

# ET: uma Estação de Trabalho para revisão, edição e avaliação de corpora anotados morfossintaticamente

Elvis de Souza  
Departamento de Letras  
PUC-Rio  
Rio de Janeiro, Brasil  
elvis.desouza99@gmail.com

Cláudia Freitas  
Departamento de Letras  
PUC-Rio  
Rio de Janeiro, Brasil  
claudiafreitas@puc-rio.br

**Palavras-chave**—revisão de corpora, métodos de avaliação, anotação automática, processamento de linguagem natural, linguística computacional

## I. INTRODUÇÃO

Sistemas de anotação morfossintática (pos-taggers e parsers) que utilizam tecnologia de aprendizado de máquina demandam corpora volumosos e bem anotados para aprenderem a anotar textos adequadamente. De modo geral, visando melhorar a qualidade dos anotadores automáticos, fez-se muito em relação à tecnologia subjacente aos sistemas, o que elevou drasticamente, no decorrer dos anos, a qualidade da anotação gramatical. No entanto, há um gargalo na qualidade do material que serve de treino, o que acaba impactando na qualidade da aprendizagem [1]. No corpus Bosque-UD [2], de textos jornalísticos em Língua Portuguesa adaptados para o formato Universal Dependencies [3], as métricas de POS, quando o modelo é treinado utilizando a ferramenta UDPipe [4], são relativamente confortáveis: 96,46% para a versão 2.3 do Bosque-UD. Contudo, esse mesmo material em outros níveis de análise linguística, como o de relações sintáticas, leva a resultados de 81,82%, indicando que ainda há um grande espaço para melhorias.

Acreditamos, junto com Manning [1], que um caminho para superar esses gargalos deve ser pela via linguística: melhorando a anotação dos corpora que servem de treino para os sistemas de anotação, tornando-os mais consistentes e eliminando possíveis erros humanos.

Nesse contexto, este trabalho relata a construção de uma Estação de Trabalho (ET) desenhada a partir da perspectiva linguística, com o objetivo de facilitar a revisão, a edição e a avaliação de corpora anotados, alinhando o trabalho feito pelos especialistas em língua, de um lado, e os resultados práticos, isto é, o desempenho de sistemas de PLN, de outro. Desse modo, discussões teóricas sobre as categorias gramaticais podem ser embasadas não apenas quanto à adequação linguística a certas teorias, mas também nos resultados empíricos de sistemas de aprendizado artificial.

Nossa Estação de Trabalho compreende, até o momento, dois eixos centrais: 1) a edição e revisão do corpus anotado; e 2) a avaliação do resultado da revisão, feita a partir da análise da anotação automática. Apresentaremos a arquitetura

das ferramentas participantes da nossa ET tendo em vista que foram desenvolvidas objetivando duas tarefas complexas, em momentos distintos: primeiro, a contabilização de sujeitos ocultos no corpus Bosque-UD (seção 2), e posteriormente, o lançamento de uma nova versão, intensamente revisada, do mesmo corpus (seção 3).

É importante ressaltar que, embora estejamos apresentando as ferramentas publicamente, nosso objetivo, no lugar de simplesmente compartilhar os códigos empacotados<sup>1</sup> e descrever como os implementamos, é relatar o que nos levou a desenvolvê-los, destacando a importância de estruturar os fenômenos nos corpora sob diferentes perspectivas que possam motivar o trabalho linguístico no PLN, fomentando o diálogo entre as duas áreas.

## II. CONTEXTUALIZAÇÃO

Começamos o desenvolvimento da ET, inicialmente, a partir de uma tarefa complexa, porque envolve contar algo que não tem materialidade: a contabilização de sujeitos ocultos no corpus Bosque-UD.

Para a quantificação de sujeitos ocultos em um corpus, precisamos, antes, tomar decisões qualitativas, isto é, foi necessário decidir quais frases consideraríamos “sujeito oculto”. Toda contagem é, antes, uma forma de qualificar os fenômenos que se nos apresentam [5], e decisões importantes como essas demandam, além de conhecimento teórico/gramatical, ferramentas que nos permitam visualizar os dados de que dispomos a fim de uma tomada de decisão com respaldo nas ocorrências da língua em uso. Evidentemente, nosso ponto de partida para a discussão foram as gramáticas, no entanto, sozinhas, elas não apresentam respostas para todos os fenômenos com que precisamos lidar em um corpus real, mesmo em casos simples como os sujeitos ocultos.

Nesse contexto, desenvolvemos o Interrogatório, um ambiente de busca em corpora no formato CoNLL-U escrito em Python e Javascript. À medida que nos aprofundamos na caça às sentenças com sujeito oculto, novas utilidades eram demandadas, de tal maneira que o ambiente foi sendo construído tendo em mente que, a qualquer momento, uma nova necessidade poderia emergir.

<sup>1</sup>Disponíveis em <https://github.com/alvelvis/ACDC-UD>

Disse que não conseguia vislumbrar artifícios fraudulentos ou prática de peculato no protocolo assinado por Quêrcia.





Figura 1. Frase encontrada como resultado da primeira busca por sujeitos ocultos

Nossa primeira tentativa foi a de criar uma regra geral que respondesse por grande parte das sentenças com sujeito oculto, tal como nos informam as teorias gramaticais. Desenvolvemos, então, um padrão de busca que procurasse por sentenças em que um token X não tivesse Y como seu filho — no nosso exemplo, nenhuma “raiz” de sentença poderia ter “sujeito” como seu filho. E ter como primeiro problema uma busca que envolve a ausência foi um ótimo desafio.

Diferentemente da maioria das ferramentas que lida com dependências sintáticas (como por exemplo [6] e [7]), preferimos ver nossos resultados (as frases devolvidas) como linhas de concordâncias “simples”, com o trecho buscado em negrito (com cores diferentes para cada elemento indicado na expressão de busca), sem evidenciar as relações/anotações de dependência sintática, como ilustra a figura 1.

O negrito nas palavras encontradas e a possibilidade de visualizar diferentes palavras com cores é essencial para que possamos focalizar a nossa atenção no que de fato importa. Clicando em “Mostrar anotação”, ainda, é possível visualizar como todos os tokens da sentença estão anotados morfossintaticamente no corpus, inclusive com suas relações de dependência. Com essa funcionalidade, conseguimos elaborar os próximos passos da tarefa, pois podemos identificar, morfossintaticamente, quais as características que queremos considerar e quais podemos descartar em buscas futuras.

A partir dos resultados da primeira busca, adicionamos duas utilidades ao Interrogatório, e de grande relevância quando se deseja buscar elementos específicos não apenas para a busca e contagem, mas também para a revisão de corpus: (a) a possibilidade de filtrar os resultados de uma busca realizando outras buscas dentro da primeira, e (b) a capacidade de fazer pesquisas por dependência sintática.

Estudando os resultados e realizando os filtros necessários, chegamos à porcentagem de 16,04% de frases com sujeito oculto no corpus Bosque-UD (todos os procedimentos e resultados relativos à pesquisa com sujeitos ocultos estão descritos em [8]). Além do resultado, tínhamos em mãos uma ferramenta pronta para ser usada em outros projetos, desenvolvida com base no uso e aberta às mudanças que quiséssemos implementar a qualquer momento.

### III. AVALIAÇÃO DO CORPUS

Posteriormente, nossa ET, que já contava com o Interrogatório — um ambiente em expansão e ainda incipiente no momento —, se viu diante da tarefa de nos auxiliar a identificar pontos em que nosso corpus poderia melhorar — seja na

Métricas oficiais				
Metric	Precision	Recall	F1 Score	AligndAcc
Tokens	100.00	100.00	100.00	
Sentences	100.00	100.00	100.00	
Words	100.00	100.00	100.00	
UPOS	96.45	96.45	96.45	96.45
XPOS	100.00	100.00	100.00	100.00
UFeats	94.81	94.81	94.81	94.81
AllTags	92.94	92.94	92.94	92.94
Lemmas	96.94	96.94	96.94	96.94
UAS	86.46	86.46	86.46	86.46
LAS	82.42	82.42	82.42	82.42
CLAS	75.33	74.89	75.11	74.89
MLAS	67.40	67.01	67.20	67.01
BLEX	72.05	71.63	71.84	71.63

Figura 2. Exemplo de métricas de avaliação do CoNLL 2018 Shared Task

qualidade da anotação ou nas categorias linguísticas propostas. O Bosque-UD estava se preparando para o lançamento de uma nova versão [9], e enquanto o Interrogatório já era um ambiente que nos ajudaria na tarefa de visualizar as sentenças, precisávamos ainda de uma motivação empírica para guiar nossas correções — ou, para guiar nossa busca por uma estratégia sistemática de correção.

Em [10] exploramos a estratégia de revisão por divergência de modelos, usando dados das matrizes de confusão. Para dar continuidade a essa forma de revisão, nossa ET passou a integrar a visualização das métricas de avaliação da anotação feitas por algum modelo à visualização e correção das sentenças no corpus em si. Como ferramenta de anotação, escolhemos o UDPipe [4] e o tomamos como régua para comparar os modelos treinados a partir de diferentes versões do nosso corpus.

Para avaliar a qualidade das revisões — medidas a partir do desempenho do UDPipe na partição teste do corpus — usamos as métricas oficiais do CoNLL 2018 Shared Task [11]. A partir dessas métricas (Figura 2), conseguimos avaliar globalmente se o material de treino, isto é, nosso corpus, está mais consistente em relação a uma versão anterior, estabelecendo comparações entre as métricas anteriores e posteriores às nossas alterações. Para cada alteração substancial no corpus, realizamos um novo treinamento e avaliação do sistema.

Embora nos informem sobre a qualidade da anotação automática (e, de maneira indireta, sobre a qualidade/consistência da anotação manual), essas métricas pouco nos ajudam sobre o que podemos fazer para melhorar a anotação. Para nos indicar quais categorias precisam de uma revisão atenta, então, elaboramos uma outra maneira de visualizar a anotação, que envolve a distribuição de acertos por categoria. Isto é, além dos números globais, vemos, para cada tipo de anotação linguística (dependência sintática, POS etc.) o quanto o modelo acertou ou errou, o que nos diz, indiretamente, o quanto a anotação daquela categoria está consistente. A Figura 3 nos mostra a tabela para os casos de erros de POS, e vemos, pela tabela, que se a identificação de

Acurácia por categoria gramatical			
UPOS	GOLDEN	ACERTOS	PORCENTAGEM
ADJ	444	386	86.93693693693693%
ADP	1623	1616	99.5686999383857%
ADV	364	350	96.15384615384616%
AUX	283	272	96.113074204947%
CCONJ	210	203	96.66666666666667%
DET	1561	1543	98.8468930172966%
NOUN	1925	1837	95.42857142857143%
NUM	248	230	92.74193548387096%
PART	17	14	82.35294117647058%
PRON	318	295	92.76729559748428%
PROPN	851	812	95.41715628672151%
PUNCT	1343	1343	100.0%
SCONJ	100	84	84.0%
SYM	30	29	96.66666666666667%
VERB	853	821	96.24853458382181%
X	30	3	10.0%
-	742	742	100.0%

Figura 3. Acurácia de parte das classe gramaticais ao se comparar a anotação do corpus (Golden) e a previsão do parser UDPipe

UD[2]	ADJ	ADP	ADV	AUX	CCONJ	DET	NOUN	NUM	PART	PRON	PROPN	PUNCT	SCONJ	SYM	VERB	X	_	All
UD[1]																		
ADJ	386	1	2	0	0	0	27	0	0	0	5	0	0	0	23	0	0	444
ADP	0	1616	1	0	0	4	0	0	0	0	2	0	0	0	0	0	0	1623
ADV	2	3	350	0	0	2	3	0	0	1	1	0	1	0	1	0	0	364
AUX	0	0	0	272	0	0	1	0	0	0	0	0	0	0	10	0	0	283
CCONJ	0	1	3	0	203	0	2	0	0	0	0	0	1	0	0	0	0	210
DET	1	1	3	0	0	1543	0	1	0	6	6	0	0	0	0	0	0	1561
NOUN	29	2	3	2	0	3	1837	0	0	0	41	0	0	0	8	0	0	1925
NUM	2	2	0	0	0	1	3	230	0	0	10	0	0	0	0	0	0	248
PART	0	3	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	17
PRON	0	0	0	0	0	7	2	1	0	295	2	0	5	0	0	0	0	318
PROPN	4	0	1	0	0	0	29	2	0	1	812	0	0	0	2	0	0	851
PUNCT	0	0	0	0	0	0	0	0	0	0	0	1343	0	0	0	0	0	1343
SCONJ	0	1	3	0	0	0	0	0	0	12	0	0	84	0	0	0	0	100
SYM	0	0	0	0	0	0	0	0	0	0	1	0	0	29	0	0	0	30
VERB	13	1	1	7	0	0	5	0	0	0	5	0	0	0	821	0	0	853
X	1	2	0	0	0	0	7	0	0	0	17	0	0	0	0	0	3	30
-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	742	742
All	438	1633	373	281	203	1560	1916	234	14	315	902	1343	91	29	865	3	742	10942

Figura 4. Matriz de confusão de POS (classes gramaticais). Nas linhas, a anotação do corpus original, e nas colunas, a previsão do parser UDPipe

preposições (ADP) não traz preocupações (99% de acertos), o mesmo não pode ser dito da classe dos adjetivos (ADJ), com quase 87% de acertos.

Uma vez identificada uma categoria com porcentagem insatisfatória, lançamos mão da análise de matrizes de confusão (Figura 4), continuando a estratégia já utilizada em [12], para identificar quais são as divergências entre a anotação no corpus e a previsão do parser. As informações são dispostas em páginas de extensão *.html* geradas automaticamente, possibilitando hiperlinks: em qualquer um dos números na matriz, visualizamos uma relação das frases que se encontram naquela interseção, com a palavra divergente em negrito para direcionar nossa atenção.

4 / 5

# sent\_id = CF852-4

# text = **Chocante**.

ADJ  PROPN - Comentários:

Figura 5. Frase encontrada na análise da divergência entre ADJ e PROPN. Neste caso, o sistema previu erroneamente, pois “Chocante” é adjetivo, e não nome próprio

Por exemplo, na interseção entre ADJ e PROPN (casos em que, no corpus, a palavra estava anotada como ADJ, e o sistema previu PROPN), encontramos a frase da Figura 5 (de apenas uma palavra). Na mesma página podemos visualizar, ainda, a anotação completa que consta no corpus (UD[1]) e a anotação realizada pelo sistema (UD[2]).

É importante ressaltar que tanto a tabela de “Acurácia por categoria gramatical” quanto a matriz de confusão podem ser geradas para todas as informações morfossintáticas no formato UD, tais como classe gramatical e relação sintática.

#### IV. CORREÇÃO DO CORPUS

Sendo possível analisar caso a caso quais foram as divergências entre golden e modelo (ou entre modelos distintos), podemos elaborar generalizações que nos ajudem a avançar com a revisão do material, fazendo uma revisão linguisticamente motivada e tornando o material mais consistente.

Com o Interrogatório (seção 2), podemos realizar pesquisas nos corpora a partir de 5 critérios de busca (descritos na Figura 6). Essas pesquisas são guiadas pelo nosso conhecimento linguístico e pelo que generalizamos serem erros comuns ao analisar as matrizes de confusão.

Vejam, por exemplo, mais uma divergência numerosa: entre adjuntos adverbiais e adjuntos adnominais. Frases como a da Figura 7 são muito comuns de figurarem entre as divergências. No corpus, acertadamente, “manhã” está anotado como adjunto adverbial dependente de um verbo, “ligar”. A anotação automática, porém, previu “manhã” como adjunto adnominal dependente de “televisão”, como sendo uma característica do objeto, assim como falamos “televisão de plasma” ou “televisão de LED”. Parece haver uma lógica no erro da máquina, e, de fato, trata-se de uma situação conhecida na linguística: ambiguidade do sintagma preposicionado. Devemos, portanto, procurar explorá-la para facilitar nossa correção, fazendo do uso de expressões específicas e dos filtros, já mencionados na seção 2.

No Interrogatório, para todas as frases que buscamos é possível “Abrir inquérito”, isto é, editá-las (Figura 8). Na interface, basta clicar na coluna que se deseja alterar e digitar a alteração. Então, o inquérito é salvo no corpus e as alterações são anexadas a um relatório com todos os “inquéritos” feitos pelo usuário.

Critério	Expressão	Resultados da busca
1 Regex	(.*dizer.*)\n(.*PROPN)	É uma situação absurda», <b>disse Taylor, 49.</b>  «Um monte de artistas vai querer ver os concertos», <b>diz</b> <b>Ohtake.</b>
2 Ausência de B apontando para A	root#8#nsubj csubj#8	Já não <b>há</b> o império do mal para combater.  «Não <b>damos</b> conta de atendê-los.
3 Regex Independ entes	VERB.*root::!nsubj::!csubj	Diariamente, <b>está</b> <b>promovendo</b> desfiles de moda para seus consumidores.  Não <b>houve</b> acordo para uma trégua durante a Copa.
4 Pais e filhos (dependê ncia)	\tNOUN\t.*nsubj :: \tADV\t	O <b>árbitro</b> Pinto Correia esteve <b>bem</b> durante toda a primeira parte e durante quase toda a segunda.  Está <b>longe</b> a <b>constituição</b> no Nou Camp de um novo «dream team», como o de Romário e Stoichkov, o que já enerva a direção.
5 Sintaxe em Python	token.upos == "NOUN" and token.head_token.upos == "ADJ"	O <b>problema</b> é <b>político</b> porque envolve, por exemplo, a gratuidade da educação, da saúde, da previdência mínima.  Mulher morre em rio <b>presa</b> ao <b>cinto</b> do carro

Figura 6. Exemplificação de critérios de busca no Interrogatório

1 / 48

# sent\_id = CP981-4

# text = Talvez o Presidente, quando liga a televisão de **manhã** e descobre que apesar de Hillary continuar ao seu lado o país insiste em discutir se ele deve ou não abandonar o cargo porque teve uma relação extraconjugal com Monica Lewinsky, pense, como uma das personagens de «Happiness»:

obl  nmod - Comentários:

Figura 7. Divergência comum entre adjunto adverbial e adnominal. No corpus, “manhã” está anotado como adverbial, mas o sistema previu adnominal

```
# text = Os três cortadores de cana eram de Alagoas e estavam na cidade havia 15 dias.
# source = CETENFolha n=269 cad=Cotidiano sec=soc sem=94a
# sent_id = CF269-2
# id = 1131
1 Os o DET _ Definite=Def|Gender=Masc|Number=Plur|PronType=Art 3 det _ _
2 três três NUM _ NumType=Card 3 nummod _ _
3 cortadores cortador NOUN _ Gender=Masc|Number=Plur 8 nsubj _ _
4 de de ADP _ _ 5 case _ _
5 cana cana NOUN _ Gender=Fem|Number=Sing 3 rmod _ _
6 eram ser AUX _ Mood=Ind|Number=Plur|Person=3|Tense=Imp|VerbForm=Fin 8 cop _ _
7 de de ADP _ _ 8 case _ _
8 Alagoas Alagoas PROPON _ Gender=Masc|Number=Sing 0 root _ _
9 e e CCONJ _ _ 8 cc _ _
10 estavam estar VERB _ Mood=Ind|Number=Plur|Person=3|Tense=Imp|VerbForm=Fin 8 conj _ _
11-12 na - - - - - - - - - -
11 em em ADP _ _ 13 case _ _
12 a o DET _ Definite=Def|Gender=Fem|Number=Sing|PronType=Art 13 det _ _
13 cidade cidade NOUN _ Gender=Fem|Number=Sing 10 iobj _ _
14 havia haver VERB _ _ 10 advcl _ _
15 15 15 NUM _ NumType=Card 16 nummod _ _
16 dias dia NOUN _ Gender=Masc|Number=Plur 14 obj _ SpaceAfter=No
17 . . PUNCT _ _ 8 punct _ _
```

Figura 8. Interface de edição de sentenças no Interrogatório

A partir de uma página com resultados de busca, também, é possível aplicar regras de transformação escritas em Python, conforme descrito na documentação<sup>2</sup>.

Voltando à etapa primeira da ET, comparamos os resultados das versões e, então, começamos novos experimentos, de tal modo que as duas etapas da ET se retroalimentam.

## V. CONCLUSÃO

Apresentamos uma Estação de Trabalho para trabalhar com corpora anotados composta por duas frentes interligadas: (1) a busca, revisão e edição de frases; e (2) a avaliação da qualidade da anotação do corpus.

Em termos gerais, a ideia é que por meio da visualização de divergências entre golden e modelo (ou entre diferentes modelos) seja possível depreender padrões que possam nos ajudar a buscar mais consistência. Além disso, por meio do Interrogatório, temos à disposição um sistema de busca e correção de corpora motivado pelo tipo de interrogações que linguistas podem vir a fazer ao conjunto de sentenças, de maneira tal que ajude pesquisas linguísticas com base em corpus, por um lado, e torne mais eficiente o processo de correção, manual ou automática, por outro.

## AGRADECIMENTOS

Elvis de Souza é bolsista de Iniciação Científica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) no projeto *Construção de datasets para o PLN de língua portuguesa*. Número do processo da bolsa: 128693/2019-3.

## REFERÊNCIAS

- [1] C. D. Manning, “Part-of-speech tagging from 97% to 100%: is it time for some linguistics?” in *International conference on intelligent text processing and computational linguistics*. Springer, 2011, pp. 171–189.
- [2] A. Rademaker, F. Chalub, L. Real, C. Freitas, E. Bick, and V. de Paiva, “Universal dependencies for portuguese,” in *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017)*, 2017, pp. 197–206.

<sup>2</sup>Disponível em <https://github.com/alvelvis/Interrogat-rio/wiki>

- [3] R. McDonald, J. Nivre, Y. Quirmbach-Brundage, Y. Goldberg, D. Das, K. Ganchev, K. Hall, S. Petrov, H. Zhang, T. Oscar *et al.*, “Universal dependency annotation for multilingual parsing,” in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2013, pp. 92–97.
- [4] M. Straka, J. Hajič, and J. Straková, “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 2016)*, 2016, pp. 4290–4297.
- [5] D. Santos, “Podemos contar com as contas?” *quot*; In *Sandra Maria Aluísio; Stella E O Tagnin (ed) New Language Technologies and Linguistic Research: A Two-Way Road Cambridge Scholars Publ 2014; 2014*, 2014.
- [6] F. M. Tyers, M. Sheyanova, and J. N. Washington, “Ud annotatrix: an annotation tool for universal dependencies,” 2017.
- [7] M. Janssen, “Dependency graphs and teitok: Exploiting dependency parsing,” in *International Conference on Computational Processing of the Portuguese Language*. Springer, 2018, pp. 470–478.
- [8] C. Freitas, E. de Souza, and L. Rocha, “Quantificando (e qualificando) o sujeito oculto em português,” in *VI Jornada de Descrição do Português, STIL 2019*, 2019.
- [9] J. Nivre, M. Abrams, Ž. Agić, L. Ahrenberg, G. Aleksandravičiūtė, L. Antonsen, K. Aplonova, M. J. Aranzabe, G. Arutie, M. Asahara, L. Ateyah, M. Attia, A. Atutxa, L. Augustinus, E. Badmaeva, M. Ballesteros, E. Banerjee, S. Bank, V. Barbu Mititelu, V. Basmov, J. Bauer, S. Bellato, K. Bengoetxea, Y. Berzak, I. A. Bhat, R. A. Bhat, E. Biagetti, E. Bick, A. Bielinskienė, R. Blokland, V. Bobicev, L. Boizou, E. Borges Völker, C. Börstell, C. Bosco, G. Bouma, S. Bowman, A. Boyd, K. Brokaitė, A. Burchardt, M. Candito, B. Caron, G. Caron, G. Cebiroğlu Eryiğit, F. M. Cecchini, G. G. A. Celano, S. Cėplő, S. Cetin, F. Chalub, J. Choi, Y. Cho, J. Chun, S. Cinková, A. Collomb, Ç. Çöltekin, M. Connor, M. Courtin, E. Davidson, M.-C. de Marneffe, V. de Paiva, A. Diaz de Ilarraza, C. Dickerson, B. Dione, P. Dirix, K. Dobrovoljc, T. Dozat, K. Droganova, P. Dwivedi, H. Eckhoff, M. Eli, A. Elkahky, B. Ephrem, T. Erjavec, A. Etienne, R. Farkas, H. Fernandez Alcalde, J. Foster, C. Freitas, K. Fujita, K. Gajdošová, D. Galbraith, M. Garcia, M. Gärdenfors, S. Garza, K. Gerdes, F. Ginter, I. Goenaga, K. Gojenola, M. Gökırmak, Y. Goldberg, X. Gómez Guinovart, B. González Saavedra, M. Grioni, N. Grūzītis, B. Guillaume, C. Guillot-Barbance, N. Habash, J. Hajič, J. Hajič jr., L. Hà Mý, N.-R. Han, K. Harris, D. Haug, J. Heinecke, F. Hennig, B. Hladká, J. Hlaváčová, F. Hociung, P. Hohle, J. Hwang, T. Ikeda, R. Ion, E. Irimia, O. Ishola, T. Jelínek, A. Johannsen, F. Jørgensen, H. Kaşıkara, A. Kaasen, S. Kahane, H. Kanayama, J. Kanerva, B. Katz, T. Kayadelen, J. Kenney, V. Kettnerová, J. Kirchner, A. Köhn, K. Kopacewicz, N. Kotsyba, J. Kovalevskaitė, S. Krek, S. Kwak, V. Laippala, L. Lambertino, L. Lam, T. Lando, S. D. Larasati, A. Lavrentiev, J. Lee, P. Lê H’ông, A. Lenci, S. Lertpradit, H. Leung, C. Y. Li, J. Li, K. Li, K. Lim, Y. Li, N. Ljubešić, O. Loginova, O. Lyashevskaya, T. Lynn, V. Macketanz, A. Makazhanov, M. Mandl, C. Manning, R. Manurung, C. Mărănduc, D. Mareček, K. Marheinecke, H. Martínez Alonso, A. Martins, J. Mašek, Y. Matsumoto, R. McDonald, S. McGuinness, G. Mendonça, N. Miekka, M. Misirpashayeva, A. Missilä, C. Mititelu, Y. Miyao, S. Montemagni, A. More, L. Moreno Romero, K. S. Mori, T. Morioka, S. Mori, S. Moro, B. Mortensen, B. Moskalevskiy, K. Muischnek, Y. Murawaki, K. Müürisepp, P. Nainwani, J. I. Navarro Horniácek, A. Nedoluzhko, G. Nešpore-Bėrzkalne, L. Nguy ễn Thj, H. Nguy ễn Thj Minh, Y. Nikaido, V. Nikolaev, R. Nitisaroj, H. Nurmi, S. Ojala, A. Olúòkun, M. Omura, P. Osenova, R. Östling, L. Øvrelid, N. Partanen, E. Pascual, M. Passarotti, A. Patejuk, G. Paulino-Passos, A. Peljak-Lapińska, S. Peng, C.-A. Perez, G. Perrier, D. Petrova, S. Petrov, J. Piitulainen, T. A. Pirinen, E. Pitler, B. Plank, T. Poibeau, M. Popel, L. Pretkalniņa, S. Prėvost, P. Prokopidis, A. Przepiórkowski, T. Puolakainen, S. Pyysalo, A. Rääbis, A. Rademaker, L. Ramasamy, T. Rama, C. Ramisch, V. Ravishankar, L. Real, S. Reddy, G. Rehm, M. Rießler, E. Rimkutė, L. Rinaldi, L. Rituma, L. Rocha, M. Romanenko, R. Rosa, D. Rovati, V. Roşca, O. Rudina, J. Rueter, S. Sadde, B. Sagot, S. Saleh, A. Salomoni, T. Samardžić, S. Samson, M. Sanguinetti, D. Särg, B. Saulite, Y. Sawanakunanon, N. Schneider, S. Schuster, D. Seddah, W. Seeker, M. Seraji, M. Shen, A. Shimada, H. Shirasu, M. Shohibussirri, D. Sichinava, N. Silveira, M. Simi, R. Simionescu, K. Simkó, M. Šimková, K. Simov, A. Smith, I. Soares-Bastos, C. Spadine, A. Stella, M. Straka, J. Strnadová, A. Suhr, U. Sulubacak, S. Suzuki, Z. Szántó, D. Taji, Y. Takahashi, F. Tamburini, T. Tanaka, I. Tellier, G. Thomas, L. Torga, T. Trosterud, A. Trukhina, R. Tsarfaty, F. Tyers, S. Uematsu, Z. Urešová, L. Uria, H. Uszkoreit, S. Vajjala, D. van Niekerk, G. van Noord, V. Varga, E. Villemonte de la Clergerie, V. Vincze, L. Wallin, A. Walsh, J. X. Wang, J. N. Washington, M. Wendt, S. Williams, M. Wirén, C. Wittern, T. Woldemariam, T.-s. Wong, A. Wróblewska, M. Yako, N. Yamazaki, C. Yan, K. Yasuoka, M. M. Yavrumyan, Z. Yu, Z. Žabokrtský, A. Zeldes, D. Zeman, M. Zhang, and H. Zhu, “Universal dependencies 2.4,” 2019, LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. [Online]. Available: <http://hdl.handle.net/11234/1-2988>
- [10] L. Rocha, I. Soares-Bastos, C. Freitas, and A. Rademaker, “Scavenger hunt: what do we find when look for confusions,” in *International Conference on the Computational Processing of Portuguese, PROPOR 2018*, 2018.
- [11] D. Zeman, J. Hajič, M. Popel, M. Potthast, M. Straka, F. Ginter, J. Nivre, and S. Petrov, “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*, 2018, pp. 1–21.
- [12] C. Freitas, L. F. Trugo, F. Chalub, G. Paulino-Passos, and A. Rademaker, “Tagsets and datasets: Some experiments based on portuguese language,” in *International Conference on Computational Processing of the Portuguese Language*. Springer, 2018, pp. 459–469.