in treebanks, a problem reported by the above-mentioned studies. There are

also considerations which challenge some of the assumptions underlying the expletive tags.

The tag *expl:pass* implicates a passive voice reading, which is not the case in Brazilian Portuguese, as the verb form is not passive. Moreover, these constructions do not admit expression of a passive agent, as is the case in passive voice constructions. A further issue in Brazilian Portuguese is increasing loose agreement between the verb form and the complement to the right, which raises discussions as to whether the complement is an object or a passive subject, as illustrated by example (10).

(10) Vê se muitas crianças sozinhas na rua
 see-PRS.3SG REFL.3 many children alone in the street
You can see many children alone on the streets

In fact, [Magalhães and Carvalho 2021] analyzed a collection of sentences from contemporary Brazilian newspapers and found that variation in subject-verb agreement for synthetic passives is around 50%, both with finite verbs and in the infinitive.² According to traditional grammar [Bechara 2012, Bechara 2018, Cunha and Cintra 2016], the lack of agreement would be considered a grammatical mistake, since in passive constructions there must be subject-verb agreement. We conclude that the phenomenon of synthetic passives is in the process of merging with that of indeterminacy in Brazilian Portuguese, a thesis also supported by authors such as [Bagno 2012], justifying the choice to merge them in the annotation as well.

Considering these cases as strategies for construing impersonal meanings would allow for strengthening comparability with other languages which have active voice constructions for similar meanings, as is the case of pronouns “you” and “one” in English, “on” in French and “uno” in Spanish.

The above considerations led us to propose to use a single tag, namely *expl:impers*, to annotate cases formerly annotated as *expl:pass*. Several advantages ensue from this proposal. Impersonal active voice constructions of the kind we have illustrated will be treated as impersonalization strategies and their complement considered an object. Agreement issues between verb form and complement will be solved as both third person singular and plural forms will be equally annotated as objects. Constructions with a direct transitive verb selecting *-se* and having a clausal complement (*cf.* example 11) will be considered impersonal constructions as well. Finally, passive voice (tagged Voice=Pass) will be a feature assigned exclusively to past participle constructions, both in reduced and fully expanded clauses.

(11) Via se que todos estavam desesperados
 see- PST.PERF.3SG REFL.3 that everyone was desperate
You could see that everyone was in despair

²Some examples from the analysis include: “A urgência de se *aprovar* leis imprescindíveis” (lit. “The urgency of *passing*(singular) essential laws(plural)”) and “não se *pode* perder os amigos” (lit. “one can’t lose(singular) their friends(plural)”).

5. Methodology

In order to demonstrate the advantages of our proposal for the purposes of treebank annotation, we carried out experiments using PetroGold, a gold standard treebank comprising academic text in the oil and gas domain [de Souza et al. 2021] (8,946 sentences, 250,605 tokens). We chose PetroGold as currently this is the only dataset in the UD project (v2.14) that clearly distinguishes the different uses of the pronoun *—se*.³

PetroGold annotates non-argumental uses of the pronoun *—se* with UD labels *expl:impers*, *expl:pass* and *expl:pv*. Annotation draws on traditional grammars of Brazilian Portuguese, which classify the pronoun *—se* as (i) a marker of subject indeterminacy (annotated as *expl:impers*), (ii) a pronoun for passive constructions (annotated as *expl:pass*) and (iii) a pronoun selected by pronominal verbs (*expl:pv*). These three cases are illustrated by examples (12), (13) and (14) retrieved from PetroGold.

- (12) *expl:impers*: Quando **se** **fala** em PHPA, **fala** **se** em a união de a acrilamida e de o acrilato de sódio através de a copolimerização.⁴
- (13) *expl:pass*: Para a síntese **usou** **se** manta aquecida, temperatura de refluxo e agitação mecânica.⁵
- (14) *expl:pv*: Este estudo **se** **baseia** em as propriedades magnéticas de os minerais que **se** **concentram** em as rochas de a crosta terrestre.⁶

In order to join occurrences of *—se* labeled as *expl:pass* and *expl:impers* in PetroGold, we took the following steps:

- 1) The *expl:impers* label was retained in all annotated cases.
- 2) Cases previously annotated with *expl:pass* were annotated as *expl:impers* (Figure 1 exemplifies an impersonal meaning previously annotated as *expl:pass*)

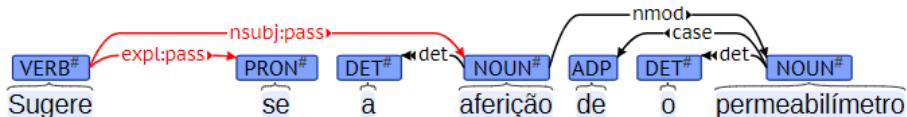


Figure 1. Original annotation of a sentence using *expl:pass*

- 3) The feature *Voice=Pass* was reserved for passive voice constructions, with a past participle form and passive voice auxiliary “ser”.
- 4) Subjects of impersonal constructions previously annotated as *expl:pass* (*nsubj:pass*) were annotated as objects of active voice constructions (*obj*) (Figure 2 shows changes made to example in Figure 1). The *nsubj:pass* tag was hence reserved for passive voice constructions, with a past participle form.

³Bosque [Rademaker et al. 2017] contains a residual annotation of *expl:pass* for only 4 verb lemmas selecting the pronoun; PUD [Zeman et al. 2017] annotates all *—se* pronouns as *expl:pv*, CINTIL [Branco et al. 2022] makes no use of any UD *expl* tags to annotate the pronoun *—se*, and Portinari [Duran et al. 2023] annotates all non-argumental *—se* as *expl*.

⁴Transl. “When we **refer** to PHPA, we **are referring** to the union of acrylamide and sodium acrylate through copolymerization.”

⁵Transl. “For the synthesis, a heated blanket, reflux temperature and mechanical stirring **were used**.”

⁶Transl. “This study **is based** on the magnetic properties of the minerals that **are concentrated** in the rocks of the Earth’s crust.”

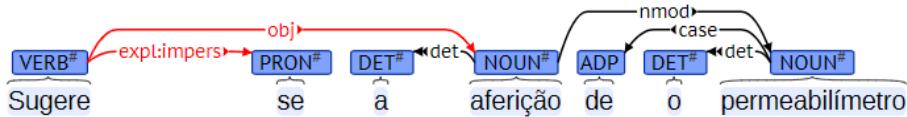


Figure 2. Current annotation of a former *expl:pass* sentence as *expl:impers*

To evaluate the quality of an NLP parsing model trained on the dataset after implementing our conversion, we trained two annotation models based on UDPipe 1.2.0 [Straka et al. 2016], one before and the other after the conversion, without any change in the model hyperparameters. The models were evaluated using the CLAS (content-label attachment score) metric, which indicates the correctness of syntactic relations and dependencies attachment for words with lexical content [Zeman et al. 2018], and the differences are shown as absolute percentage points.

Furthermore, to detail each of the tags impacted by this modification to the tagset, we used the metrics UAS (unlabeled-attachment score) and LAS (labeled-attachment score), which measure the hit rate of syntactic dependencies attachment ignoring the relation label and considering it, respectively [Zeman et al. 2018]. Other metrics are UPOS (universal part-of-speech), UFeats (universal morphological features) and Lemmas (correct prediction of lemmas), which are not directly related to the experiments conducted but may present minor impact due to the nature of neural networks training in a pipeline fashion [Straka et al. 2016].

6. Results

The treebank originally contained a total of 578 subjects (nsubj:pass) of impersonal constructions previously labeled as *expl:pass*, which were converted into active voice objects of verbs, and a total of 793 pronouns –*se* annotated with the label *expl:pass*, which were re-annotated as *expl:impers*. In total, 1,371 tokens were impacted, which corresponds to 0.5% of the tokens in the treebank. In terms of number of sentences in the corpus, 710 sentences were re-annotated, which corresponds to 7.9% of the sentences in the corpus.

Table 1 shows the results for the evaluation metrics of the trained model after the conversions were performed. Δ is the difference between the current performance value and the value before the conversions were performed. We can see a slight improvement in all metrics related to morphological features (UPOS, UFeats, Lemmas). Regarding these metrics, the only expected difference would be “UFeats”, since we removed the feature *Voice=Pass* from finite verb forms previously categorized with it. Metrics related to syntactic analysis (UAS, LAS and CLAS) show larger improvement, of up to 1.27 CLAS increase – a considerable increase, taking into account that only 0.5% of the tokens were impacted in the corpus.

To examine the parser’s performance for each of the linguistic categories in the dataset, we inspected the annotation model’s predictions on the PetroGold test partition (1,039 of the 8,946 sentences, or 11.6%). This partition contains 147 labels of type *expl:impers* resulting from the conversion of *expl:pass* and 78 objects (obj) resulting from the conversion of passive subjects (nsubj:pass).

Table 2 indicates the performance of the trained parser for each of the treebank’s syntactic categories, where “Freq.” stands for the number of times the label appears in

Metric	%	Δ
UPOS	98.43	0.02
UFeats	98.17	0.09
Lemmas	98.81	0.04
UAS	91.04	0.60
LAS	89.38	0.77
CLAS	84.57	1.27

Table 1. Evaluation metrics for the model trained on the new annotation

the test portion of the dataset, “Hits” for the number of times the parser annotated the dependency relation correctly, and “LAS” when it predicted the correct dependency attachment as well. The three last columns indicate the difference when comparing the data and the model performance before and after the conversions, and column Δ Freq. confirms that only tokens with relations *nsubj:pass*, *obj* and *expl:impers* were re-annotated. For space reasons, we display in the table only categories that have increased or decreased performance in a number higher than 1.27, which is the increase in CLAS of the model.

Relation	Freq.	Hits	LAS	Δ Freq.	Δ Hits	Δ LAS
acl	592	93.58%	79.90%	0	0.84	2.37
acl:relcl	212	88.21%	68.87%	0	-2.83	0.47
advmod	746	93.43%	79.49%	0	-0.81	1.61
appos	234	79.06%	61.11%	0	-1.71	0.85
aux	40	72.50%	70.00%	0	-10.00	-10.00
cc	765	97.52%	88.50%	0	0.53	1.83
ccomp	140	75.00%	74.29%	0	2.86	3.58
conj	1025	88.98%	60.78%	0	0.39	3.02
cop	339	96.76%	92.63%	0	2.07	1.77
expl:impers	224	89.73%	89.29%	147	68.95	68.51
expl:pv	71	92.96%	91.55%	0	4.23	4.23
flat:name	181	83.98%	83.43%	0	-1.10	-1.65
nsubj:pass	380	88.68%	88.16%	-78	0.69	1.48
nummod	496	94.96%	92.34%	0	1.41	1.82
obj	819	93.04%	91.33%	78	5.46	5.50
obl	1486	85.94%	78.60%	0	1.55	1.88
obl:agent	102	95.10%	95.10%	0	1.96	1.96
obl:arg	142	49.30%	48.59%	0	6.34	5.63
parataxis	119	77.31%	68.91%	0	0.00	3.36
xcomp	298	87.58%	87.58%	0	1.67	2.01

Table 2. Performance of the annotation model for some dependency tags

To verify our initial hypothesis, i.e., that merging impersonal types of *-se* is promising for improving results for NLP tasks such as question and answer (Q&A) and information extraction, results for specific categories need to be taken into account. These were verb arguments – *nsubj* (+0.57), *nsubj:pass* (+1.48), *obj* (+5.50), *obl:arg* (objects introduced by preposition, +6.34), *ccomp* (objects in form of subordinate clauses, +3.58). Hence, annotation by the trained model after merging impersonal *-se* showed improve-

ment for all classes that implicate arguments, even those that were not directly affected in the gold annotation.

One of the classes that showed the highest improvement was *obj*, which can be accounted for by the fact that the proposed annotation labels arguments to the right of verbs associated with the pronoun *-se* as objects, with no more room for the model to consider them subjects. Another class with a considerable increase was *ccomp*, which can also be accounted for by the fact that constructions with a direct transitive verb selecting *-se* and having a clausal complement are not passive constructions, as formerly considered, and their complement should be tagged as *ccomp*.

The class *obl:arg*, used for verbal arguments introduced by prepositions, increased 6.34 points. Although an improvement in verbal arguments was expected, prepositional arguments were not impacted in the dataset, since only direct objects (former patient subjects) were targeted.⁷ Moreover, a class which had a 10 point decrease in the model performance in comparison to the model trained before our experiment was *aux*, used in Brazilian Portuguese to construe tense through auxiliary verbs (“ter”, “haver”, “ir”, “estar”). Such large differences merit further investigation, being possibly due to the parser neural network training procedure, which happens in a pipeline fashion, yielding no explainability for results.

7. Conclusion

Constructions with *-se* traditionally seen as construing passive meanings in Brazilian Portuguese pose a challenge to annotators, as they behave differently from regular passive constructions. They do not allow for the expression of an agent; they take a complement to the right of the verb, a position typically filled by objects; they have an active verb form, and exhibit an increasingly loose agreement between verb and complement. Our annotation considers these constructions as impersonalization strategies which can be readily annotated by means of the *expl:impers* tag. Our results show improvements in the model’s recognition of classes and our proposal is expected to allow for quicker decision-making and better consistency in treebank annotation.

There are some limitations, though, to the proposal implemented and evaluated in our study. One is the fact that there are no occurrences of pronoun *-se* as an object (*obj*) or an indirect object (*iobj*) in PetroGold, which restricts the scope of the analysis. Another one is the fact that no extrinsic evaluation was performed to verify the impact of the proposed simplification. Future work should be done to evaluate the effect of the proposed annotation of the pronoun *-se* in NLP applications.

Acknowledgments

This work was carried out at the Center for Artificial Intelligence, University of São Paulo (<http://c4ai.inova.usp.br/>), with support of São Paulo Research Foundation FAPESP Grant 2019/07665-4) and ABM Corporation. The project was supported by the Ministry of Science, Technology and Innovation, with resources of the National Council of Science and Technology (CNPQ) (313103/2021-6).

⁷Prepositional objects are tagged as *obl:arg* in UD, which roughly correspond to indirect objects in traditional grammars.

References

[Bagno 2012] Bagno, M. (2012). *Gramática pedagógica do português brasileiro*. Parábola Ed.

[Bechara 2012] Bechara, E. (2012). *Moderna gramática portuguesa*. Nova Fronteira.

[Bechara 2018] Bechara, E. (2018). *Lições de português pela análise sintática*. Nova Fronteira.

[Bouma et al. 2018] Bouma, G., Hajic, J., Haug, D., Nivre, J., Solberg, P. E., and Øvrelid, L. (2018). Expletives in Universal Dependency treebanks. In de Marneffe, M.-C., Lynn, T., and Schuster, S., editors, *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 18–26, Brussels, Belgium. Association for Computational Linguistics.

[Branco et al. 2022] Branco, A., Silva, J., Gomes, L., and Rodrigues, J. (2022). Universal grammatical dependencies for portuguese with cintil data, lx processing and clarin support. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5617–5626.

[Cunha and Cintra 2016] Cunha, C. and Cintra, L. (2016). *Nova gramática do português contemporâneo*. LEXIKON Editora Digital ltda.

[De Marneffe et al. 2021] De Marneffe, M.-C., Manning, C. D., Nivre, J., and Zeman, D. (2021). Universal dependencies. *Computational linguistics*, 47(2):255–308.

[de Souza and Freitas 2023] de Souza, E. and Freitas, C. (2023). Explorando variações no tagset e na anotação universal dependencies (ud) para português: Possibilidades e resultados com base no treebank petrogold. In *Anais do XIV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 125–134. SBC.

[de Souza et al. 2021] de Souza, E., Silveira, A., Cavalcanti, T., Castro, M., and Freitas, C. (2021). Petrogold – corpus padrão ouro para o domínio do petróleo. In *Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 29–38, Porto Alegre, RS, Brasil. SBC.

[Degraeuwe and Goethals 2020] Degraeuwe, J. and Goethals, P. (2020). Reflexive pronouns in Spanish Universal Dependencies. *PROCESAMIENTO DEL LENGUAJE NATURAL*, 64(64):77–84.

[Duran et al. 2023] Duran, M. S., Lopes, L., Nunes, M. d. G. V., and Pardo, T. A. S. (2023). The dawn of the porttinari multigenre treebank: introducing its journalistic portion. *Anais*.

[Magalhães and Carvalho 2021] Magalhães, H. L. P. and Carvalho, H. M. d. (2021). Uso variável da concordância verbal em construções de voz passiva sintética na escrita de textos jornalísticos cearenses.

[Marković and Zeman 2018] Marković, S. and Zeman, D. (2018). Reflexives in universal dependencies.

[Rademaker et al. 2017] Rademaker, A., Chalub, F., Real, L., Freitas, C., Bick, E., and De Paiva, V. (2017). Universal dependencies for Portuguese. In *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017)*, pages 197–206.

[Straka et al. 2016] Straka, M., Hajic, J., and Straková, J. (2016). UDPipe: trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, pos tagging and parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4290–4297.

[Zeman et al. 2018] Zeman, D., Hajic, J., Popel, M., Potthast, M., Straka, M., Ginter, F., Nivre, J., and Petrov, S. (2018). Conll 2018 shared task: Multilingual parsing from raw text to universal dependencies. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual parsing from raw text to universal dependencies*, pages 1–21.

[Zeman et al. 2017] Zeman, D., Popel, M., Straka, M., Hajic, J., Nivre, J., Ginter, F., Luotolahti, J., Pyysalo, S., Petrov, S., Potthast, M., Tyers, F., Badmaeva, E., Gokirmak, M., Nedoluzhko, A., Cinkova, S., Hajic jr., J., Hlavacova, J., Kettnerová, V., Uresova, Z., Kanerva, J., Ojala, S., Missilä, A., Manning, C. D., Schuster, S., Reddy, S., Taji, D., Habash, N., Leung, H., de Marneffe, M.-C., Sanguinetti, M., Simi, M., Kanayama, H., dePaiva, V., Droganova, K., Martínez Alonso, H., Çöltekin, c., Sulubacak, U., Uszkoreit, H., Macketanz, V., Burchardt, A., Harris, K., Marheinecke, K., Rehm, G., Kayadelen, T., Attia, M., Elkahky, A., Yu, Z., Pitler, E., Lertpradit, S., Mandl, M., Kirchner, J., Alcalde, H. F., Strnadová, J., Banerjee, E., Manurung, R., Stella, A., Shimada, A., Kwak, S., Mendonca, G., Lando, T., Nitisoroj, R., and Li, J. (2017). 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*, pages 1–19, Vancouver, Canada. Association for Computational Linguistics.