

Beyond the Data: The Impact of Annotation Inconsistencies in UD Treebanks on Typological Universals and Complexity Assessment

Antoni Brosa-Rodríguez and M. Dolores Jiménez-López

Universitat Rovira i Virgili

Avda. Catalunya 35, 43002, Tarragona (Spain)

{antoni.brosa, mariadolores.jimenez}@urv.cat

Abstract

This study explores the impact of annotation inconsistencies in Universal Dependencies (UD) treebanks on typological research in computational linguistics. UD provides a standardized framework for cross-linguistic annotation, facilitating large-scale empirical studies on linguistic diversity and universals. However, despite rigorous guidelines, annotation inconsistencies persist across treebanks. The objective of this paper is to assess how these inconsistencies affect typological universals, linguistic descriptions, and complexity metrics. We analyze systematic annotation errors in multiple UD treebanks, focusing on morphological features. Case studies on Spanish and Dutch demonstrate how differing annotation decisions within the same language create contradictory typological profiles. We classify the errors into two main categories: overgeneration errors (features incorrectly annotated, since do not actually exist in a language) and data omission errors (inconsistent or incomplete annotation of features that do exist). Our results show that these inconsistencies significantly distort typological analyses, leading to false generalizations and miscalculations of linguistic complexity. We propose methodological safeguards for typological research using UD data. Our findings highlight the need for methodological improvements to ensure more reliable cross-linguistic generalizations in computational typology.

1 Introduction

Multilingual corpora with consistent annotation schemes have become invaluable resources for typological research in computational linguistics (O’Horan et al., 2016; Ponti et al., 2019). Among these, Universal Dependencies (UD) (Nivre et al., 2023) stands out as one of the most comprehensive collections of consistently annotated treebanks across diverse languages. The standardized annotation framework of UD has enabled researchers to conduct large-scale cross-linguistic

comparisons and formulate typological universals based on empirical data rather than theoretical assumptions (Brosa-Rodríguez and Jiménez-López, 2023; Gerdes et al., 2019). This development has significantly advanced our understanding of linguistic diversity and universals.

However, the promise of consistent cross-linguistic annotation faces substantial challenges in practice. Despite rigorous guidelines and quality control measures, inconsistencies and errors in annotation persist across different treebanks, even within the same language. These inconsistencies, while perhaps minor when considering individual treebanks in isolation, can have significant implications when aggregated for typological studies, potentially leading to incorrect characterizations of languages and flawed formulations of linguistic universals. Our research identified several systematic annotation errors across multiple UD treebanks that directly impact typological characterizations based on UD data.

This paper examines how these annotation inconsistencies affect the formulation of typological universals, description of languages or information regarding linguistic complexity (Brosa-Rodríguez et al., 2024) with a particular focus on morphological features such as gender, number, and verbal mode/tense. We establish a correlation between the concepts of linguistic complexity and linguistic universals. We understand the concept of complexity in terms of the difficulty of learning one language from another (second language acquisition); we interpret universals as structures/categories present in all languages. From this standpoint, we establish an inversely proportional relationship between the two concepts: The greater the degree of shared characteristics between two languages, the less challenging it will be to learn one from the other. In essence, the higher the universality of a language, the lower its complexity level when learned as a second language. Given the interrelationship between typological

logical universals and complexity, and considering that typological universals are calculated from treebanks, eliminating inconsistencies in UD treebanks is crucial for accurately calculating linguistic complexity, as these inconsistencies can distort typological profiles, affecting the relationship between universals and complexity, particularly when measuring the ease of learning languages based on their shared features.

We exemplify our research with cases from Spanish and Dutch treebanks to demonstrate how annotation decisions in one treebank can differ substantially from another for the same language, creating contradictory typological profiles. Furthermore, we explore how conversion processes from legacy annotation schemes to UD can introduce systematic biases if not carefully supervised.

The research questions guiding this investigation are: (1) How do annotation inconsistencies in UD treebanks affect typological characterizations of languages? (2) What methodological safeguards can researchers implement to account for these biases when conducting typological studies using UD data?

By addressing these questions, we aim to strengthen the foundation of computational typology while acknowledging the inherent challenges in creating truly consistent cross-linguistic annotation schemes. Rather than diminishing the value of resources like UD, our goal is to enhance their utility by promoting awareness of potential biases and suggesting practical approaches to mitigate their effects on typological research.

2 Theoretical Framework

UD (Nivre et al., 2023) has established itself as a standard framework for syntactic annotation across languages (Marneffe et al., 2013; Zeman, 2008; Petrov et al., 2012), with its primary goal being to capture linguistic universals while accommodating language-specific phenomena. The standardized annotation schema enables cross-linguistic comparison and facilitates typological research on an unprecedented scale in corpora (Haspelmath, 2010). However, the application of a universal schema to typologically diverse languages inevitably creates tension between universal applicability and language-specific accuracy.

The challenges of cross-linguistic annotation have been documented in the literature (Kahane et al., 2021; Gerdes et al., 2018, 2022; Yan and

Liu, 2022; Osborne and Gerdes, 2019). These challenges include the difficulty of establishing truly universal categories, the problem of forcing language-specific phenomena into universal frameworks, and the lack of correspondence between UD annotation guidelines and classical linguistic claims or theories. While UD has made significant progress in addressing these issues through detailed guidelines and collaborative development, other authors have proposed alternative proposals in order to enhance these detected problems (Gerdes et al., 2022).

In particular, morphological features present unique challenges for cross-linguistic annotation. Features such as gender or number vary significantly across languages, both in terms of their existence and their manifestation. UD addresses this variability through a feature inventory that distinguishes universal from language-specific features. Even if the annotation scheme is adaptable enough, the problems still arise due to annotators (or annotating) action. In this case we do not find as much error analysis as in the case of the revision of the annotation scheme from a theoretical perspective (Arista, 2022; Oh et al., 2020). The only frequent review is a specific review of problems inherent to certain languages, without being general or extendable.

3 Typology of Annotation Errors

Based on our analysis of UD treebanks, we propose a typology of annotation errors that affects typological generalizations. These errors can be classified into two broad categories:

1. **Overgeneration errors:** These occur when features that do not exist in a language (or structure) are incorrectly annotated. We have identified two primary sources of overgeneration:
 - *Automatic conversion artifacts:* When non UD-native treebanks are converted from legacy annotation schemes to UD one, features may be erroneously carried over or generated based on superficial similarities with other languages or parts of speech.
 - *Overgeneralization of specific contexts:* Annotators may apply features appropriate in one context and these are (probably automatically) propagated to con-

texts where they are linguistically unmotivated.

2. **Data omission errors:** These occur when features that do exist in a language (or structure) are inconsistently or incompletely annotated. Sources include:

- *Annotation fatigue:* Manual annotation of features that are not morphologically marked may be inconsistent due to human error or oversight.
- *Implicit vs. explicit marking:* Disagreement among annotators regarding whether features should be annotated only when explicitly marked or also when implicitly present through other patterns.

3.1 Implications for Typological Universals

These annotation inconsistencies have direct implications for the identification and validation of typological universals and, even, linguistic (structural) complexity. In the context of UD-based typology, we will use as example a revisit [Greenberg \(1963\)](#) universals.

Specifically, we examine how annotation errors may affect the validity of linguistic type knowledge based on three universals we select for exemplifying:

- **Universal 30:** If the verb has categories of person-number or if it has categories of gender, it always has tense-mode categories.
- **Universal 31:** If either the subject or object noun agrees with the verb in gender, then the adjective always agrees with the noun in gender.
- **Universal 42:** All languages have pronominal categories involving at least three persons and two numbers.

We consider that both types of errors—overgeneration and data omission—can artificially strengthen or weaken the evidence for these universals. Overgeneration errors may create false examples supporting a universal, while data omission may obscure examples that would contradict it. The combined effect can significantly distort our understanding of cross-linguistic patterns or how we can characterize the studied languages.

In the following sections, we present empirical evidence of these error types from Spanish and Dutch treebanks and demonstrate their impact on the universals listed above.

4 Methodology

Our investigation of annotation inconsistencies in UD treebanks follows a systematic methodology designed to identify, categorize, and assess the impact of annotation errors on typological generalizations. This section describes our data selection, query methods, and analytical approach. We analyzed all available treebanks from UD (version 2.15) querying information contained in [Greenberg \(1963\)](#) universals.

4.1 Query Methodology

To systematically identify annotation inconsistencies, we utilized Grew-Match ([Guillaume, 2021](#)), a query tool specifically designed for UD treebanks. Grew-Match allows for precise pattern matching across morphosyntactic features and dependencies, making it ideal for cross-treebank comparison.

We formulated targeted queries to detect potential annotation errors related to the universals under investigation. For example, some of the formalisations used in connection with Greenberg universals that have allowed us to uncover errors are:

```
pattern {A[upos=ADJ, !Gender]}
pattern {A[upos=VERB, Gender=Masc]}
pattern {A[upos=PRON, Person=1, !Number]}
```

The first pattern identifies adjectives lacking gender feature, which may indicate data omission errors relevant to Universal 31. The second pattern identifies verbs with masculine gender in agreement with masculine nominal subjects, which may represent overgeneration errors affecting Universal 31. The third pattern locates first-person pronouns without number annotation, potentially impacting Universal 42.

For each query, we:

1. Executed the pattern across all selected treebanks.
2. Counted matches to quantify the prevalence of each pattern.
3. Extracted contextual examples for qualitative analysis.
4. Compared results across different treebanks of the same language.

4.2 Analytical Framework

Our analysis proceeded in two stages:

Stage 1: Identification of Candidate Errors

We first identified candidate errors by looking for patterns that: (1) appeared inconsistently across different treebanks of the same language; (2) contradicted known typological features of the language; (3) showed signs of automatic conversion artifacts, such as systematic misapplication of features.

Stage 2: Impact Assessment

Then, we assessed the impact of confirmed errors on typological universals by: (1) quantifying how the error affects statistical generalizations; (2) determining whether the error would lead to misclassification of a language with respect to a universal; (3) estimating the potential cascade effect on related typological claims.

4.3 Reproducibility

To ensure reproducibility of our findings, we provide all Grew-Match queries used in our analysis. They are available in [GitHub](#).

5 Case Studies

This section presents detailed analyses of specific annotation inconsistencies identified in our investigation and their implications for typological research. We focus on three representative cases that illustrate both overgeneration and data omission errors across different morphological features.

5.1 Gender in Spanish Verbs: An Overgeneration Error

Our analysis revealed a systematic overgeneration error in Spanish treebanks, where perfect participles in compound verb forms are incorrectly annotated with gender features. For example, in the AnCora treebank, sentences like "*Microsoft ha cometido repetidamente graves violaciones legales*" ('Microsoft has repeatedly committed serious legal violations'), show the participle *cometido* annotated with Gender=Masc, as can be seen in figure 1.

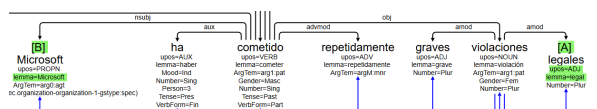


Figure 1: Annotation of "ha cometido" in Spanish-AnCora treebank

This pattern of marking a gender is not limited to AnCora but appears across multiple Spanish tree-

banks, as evidenced by examples such as "*He dicho con una botella*" ('I have said with a bottle') from COSER, "*Han muerto todos*" ('They have all died') from GSD, and "*Hemos pedido a otros países*" ('We have asked other countries') from PUD.

The error appears to stem from an overgeneralization of specific contexts where gender marking on participles is linguistically motivated (such as in passive constructions like "*fue cometida*" - 'it was committed'). In compound tenses with *haber*, however, the participle functions purely as a verbal element without nominal or adjectival properties, making gender marking inappropriate in these contexts, as Spanish does not express gender in verbs.

Impact on Typological Universals This inconsistency directly affects Universal 31, which concerns patterns of gender agreement. When analyzing Spanish based on these treebanks, we would incorrectly conclude that Spanish exhibits gender marking on verbs in all perfect constructions, potentially classifying it with languages that genuinely mark gender on verbs. Thus, we could also wrongly conclude that there is gender agreement between subjects and verbs. This misclassification could skew cross-linguistic patterns and lead to incorrect typological generalizations about the distribution of gender features across different parts of speech.

5.2 Gender in Roman Languages Adjectives: Implicit vs. Explicit Marking

We identified a systematic data omission problem regarding gender features in invariant adjectives across Spanish treebanks. In the AnCora treebank, noun phrases like "*La admisión oficial*" ('The official admission'), the adjective *oficial* lacks gender annotation, as can be seen in figure 2.

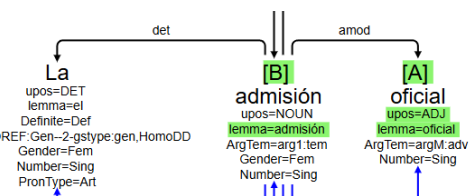


Figure 2: Annotation of "la admisión oficial" in Spanish-AnCora treebank

This contrasts with other Spanish treebanks, which show inconsistent approaches to the same adjective. For example, in GSD we find "*Es el segundo oficial organizado por*" ('It is the second official [event] organized by') with the adjective *oficial* marked as Gender=Masc, while in PUD "*Las*

fotos oficiales" ('The official photos') shows *oficiales* with Gender=Fem annotation, as can be seen in figure 3.

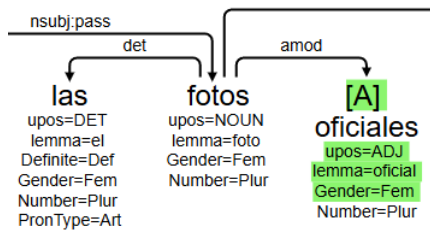


Figure 3: Annotation of "las fotos oficiales" in Spanish-PUD treebank

This inconsistency extends beyond Spanish. Comparing with closely related languages reveals that Portuguese systematically annotates gender on invariant adjectives, as in "*O nome oficial*" ('The official name') from the Bosque treebank, which includes Gender=Masc. Similarly, Italian treebanks show the same inconsistency pattern, with ISDT containing examples of "*ufficiale*" without gender feature while PUD consistently includes the feature.

The omission appears to stem from the lack of overt morphological marking for gender in invariant adjectives like *oficial*, which has the same form for both masculine and feminine. That is why we consider this to be a problem of disparity of annotators, who in an uncoordinated way interpret whether the morphological marking on the adjective takes precedence in order to decide not to mark the gender of that adjective which is in agreement with the noun it modifies.

Impact on Typological Universals This inconsistency affects Universal 31, which address adjectival agreement patterns. The inconsistent annotations would suggest that Spanish is from one specific type of language depending on the corpus the researcher uses. Additionally, the cross-linguistic inconsistency makes comparative analysis of gender agreement patterns difficult across related Romance languages.

5.3 Number in Dutch Pronouns: A Data Omission Error

Analysis of Dutch treebanks revealed a systematic omission of number features in pronouns across both Alpino and LassySmall treebanks. For example, the first-person plural pronoun *we* ('we') in sentences like "*We hebben een concept*" ('We

have a concept') from Alpino consistently lacks the Number=Plur feature. Similarly, the third-person plural pronoun *zij* ('they') in "*schamen voor wat zij*" ('ashamed of what they') from LassySmall is annotated without number, and the first-person singular *ik* ('I') in "*Ik geloof niet*" ('I don't believe') from Alpino lacks the Number=Sing feature.

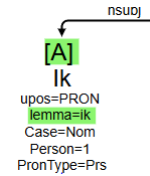


Figure 4: Annotation of "ik" in Dutch-Alpino treebank

This pattern extends to all personal pronouns in both Dutch treebanks, creating a systematic gap in the annotation of a fundamental morphological feature. The omission makes it difficult to compare Dutch pronominal systems with those of other languages, where number is consistently annotated on pronouns.

Impact on Typological Universals This fact of not marking number has significant implications for Universal 42, which deals with pronominal number distinctions. Based on Dutch treebanks alone, one might incorrectly conclude that Dutch pronouns lack explicit number marking, placing Dutch typologically with languages that genuinely lack such distinctions. This would represent a substantial mischaracterization of the Dutch pronominal system, which clearly distinguishes singular from plural forms both morphologically and syntactically. Furthermore, the systematic nature of this omission across all pronouns could significantly distort typological comparisons involving pronominal systems.

6 Discussion

Our investigation into annotation inconsistencies in UD treebanks reveals several important implications for typological research. This section examines the broader significance of our findings and proposes approaches to mitigate the impact of annotation errors on typological generalizations.

6.1 Implications for Typological Research

6.1.1 Reliability of Treebank-Based Typology

The systematic errors identified in our case studies raise legitimate concerns about the reliability of typological generalizations derived solely from

treebank data. However, this does not invalidate treebank-based approaches to typology. Rather, it highlights the need for methodological caution when using these resources for cross-linguistic comparison.

Our qualitative analysis suggests that annotation inconsistencies vary across treebanks, with converted resources generally showing more problematic patterns than natively UD-annotated corpora. This is particularly evident in the contrast between AnCora and PUD for Spanish, where PUD exhibits more linguistically motivated annotation of gender on adjectives. This may suggest that typological studies should account for the origin of treebanks when evaluating evidence.

6.1.2 Impact on Specific Universals

Our findings have varying implications for the universals under examination:

Universal 30 (Verbal Features) Universal 30 focuses on verbal tense-mood-aspect systems and is affected by two contrasting error types in Spanish treebanks. First, we observed undergeneration where some verbs receive incomplete tense-mood-aspect annotation while others show complete feature attribution. This inconsistent annotation makes it difficult to accurately characterize the Spanish verbal system in cross-linguistic analysis. Second, we identified an overgeneration problem where verbal participles in perfect constructions are incorrectly assigned gender features. This error conflates verbal and adjectival properties, making Spanish appear to have gender-marking on verbs in contexts where such marking is linguistically unmotivated. Together, these inconsistencies distort the typological classification of Spanish verbal morphology, potentially placing it incorrectly in relation to other languages based on both features it lacks and features it falsely appears to have.

Universal 31 (Gender Agreement) Universal 31 addresses patterns of gender agreement and is significantly impacted by the annotation errors we identified. The spurious assignment of gender to verbs in Spanish compound tenses, observed across multiple treebanks (AnCora, COSER, GSD, PUD), creates the false impression that Spanish typologically aligns with languages that genuinely mark gender on verbs. This overgeneration error artificially expands the scope of gender agreement in Spanish. Conversely, the omission of gender features on invariant adjectives could underrepresent

the extent of gender agreement in the language. The cross-linguistic inconsistency in handling invariant adjectives, as seen in our comparison between Spanish, Portuguese, and Italian treebanks, further complicates typological comparisons, as the same linguistic phenomenon receives different treatments across related languages.

Universal 42 (Pronominal Number) Universal 42 concerns pronominal number distinctions and is undermined by the systematic omission of number features on pronouns in Dutch treebanks. Our analysis revealed that both all pronouns, like *we* ('we') or *ik* ('I'), as well as third-person pronouns like *zij* ('they'), consistently lack number annotation in both Alpino and LassySmall treebanks. This pervasive omission could lead to the misclassification of Dutch as having a pronominal system without number distinctions, which would be a fundamental mischaracterization. This is particularly problematic for typological studies that rely on pronoun features to establish diachronic or areal patterns. The systematic nature of this omission across all pronouns in both treebanks suggests a guideline interpretation issue rather than random annotation errors, potentially affecting how Dutch relates typologically to other Germanic and European languages.

6.1.3 Methodological Implications

Our findings suggest that computational typologists should implement several methodological safeguards when working with UD data:

1. **Multi-treebank verification:** When multiple treebanks exist for a language, researchers should compare annotation patterns across resources to identify potential inconsistencies, as demonstrated by our comparison of different Spanish treebanks, even if they are not interested in all textual typologies.
2. **Conversion awareness:** Studies should explicitly account for whether treebanks were natively annotated in UD or converted from legacy formats, as conversion artifacts represent a significant source of errors.
3. **Cross-linguistic consistency checks:** Researchers should verify whether similar linguistic phenomena receive consistent annotation across related languages, as shown in our comparison of invariant adjectives across Romance languages.

4. **Annotation guideline consultation:** When discrepancies are found, reference to the UD guidelines can help determine which approach better reflects the intended annotation standard.

6.2 Improving UD for Typological Research

While our study identifies several challenges, we believe that UD remains an invaluable resource for computational typology. Based on our findings, we propose several improvements to enhance the reliability of UD for typological research:

6.2.1 Clearer Guidelines for Implicit Features

Many of the data omission errors identified stem from ambiguity regarding whether features should be annotated only when morphologically marked or also when syntactically relevant but not overtly marked. The UD guidelines could be enhanced with more explicit guidance on:

- Annotation of agreement features on invariant forms, as seen in the case of Spanish adjectives.
- Systematic annotation of inherent features on pronouns, as highlighted by the Dutch examples.

6.3 Balancing Universality and Accuracy

The tension between universal application and language-specific accuracy represents a fundamental challenge for cross-linguistic annotation projects. Our case studies illustrate how this tension can manifest in specific annotation decisions, such as whether to annotate gender on invariant adjectives or number on pronouns.

7 Conclusion

This study has identified and analyzed systematic annotation inconsistencies in UD treebanks that affect typological generalizations, with a focus on exemplifying it by morphological features in Spanish and Dutch. Our investigation revealed two primary categories of errors: overgeneration, where features are incorrectly applied to elements that should not have them, and data omission, where features are inconsistently or incompletely annotated. These errors have direct implications for the validity of typological universals derived from UD data.

7.1 Summary of Findings

Our case studies demonstrated specific instances of annotation inconsistencies with typological consequences:

- Incorrect assignment of gender features to verbal participles in Spanish compound tenses across multiple treebanks, creating a false impression that Spanish verbs carry gender marking.
- Inconsistent annotation of gender on invariant adjectives across Spanish treebanks, creating artificial variation within the same language.
- Cross-linguistic inconsistency in handling invariant adjectives across Romance languages.
- Systematic omission of number features on pronouns in Dutch treebanks, potentially leading to incorrect characterization of Dutch pronominal number distinctions.

7.2 Implications for Typology and Linguistic Complexity

These annotation inconsistencies significantly impact both typological research and linguistic complexity studies. As [Brosa-Rodríguez et al. \(2024\)](#) state, the relationship between typological universals and linguistic complexity is inversely proportional—languages sharing more universal features are generally considered less complex to learn as second languages.

The inconsistencies we identified distort complexity metrics by artificially inflating or deflating the morphological complexity of language systems. For instance, spurious gender assignments to Spanish verbs increase the apparent verbal complexity, while omitted number features in Dutch pronouns potentially underestimate pronominal complexity. Such distortions compromise cross-linguistic comparisons and may lead to incorrect predictions about second language acquisition challenges.

These implications underscore the need for researchers to carefully account for annotation inconsistencies when using UD data for both typological research and complexity measurements.

7.3 Contributions

This research makes several contributions to the field of computational typology:

- A typology of annotation errors that affect typological generalizations.

- Empirical evidence of specific inconsistencies in widely used UD treebanks.
- Methodological recommendations for typological research using UD.

7.4 Future Directions

Building on our findings, several promising directions for future research emerge:

- Development of validation procedures to identify typologically relevant annotation inconsistencies.
- Expansion of this analysis to other languages and language families.
- Investigation of how annotation inconsistencies affect typological metrics, language classification, and complexity measurements.
- Collaboration with the UD community to refine annotation guidelines.

7.5 Final Remarks

Despite the challenges identified, we remain optimistic about the value of UD for typological research and complexity studies. By acknowledging and addressing annotation inconsistencies, the computational linguistics community can enhance the reliability of treebank-based analyses, ultimately leading to more accurate characterizations of linguistic diversity, universals, and complexity. As multilingual NLP advances, improved consistency in linguistic annotations will strengthen both our theoretical understanding and the foundation for truly multilingual language technologies.

8 Limitations

While our study provides valuable insights into annotation inconsistencies in UD treebanks, several limitations should be acknowledged.

Our investigation relied primarily on an initial explorative qualitative analysis of specific examples rather than comprehensive quantitative assessment. This approach allowed for detailed linguistic analysis but limits our ability to make broad generalizations about the overall prevalence of these inconsistencies across UD treebanks.

The study focused on exemplifying in Spanish, Italian, Portuguese, and Dutch, Indo-European languages with similar typological profiles. This limited language sample may not capture the full range

of annotation challenges present across typologically diverse languages. Additionally, our analysis concentrated on morphology, leaving other syntactic features unexplored.

We have theorized about potential effects on universals 30, 31, and 42, but, due to lack of space, we have not empirically validated how correction of these errors would alter cross-linguistic generalizations in practice. This makes it difficult to assess the practical significance of these inconsistencies for typological research.

Our study offers limited insight into the underlying causes of these inconsistencies beyond the broad distinction between conversion artifacts and manual annotation variability. A more detailed understanding of annotation decision processes would provide valuable context for addressing these issues.

Finally, in some cases, multiple theoretically justified annotation approaches may exist for certain features. We did not systematically explore where annotation differences might reflect legitimate theoretical disagreements rather than errors, nor did we propose mechanisms for accommodating such variation within the UD framework.

References

- Javier Martín Arista. 2022. [Toward the morpho-syntactic annotation of an old english corpus with universal dependencies](#). *Revista de Linguística y Lenguas Aplicadas*, 17:85–97.
- Antoni Brosa-Rodríguez, M. Dolores Jiménez-López, and Adrià Torrens-Urrutia. 2024. [Exploring the complexity of natural languages: A fuzzy evaluative perspective on greenberg universals](#). *AIMS Mathematics*, 9:2181–2214.
- Antoni Brosa-Rodríguez and María Dolores Jiménez-López. 2023. [A typometrical study of greenberg’s linguistic universal 1](#). In *Distributed Computing and Artificial Intelligence. Lecture Notes in Networks and Systems*, pages 186–196. Springer.
- Kim Gerdes, Bruno Guillaume, Sylvain Kahane, and Guy Perrier. 2018. [Sud or surface-syntactic universal dependencies: An annotation scheme near-isomorphic to ud](#). In *Proceedings of the Second Workshop on Universal Dependencies*, pages 66–74. Association for Computational Linguistics.
- Kim Gerdes, Bruno Guillaume, Sylvain Kahane, and Guy Perrier. 2022. [Starting a new treebank? go sud! theoretical and practical benefits of the surface-syntactic distributional approach](#). In *SyntaxFest Depling 2021 - 6th International Conference on Dependency Linguistics*, pages 35–46.

- Kim Gerdes, Sylvain Kahane, and Xinying Chen. 2019. [Rediscovering greenberg’s word order universals in ud](#). In *Proceedings of the Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019)*, pages 124–131.
- Joseph H. Greenberg. 1963. *Universals of Language*. The M.I.T. Press. Citado intro.
- Bruno Guillaume. 2021. [Graph matching and graph rewriting: Grew tools for corpus exploration, maintenance and conversion](#). In *EACL 2021 - 16th conference of the European Chapter of the Association for Computational Linguistics*, pages 1–9.
- Martin Haspelmath. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language*, 86:663–687.
- Sylvain Kahane, Bernard Caron, Emmett Strickland, and Kim Gerdes. 2021. [Annotation guidelines of ud and sud treebanks for spoken corpora: a proposal](#). In *Proceedings of the 20th International Workshop on Treebanks and Linguistic Theories*.
- Marie-Catherine De Marneffe, Miriam Connor, Natalia Silveira, Samuel R Bowman, Timothy Dozat, and Christopher D Manning. 2013. [More constructions, more genres: Extending stanford dependencies](#). In *Proceedings of the Second International Conference on Dependency Linguistics*, pages 187–196.
- Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher D. Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. 2023. [Universal dependencies](#).
- Tae Hwan Oh, Ji Yoon Han, Hyonsu Choe, Seokwon Park, Han He, Jinho D. Choi, Na-Rae Han, Jena D. Hwang, and Hansaem Kim. 2020. [Analysis of the penn korean universal dependency treebank \(pkt-ud\): Manual revision to build robust parsing model in korean](#). In *Ithaca arXiv*.
- Helen O’Horan, Yevgeni Berzak, Ivan Vulić, Roi Reichart, and Anna Korhonen. 2016. Survey on the use of typological information in natural language processing. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1297–1308.
- Timothy Osborne and Kim Gerdes. 2019. [The status of function words in dependency grammar: A critique of universal dependencies \(ud\)](#). *Glossa*, 4.
- Slav Petrov, Dipanjan Das, and Ryan McDonald. 2012. [A universal part-of-speech tagset](#). In *Proceedings of the conference on Language Resources and Evaluation*, pages 2089–2096.
- Edoardo Maria Ponti, Helen O’Horan, Yevgeni Berzak, Ivan Vulić, Roi Reichart, Thierry Poibeau, Ekaterina Shutova, and Anna Korhonen. 2019. [Modeling language variation and universals: A survey on typological linguistics for natural language processing](#). *Computational Linguistics*, 45:1–56. Citado intro.
- Jianwei Yan and Haitao Liu. 2022. [Semantic roles or syntactic functions: The effects of annotation scheme on the results of dependency measures](#). *Studia Linguistica*, 76:406–428.
- Daniel Zeman. 2008. Reusable tagset conversion using tagset drivers. In *Proceedings of LREC*.