Clause Linkage in Brazilian Portuguese Narratives: Evidence from Enhanced Dependency Annotation

Letícia Guedes Guimarães[1], Adriana Pagano[1], André V. Lopes Coneglian[1], Magali Duran[2]

[1] Universidade Federal de Minas Gerais
[2]Universidade de São Paulo
{letigg, apagano, coneglian}@ufmg.br
magali.duran@gmail.com

Abstract. *This paper explores the potential of Enhanced and Extended Enhanced Universal Dependencies for narrative analysis in Brazilian Portuguese, focusing on how enhanced dependency relations contribute to clause linkage. The study examines narrative sentences annotated in CoNLL-U format and compares basic and enhanced dependency layers. The results show that enhanced dependencies are particularly useful for recovering implicit subject relations in Portuguese and for enriching clause relations through information from prepositions and conjunctions. These enhancements enable a more explicit representation of participant continuity and event relations in narratives, supporting the transition from syntactic dependency analysis to richer semantic annotation.*

1. Introduction

Clause linkage is a central phenomenon for the linguistic description of discourse, since it concerns the ways in which clauses are combined, coordinated, embedded, or otherwise connected within larger sentence units [Lehmann 1988, Matthiessen and Thompson 1988]. From a systemic-functional perspective, clause complexes are one of the main grammatical resources through which logico-semantic relations between clauses are organized, including elaboration, extension, enhancement, projection, and different degrees of interdependency [Halliday and Matthiessen 2014]. In narrative texts, such as fables, these resources are especially relevant because they contribute to the sequencing of events, the backgrounding and elaboration of circumstances, the attribution of speech and the tracking of participants across clauses.

From a computational perspective, clause linkage also raises important challenges for syntactic annotation. Basic dependency representations usually encode the local head-dependent relations that hold between words, but they do not always make explicit the full range of syntactic and semantic dependencies involved in complex clause combinations. This is particularly relevant in cases such as coordinated predicates with shared subjects, control constructions, relative clauses, and constructions involving argument sharing across clauses. In such cases, some dependencies are recoverable by human interpretation but remain implicit in the basic syntactic annotation.

Enhanced Universal Dependencies (EUD) [Schuster and Manning 2016] provide a useful framework for addressing these phenomena. By allowing additional dependency relations to be represented beyond the basic tree structure, enhanced dependencies can

make implicit syntactic links more explicit. For example, they can represent shared arguments in coordination, propagated subjects in control constructions, and additional relations involving relative clauses. These enriched representations are therefore relevant not only for linguistic analysis, but also for downstream natural language processing tasks that depend on more explicit predicate-argument structures.

This paper explores the potential of EUD for informing narrative analysis in Brazilian Portuguese. The study examines how enhanced dependency annotation can make explicit relations relevant to narrative interpretation, with a focus on clause linkage and related phenomena, such as participant continuity, reported speech, relative-clause coreference, and the semantic specification of coordinating and subordinating relations. The analysis is intended as an intermediate step in a broader annotation pipeline. By examining which discourse-relevant relations can be recovered through EUD, we identify correspondences between enhanced dependency structures and phenomena represented more explicitly in subsequent semantic annotation, particularly in UMR [Van Gysel et al. 2021], such as coreference, temporal relations, modal dependencies, and event structure. In this sense, the paper situates enhanced dependency annotation as an intermediate layer between syntactic dependency analysis and richer semantic representations of narrative discourse.

2. Related work

In contrast to constituency-based representations, which organize words into hierarchically nested phrases, basic dependency trees, such as those used in the Universal Dependencies framework [Nivre et al. 2020], represent morphosyntactic structure through labeled binary relations between individual words. For example, the relation *nsubj(arrancou, cigarra)* encodes a direct dependency between the verb *arrancou* and its nominal subject, *cigarra*. A constituency analysis would represent the same grammatical relation indirectly, through the position of the noun phrase containing *cigarra* within the phrase structure. Dependency representations are widely used in natural language processing, particularly in multilingual research, because they provide a common framework for describing grammatical relations across languages with different word-order patterns. However, basic dependency trees have long been recognized as limited in their ability to represent semantically relevant relations between content words when these relations are not directly expressed in surface syntactic structure [Schuster and Manning 2016].

This limitation was progressively addressed through richer annotation schemes: first with the collapsed relations of the Stanford Dependencies framework [de Marneffe et al. 2006, de Marneffe and Manning 2008], which expanded the representation of prepositional and conjunctive relations, and later with Enhanced Universal Dependencies (EUD), proposed by [Schuster and Manning 2016], which extended this approach to a broader range of grammatical phenomena within the UD framework, such as shared arguments, elided subjects, and dependencies in relative clauses.

According to the UD guidelines [Nivre et al. 2020, de Marneffe et al. 2021], EUD complement basic dependencies by making implicit information explicit and by enriching dependency labels to better disambiguate arguments and modifiers. This can be implemented either through the addition of new relations or through the modification of existing ones. These two processes differ not only in what they represent linguistically, but also

in how amenable they are to automatic annotation: additive relations tend to be more easily converted automatically, while modification-based ones generally demand human revision [Pagano et al. 2023].

EUD was first introduced in Brazilian computational linguistics by [Pagano et al. 2023], providing examples in Portuguese for each type of enhanced dependency and discussing strategies for implement them. A detailed study was then elaborated by [Duran 2024] and [Duran et al. 2024], defining several rules for enhanced dependencies in Portuguese, including the propagation of subjects elided in pro-drop languages. These rules were used to customize a converter from UD to EUD for Portuguese [Souza et al. 2024]. This converter was improved one year later with rules developed for pro-drop languages.These rules, referred to as Extended Enhanced Universal Dependencies or EEUD, extended the original framework, introducing additional relations to represent implicit arguments, particularly omitted subjects in languages where verbal morphology allows their recovery. The converter was used to produce a layer with both EUD and EEUD in the corpus Porttinari-base [Duran et al. 2025], the first corpus in Portuguese annotated with enhanced dependencies. Together, these efforts laid the groundwork for the annotation reported in this paper.

EUD annotation is especially relevant to the analysis of narrative texts because enhanced dependencies capture relations that are more explicit and semantically informative than those represented in basic dependency structures, thereby offering useful insights into narrative organization and interpretation. Despite this potential, the application of EUD to narrative analysis remains largely unexplored, a gap that the present study seeks to address. Fables are particularly suitable for this purpose because they typically combine two linguistically distinct components—a narrative sequence and a moral commentary—creating opportunities for future comparisons of the relations instantiated in each. From a computational perspective, Brazilian Portuguese narrative datasets annotated according to internationally established frameworks such as UD and EUD also remain scarce, since existing UD resources for Brazilian Portuguese predominantly represent other registers. By providing EUD annotation for a corpus of fables, this study contributes to filling this resource gap and broadens the possibilities for computational and linguistic analyses of narrative discourse.

3. Methodology

3.1. Corpus compilation and annotation

Our corpus is part of a larger multilingual corpus of Aesop's fables translated from Greek into Brazilian Portuguese, French, English, Spanish, Russian, Chinese and Italian. The Brazilian Portuguese translations were published in *Esopo: fábulas completas* [Dezotti et al. 2013] and are available in print only. The text was manually typed and then saved in two separate plain-text files, one compiling the narrative section of the fables and another one compiling the moral or epimythium section [Dezotti 1988].

The full corpus is outlined in Table 1. In this paper, we report on the narrative subset, which accounts for 85% of the total corpus.

Table 1. Corpus composition and basic statistics.

Corpus	Sentences	Tokens	Unique words	Lemmas	Avg. sentence length
Narratives	355	8,836	2,163	1,679	24.89
Epimythia	63	1,486	512	444	23.59
Full corpus	418	10,322	2,440	1,857	24.69

3.1.1. Basic dependencies annotation

The corpus was automatically annotated for basic dependencies using the online parser UD-Pipe [Straka et al. 2016] trained on the PetroGold corpus [Souza et al. 2021]. The resulting CoNLL-U files were uploaded to Arborator-Grew [Guibon et al. 2020] for manual revision. Tokenization errors involving Brazilian Portuguese contractions that failed to be properly detokenized (e.g., 'comigo' instead of 'com mim', 'dele' instead of 'de ele') due to adjacent punctuation marks were corrected before the revision of the basic UD annotation, which was conducted in two rounds: an initial pass by one language specialist, followed by adjudication by two qualified reviewers.

Revision followed the Brazilian Portuguese UD annotation guidelines [Duran 2021, Duran 2022]. During this stage, the automatic parses were corrected and supplemented with dependency subrelations. For instance, reported-speech clauses were manually annotated with the `ccomp:speech` subrelation.

3.2. Enhanced dependencies annotation

Enhanced annotation was carried out using a converter customized for Portuguese by [Souza et al. 2024], which includes, in its last version, all EEUD proposed by [Duran et al. 2025]. A description of each category and examples can be found in [Pagano et al. 2023] and [Duran et al. 2025].

The repository was cloned locally and the scripts were run in Visual Studio Code over the CoNLL-U files used in this study. The output consists of CoNLL-U files in which the basic dependency layer is preserved and the enhanced dependency graph is added in the `DEPS` column. Manual revision was carried out on all sentences using a customized UD annotation and visualization tool [Souza 2026]. The quality and limitations of the automatic conversion exceed the scope of this paper and will be evaluated in future work.

3.3. Data extraction and analysis

Quantitative data extraction was performed using a customized Python script in Visual Studio Code. The script was designed to extract corpus-level statistics, part-of-speech frequencies, basic dependency relations, and enhanced dependency relations from the CoNLL-U files. It also compared the basic `HEAD:DEPREL` pairs with the relations encoded in the enhanced dependency column in order to identify how the enhanced layer contributes additional syntactic and semantic information.

The analysis distinguishes between enhanced relations that add new dependency links and enhanced relations that specify existing basic links. Added links include, for example, propagated subjects, propagated conjunct dependencies, relative-clause coreference links, and reattached complements. By contrast, relations such as `obl:de`, `nmod:de`, `advcl:quando`, and `conj:e` usually preserve the same head as the corresponding basic relation and add lexical or semantic information after the colon. These

were therefore treated as enhanced specifications of basic relations, rather than as newly added links.

4. Results

This section presents and discusses the results obtained from the annotated corpus. We first describe the distribution of part-of-speech tags, before examining basic and enhanced dependency relations relevant to clause linkage and narrative interpretation, with a focus on the contributions of enhanced representations.

4.1. Part-of-speech distribution

We first examine the distribution of part-of-speech tags in the corpus. As shown in Table 2, the corpus contains a high proportion of nominal and verbal tags. The most frequent lexical categories are NOUN, with 1,292 occurrences, and VERB, with 1,222 occurrences. Function words are also frequent, especially DET with 1,115 occurrences, ADP with 879 occurrences, and PRON with 762 occurrences. This distribution is consistent with the narrative character of the corpus. Nouns contribute to the introduction and tracking of participants, verbs encode events and actions, and function words contribute to the grammatical organization of clauses and clause complexes. The frequencies of CCONJ (374) and SCONJ (278) are particularly relevant for the present study, since they point to the importance of coordination and subordination in the corpus.

Table 2. UPOS frequency distribution in the corpus.

UPOS	Frequency
PUNCT	1,327
NOUN	1,292
VERB	1,222
DET	1,115
ADP	879
PRON	762
ADV	384
CCONJ	374
SCONJ	278
ADJ	253
AUX	247
PROPN	31
NUM	20
INTJ	3

The relatively low number of proper nouns is also noteworthy. The corpus contains only 31 PROPN tokens. As shown in Table 3, these correspond to a small set of mythological figures and geographical names.

Table 3. Proper nouns in the corpus.

PROPN form	Frequency
Zeus	15
Hermes	8
Prometeu	3
Pireu	3
Súnion	1
Ática	1
Total	31

The low frequency of PROPN tokens indicates that named entities play only a limited role in the corpus. This is consistent with the type of narratives analyzed here, where participants are often referred to by common nouns, such as animal names and social roles, rather than by individual proper names. For this reason, the analysis of participant reference cannot be based primarily on named entities. Instead, it requires attention to how participants are introduced and tracked through nominal expressions, pronouns, and syntactic relations across clauses.

4.2. Basic dependency relations

The basic dependency layer provides the first level of syntactic representation for the corpus. Since the corpus contains 355 sentences, the basic annotation includes 355 `root` relations. Table 4 shows the most frequent basic dependency relations in the corpus.

Table 4. Most frequent basic dependency relations.

Basic dependency relation	Frequency
punct	1,327
det	1,083
case	711
nsubj	558
obj	513
obl	424
mark	410
cc	378
advmod	361
root	355
advcl	333
conj	293
nmod	188
xcomp	163
amod	129

Excluding punctuation, the most frequent relations are `det`, `case`, `nsubj`, `obj`, `obl`, `mark`, and `cc`. These relations reflect the high frequency of nominal modification, prepositional marking, participant encoding, and clause-linking material in the corpus. The presence of frequent `advcl` and `conj` relations is particularly relevant for this study, since these relations directly contribute to the syntactic representation of subordination and coordination.

Relations associated with clause linkage and narrative interpretation in the corpus can be seen in Table 5.

Table 5. Basic dependency relations relevant to clause linkage.

Basic dependency relation	Frequency
advcl	333
conj	293
xcomp	163
ccomp	93
acl:relcl	88
ccomp:speech	64
acl	61
csubj	15
parataxis	10

The basic layer therefore already captures a substantial number of clause-linking

configurations. The high number of advcl relations indicates frequent adverbial subordination, while conj relations point to recurrent coordination between predicates or larger syntactic units. Complementation is represented by both ccomp and xcomp, and the distinction between ccomp and ccomp:speech makes it possible to separate general clausal complements from reported speech complements. Relative and adnominal clauses are represented by acl:relcl and acl, respectively.

Subject-type relations are particularly important for the analysis of participant continuity in narratives. In the basic layer, there are 588 subject-type relations, including 558 nsubj, 15 nsubj:pass, and 15 csubj relations. Object-type relations are also frequent in the basic layer. In the basic dependency layer, object-type relations are represented mainly by obj, with 513 occurrences, and, less frequently, by iobj, with 55 occurrences. Together, these account for 568 core object-type relations.

Taken together, the basic dependency results show that the corpus contains a high number of syntactic structures relevant to clause linkage. Relations such as advcl, conj, xcomp, ccomp, ccomp:speech, acl:relcl, and acl provide the basic syntactic representation of subordination, coordination, complementation, reported speech, and adnominal modification. The following subsection examines how the enhanced dependency layer adds further relations that make participant continuity and clause-linking meanings more explicit.

4.3. Enhanced dependencies

We examined enhanced relations by distinguishing newly introduced links from specifications of existing basic dependencies. The former make implicit syntactic relations explicit and support the analysis of participant tracking and event structure, whereas the latter add semantic detail relevant to clause linkage, including information contributed by prepositions and subordinating or coordinating conjunctions.

Table 6 reports the most frequent enhanced relations observed in the corpus.

Table 6. Most frequent enhanced dependency relations.

Enhanced relation	Frequency
nsubj	591
conj:e	254
nmod:de	169
obl:em	107
ref	86
obl:de	76
advcl:quando	71
obl:com	50
obl:por	50
mark	50
obl:a	48
advcl:se	33
obl:para	32
advcl:a_o	29
obl:arg:a	29
obj	28
advcl:para	24
advcl:pois	23
ccomp	22

4.3.1. Enhanced dependencies and participant tracking

The enhanced layer adds 610 subject-type relations, slightly more than the 588 subject-type relations present in the basic layer. This increase reflects the pro-drop nature of Brazilian Portuguese, elided subjects being frequently recoverable from context. By recovering implicit subject dependencies and propagating subject links across coordinated, subordinate, and non-finite clauses, EEUD makes participant–event relations more explicit. The most frequent added enhanced relation is `nsubj`, with 591 instances, indicating that subject continuity is a major contribution of the enhanced layer. These additional links are particularly relevant in clause-linkage configurations, such as coordinated predicates and control constructions, where the same participant is associated with multiple events. Table 7 compares subject-type relations in basic and enhanced dependencies.

Table 7. Comparison of subject-type relations in the basic and enhanced layers.

Subject-type relation	Basic	Enhanced
`nsubj`	558	591
`nsubj:pass`	15	13
`csubj`	15	6
Total	588	610

Figure 1 illustrates a case of enhanced subject relations, which indicates that the enhanced layer frequently adds participant–predicate links that are not explicit in the basic tree[1].

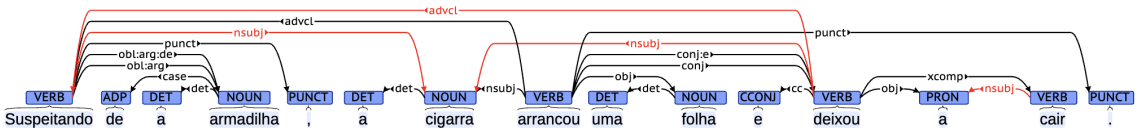

Figure 1. Basic and enhanced dependencies in a sample sentence.

In the basic layer, *a cigarra* is the subject of the finite verb *arrancou*; the gerund-dive clause headed by *Suspeitando* is attached as `advcl`; and *deixou* is coordinated with *arrancou* through `conj`. The enhanced layer makes additional relations explicit: *a cigarra* is also linked as subject to *Suspeitando* and *deixou*; the adverbial clause relation is propagated to the coordinated predicate; and the coordinating conjunction is specified in `conj:e`. In addition, the oblique argument *da armadilha* is specified as `obl:arg:de`, making the prepositional marking explicit. The example also illustrates the assignment of an external subject to an `xcomp`. In the basic layer, the clitic pronoun *a* is annotated as the object of *deixou*. In the enhanced layer, it is also linked as `nsubj` of the infinitive *cair*. This makes explicit that the entity undergoing the falling event is the leaf, resumed by the clitic *a*, rather than the matrix subject *a cigarra*. Thus, the enhanced relation `14:nsubj` avoids an incorrect interpretation in which the subject of the matrix clause would be propagated to the `xcomp`.

[1]In the figure, red arcs highlight enhanced dependencies that add new links to the graph. Black arcs include the basic dependency structure and enhanced relations that duplicate or specify existing basic links.

Unlike subject relations, enhanced dependencies add relatively few object relations. The basic layer contains 568 object-type relations, while the enhanced layer adds only 28 obj relations and no additional iobj relations.

Table 8. Comparison of object-type relations in the basic and enhanced layers.

Object-type relation	Basic	Enhanced
obj	513	28
iobj	55	0
Total	568	28

This contrast suggests that, in this corpus, enhanced dependencies contribute more strongly to the recovery of subject continuity than to the recovery of object relations. Added subject links help identify which participant is associated with each predicate, especially in coordinated, subordinate, and non-finite clauses where the participant may be shared but not repeated.

4.3.2. Enhanced dependencies and clause linkage

Enhanced relations also provide finer-grained information about clause linkage. Coordination is particularly prominent. Table 9 shows coordination relations specified in the enhanced layer. These specifications make the semantic or discourse value of coordination more explicit, distinguishing additive, contrastive, comparative, and alternative links. The most frequent coordination relation is additive, linking events or propositions through conjunction. Its frequency may reflect the relatively linear progression of fable narratives, where successive actions and developments are commonly connected through simple additive relations. However, this distribution should be interpreted with caution, since the conjunction e can also introduce relations with additional semantic values, such as contrast or temporal succession, depending on the context.

Table 9. Enhanced coordination relations.

Added coordination relation	Frequency
conj:e	254
conj:mas	19
conj:como	2
conj:ou	1
Total	276

Adverbial clause relations are also highly informative. As Table 10 shows, the enhanced layer specified relations which allow for an analysis of the semantic relations construed by adverbial clauses, such as temporal sequencing, conditionality, purpose, cause, comparison, and concession. These relations are especially relevant for narrative interpretation because they help identify how events are temporally ordered, causally motivated, conditionally framed, or contrastively related.

Temporal relations such as advcl:quando, advcl:enquanto, advcl:assim_que, advcl:depois_que, and advcl:logo_que, accounting for most adverbial relations in the corpus, contribute to event sequencing, which is central to narrative texts. Conditional relations such as advcl:se help represent

hypothetical or contingent developments. Causal and explanatory relations such as advcl:pois and advcl:porque contribute to the interpretation of motivation and consequence.

Table 10. Selected enhanced adverbial clause relations.

Selected added adverbial clause relation	Frequency
advcl:quando	71
advcl:se	33
advcl:a_o	29
advcl:para	24
advcl:pois	23
advcl:enquanto	19
advcl	19
advcl:como	14
advcl:a	10
advcl:porque	10
advcl:que	9
advcl:assim_que	9
advcl:por	7
advcl:sem	6
advcl:a_o_passo_que	6
advcl:até_que	4
advcl:para_que	4
advcl:depois_que	3
advcl:de_o_que	3
advcl:após	3

Complement clauses also appear among enhanced relations as seen in Table 11. These relations are relevant for representing speech, thought, perception, intention, and control structures, all of which are important in narrative texts.

Table 11. Enhanced complement relations.

Added complement relation	Frequency
ccomp	22
ccomp:speech	12
xcomp	12
Total	46

5. Conclusion

This paper explored the potential of EUD, specifically EEU, for informing the analysis of clause linkage in Brazilian Portuguese narratives. Overall, the enhanced layer contributes information that is implicit in the basic dependency representation. In this corpus, its most important contribution concerns subject continuity, showing that EEUD are valuable for recovering participant–event relations in a pro-drop language such as Brazilian Portuguese. By making implicit relations more explicit and enriching existing dependencies with additional linguistic information, they provide a more detailed account of clause linkage and narrative structure.

Future work will extend this analysis toward UMR annotation [Van Gysel et al. 2021], with particular attention to coreference, temporal relations, modality, and event structure. In this broader pipeline, EEUD can serve as an intermediate layer between syntactic dependency annotation and richer semantic representations of narrative discourse. Upon completion, the corpus will be made available through a GitHub repository.

6. Acknowledgments

The authors thank the anonymous reviewers for their comments, which helped improve this paper. Letícia Guedes is currently supported by Minas Gerais State Agency for Research and Development (FAPEMIG). Adriana Pagano has grants from FAPEMIG and the National Council for Scientific and Technological Development (CNPq 404722/2024-5; 313103/2021-6). André V. Lopes Coneglian's research is funded by CNPq (153972/2025-4). Their work is supported by the National Institute of Science and Technology in Responsible Artificial Intelligence for Computational Linguistics, Information Treatment, and Dissemination (INCT-TILDIAR), funded by CNPq (grant no. 408490/2024-1). Magali Sanches Duran is a member of the Artificial Intelligence Center of the University of São Paulo, funded by the São Paulo Research Foundation (FAPESP grant no. 2019/07665-4) and IBM. She is also supported by the Ministry of Science, Technology and Innovations, with resources from Law No. 8,248 of October 23, 1991, within the scope of PPI-Softex, coordinated by Softex and published as ICT Residency 13, DOU 01245.010222/2022-44.

7. Limitations

The study has some limitations. The corpus is relatively small and restricted to narrative sentences, and the analysis is primarily descriptive and annotation-oriented. In addition, the enhanced dependencies were produced by rule-based procedures, which means that some cases require manual validation, especially where enhanced relations depend on contextual interpretation. The impact of different automatic parsers on the efficiency of manual revision and on the quality of the resulting enhanced annotations remains to be investigated.

8. Ethics statement

This study does not involve collecting data from human participants or sensitive information of any kind. The corpus consists of translated texts made available for research purposes with the translator's explicit permission and in compliance with copyright requirements. All tools used for corpus compilation and annotation, code generation, and data analysis are freely available for research use.

9. Generative AI usage

No scientific content, analyses, or interpretations were produced by AI.

References

de Marneffe, M.-C., MacCartney, B., and Manning, C. D. (2006). Generating typed dependency parses from phrase structure parses. In Calzolari, N., Choukri, K., Gangemi, A., Maegaard, B., Mariani, J., Odijk, J., and Tapias, D., editors, *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).

de Marneffe, M.-C. and Manning, C. D. (2008). The Stanford typed dependencies representation. In Bos, J., Briscoe, E., Cahill, A., Carroll, J., Clark, S., Copestake, A., Flickinger, D., van Genabith, J., Hockenmaier, J., Joshi, A., Kaplan, R., King, T. H.,

Kuebler, S., Lin, D., Lønning, J. T., Manning, C., Miyao, Y., Nivre, J., Oepen, S., Sagae, K., Xue, N., and Zhang, Y., editors, *Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation*, pages 1–8, Manchester, UK. Coling 2008 Organizing Committee.

de Marneffe, M.-C., Manning, C. D., Nivre, J., and Zeman, D. (2021). Universal dependencies. *Computational Linguistics*, 47(2):255–308.

Dezotti, M. C. C. (1988). A fábula esópica anônima: Uma contribuição ao estudo dos "atos de fábula". Master's thesis, Universidade Estadual Paulista, Araraquara, Brazil.

Dezotti, M. C. C., Berliner, E., and Duarte, A. (2013). *Esopo: fábulas completas*. Cosac Naify, São Paulo, Brazil.

Duran, M. (2021). Manual de anotação de pos tags: orientações para anotação de etiquetas morfossintáticas em língua portuguesa, segundo as diretrizes da abordagem universal dependencies. Technical Report Relatório Técnico No. 434, Universidade de São Paulo, ICMC. Accessed: 2026-04-10.

Duran, M. (2022). Manual de anotação de relações de dependência: versão revisada e estendida. Technical Report Relatório Técnico No. 440, Universidade de São Paulo, ICMC. Accessed: 2026-04-30.

Duran, M., Souza, E., das Graças Volpe Nunes, M., Pagano, A. S., and Pardo, T. A. S. (2025). Extending the enhanced universal dependencies: Addressing subjects in pro-drop languages. In *Proceedings of the Eighth Workshop on Universal Dependencies (UDW, SyntaxFest 2025)*, pages 143–152, Ljubljana, Slovenia. Association for Computational Linguistics.

Duran, M. S. (2024). Anotação de enhanced dependencies. Available at: `https://repositorio.usp.br/directbitstream/ebb807a6-8543-4616-adb9-2f1715236733/3209188.pdf`. Accessed: 31 May 2026.

Duran, M. S., Nunes, M. d. G. V., and Pardo, T. A. S. (2024). Inferências baseadas em sintaxe: a anotação de sujeitos implícitos. In Claro, D. B. and Pagano, A., editors, *Proceedings of the 15th Brazilian Symposium in Information and Human Language Technology*, pages 356–364, Belém do Pará, Brazil. Association for Computational Linguistics.

Guibon, G. et al. (2020). When collaborative treebank curation meets graph grammars. In *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC)*, pages 5291–5300.

Halliday, M. A. K. and Matthiessen, C. M. I. M. (2014). *Halliday's Introduction to Functional Grammar*. Routledge, London, 4 edition.

Lehmann, C. (1988). Towards a typology of clause linkage. *Studies in Language*, 12(2):181–225.

Matthiessen, C. and Thompson, S. A. (1988). The structure of discourse and "subordination". In Haiman, J. and Thompson, S. A., editors, *Clause Combining in Grammar and Discourse*, pages 275–329. John Benjamins, Amsterdam.

Nivre, J., de Marneffe, M.-C., Ginter, F., Hajič, J., Manning, C. D., Pyysalo, S., Schuster, S., Tyers, F., and Zeman, D. (2020). Universal Dependencies v2: An evergrowing multilingual treebank collection. In Calzolari, N., Béchet, F., Blache, P., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4034–4043, Marseille, France. European Language Resources Association.

Pagano, A. S., Duran, M. S., and Pardo, T. A. S. (2023). Enhanced dependencies para o português brasileiro. In Pardo, T. A. S., Duran, M. S., and Lopes, L., editors, *Proceedings of the 2nd Edition of the Universal Dependencies Brazilian Festival*, pages 461–470, Belo Horizonte, Brazil. Association for Computational Linguistics.

Schuster, S. and Manning, C. D. (2016). Enhanced english universal dependencies: An improved representation for natural language understanding tasks. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*.

Souza, E. (2026). Anotador de dependências sintáticas enriquecidas (eud) para o português. Web interface and source code. Source code available at `https://github.com/alvelvis/eud-portugues`. Accessed: 31 May 2026.

Souza, E., Duran, M., Nunes, M. d. G. V., Sampaio, G., Belasco, G., and Pardo, T. (2024). Automatic annotation of enhanced Universal Dependencies for Brazilian Portuguese. In Claro, D. B. and Pagano, A., editors, *Proceedings of the 15th Brazilian Symposium in Information and Human Language Technology*, pages 22–31, Belém do Pará, Brazil. Association for Computational Linguistics.

Souza, E., Silveira, A., Cavalcanti, T., Castro, M., and Freitas, C. (2021). PetroGold – corpus padrão ouro para o dominio do petroleo. In Ruiz, E. E. S. and Torrent, T. T., editors, *Proceedings of the 13th Brazilian Symposium in Information and Human Language Technology*, pages 29–38, Porto Alegre, Brazil. Association for Computational Linguistics.

Straka, M., Hajič, 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)*, pages 4290–4297.

Van Gysel, J. E., Vigus, M., Chun, J., Lai, K., Moeller, S., Yao, J., O'Gorman, T., Cowell, A., Croft, W., Huang, C.-R., et al. (2021). Designing a uniform meaning representation for natural language processing. *KI-Künstliche Intelligenz*, 35(3):343–360.