

Greenberg’s Universal 45 in Universal Dependencies: Gender Distinctions and Annotation Challenges

Antoni Brosa-Rodríguez & M. Dolores Jiménez-López

Universitat Rovira i Virgili

Avda. Catalunya 35, 43002 - Tarragona (Spain)

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

Abstract

This paper revisits and extends Greenberg’s Universal 45 on gender distinctions using Universal Dependencies 2.17, comprising 339 treebanks across 186 languages. A systematic analysis of morphosyntactic patterns confirms the implicational hierarchy (singular > plural gender marking), with 98.6 % conformity in pronominal categories. Only two potential exceptions are detected, both with minimal occurrences and likely attributable to annotation errors. Extending the analysis beyond pronouns to 13 UPOS categories shows that core categories maintain near-perfect compliance, while peripheral categories exhibit higher violation rates, primarily driven by annotation inconsistencies rather than genuine linguistic exceptions. A total of 90 treebanks display gender-number features in traditionally invariable categories (e.g., adpositions, conjunctions, adverbs), indicating annotation issues such as prepositional contraction handling, homophone merging, and erroneous feature assignment. The study establishes a replicable computational methodology for large-scale typological validation, highlighting both the potential of corpus-based approaches and key limitations, including genealogical sampling biases, annotation heterogeneity despite universal schemas, and the false sense of comparability across treebanks.

Keywords: Greenberg’s universals, gender-number marking, corpus annotation errors, typological validation

1. Introduction

In this study, we propose an effective method for carrying out a specific typological check using Universal Dependencies, linked to Greenberg’s universal 45: “If there are any gender distinctions in the plural of the pronoun, there are some gender distinctions in the singular also” (Greenberg, 1963). Universal 45 has been extensively studied in traditional typological studies, with different considerations based on grammars. Its interest lies in its great predictive power. However, the full potential of Universal Dependencies (Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajic, Christopher Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers and Daniel Zeman, 2020) for the study of gender distinctions in plural and singular pronouns has not yet been fully explored in the typological literature. That is why we conducted an exploratory study of the behavior of languages in this regard using version 2.17, which contains 339 treebanks corresponding to 186 different languages.

Beyond revisiting classic typological claims, we want to delve into a systematic method that allows us to provide new universals that offer detailed typological information on the behavior of the world’s languages. Therefore, specifically, our objectives are: To review Greenberg’s universal 45 based on occurrences in real texts in a broad sample of languages; to analyze the behavior of the languages available in UD in relation to gender distinction in different numbers of different parts of speech not

included in Greenberg’s claim; and to detect which UD corpora have annotations that generate exceptions to this universal so that they can be analyzed.

Apart from confirming a well-established typological generalization, the use of Universal Dependencies for this purpose serves two additional goals that justify the methodological choice. First, corpus-based verification of morphosyntactic universals such as Universal 45 allows the detection of annotation inconsistencies at scale. Second, while consulting descriptive grammars remains a valid strategy for typological research, especially in cases involving universals on simple features like this one, it is methodologically costly and can be captured through explicit morphosyntactic queries. Embedding such verifications within a UD-based pipeline makes it possible to integrate them with other corpus-derived typological data efficiently and systematically. In this sense, Universal 45 is not only an end in itself, but also a proof of concept for a replicable methodology applicable to a wide range of implicational universals.

2. State of the Art

Greenberg (1963), with a relatively small sample of languages and data from grammars (but with a precise and systematic methodology), formulated 45 universals that characterize the behavior of the world’s languages (or specific groups of them). In the case of the last universal, Greenberg showed

how, in pronouns, gender distinctions in the plural only occur if they also exist in the singular (Corbett, 1991, 2000).

We found multiple studies subsequent to this one that attempt to revisit this linguistic behavior with different samples and other assessments. Plank and Schellinger (1997) focus on expanding the study to include data from the grammars of many more languages in an attempt to find exceptions to this universal. Above all, contrasts were found between human and non-human in the plural (and not in the singular) in some (very few) languages, which could be equated with Greenberg's claim, although he did not specify this distinction. These types of relationships are particularly conditioned by person in the case of personal pronouns (Siewierska, 2013). Berg (2024), in fact, shows how the universal of 45 is particularly robust in relation to the third person, but not in the first person. In these cases, for reasons of informativeness, he argued (again based on grammars, in a very large sample) that it is more common to find gender distinctions in third-person singular than in plural, but also more in first-person plural than in singular. However, this approach would not constitute an amendment to Greenberg (since his claim would still hold true), but simply provide greater precision. Plank (2025) criticized and amended this paper, arguing that what is presented as new is something already anticipated by Greenberg. Therefore, in summary, despite the addition of some clarifications and improvements to the samplings, Greenberg's proposal remains unchanged.

In studies of token-based typology (Levshina, 2019), there are different approaches to studies based on Greenberg's word order universals (Liu, 2010; Guzmán-Naranjo and Becker, 2018; Gerdes et al., 2021; Brosa-Rodríguez and Jiménez-López, 2023; Brosa-Rodríguez and Kahane, 2024). However, universals such as 45 have been less explored, despite the apparent ease of formalization and query. Although the availability of languages in UD is not yet comparable to that available through grammars, the methodological advantages it offers, both in terms of performing classic typological checks and illuminating new correlations, make it an attractive and efficient perspective that has not been fully examined.

3. Methodology

To conduct the study, we adopt a structured computational approach that explicitly formalizes the analytical steps. The methodology is designed to ensure transparency, consistency, and full replicability.

3.1. Sample

Version 2.17 of Universal Dependencies contains 339 treebanks corresponding to 186 languages. Although this number is interesting, it is still not sufficient to carry out sampling other than convenience sampling (Miestamo et al., 2016). However, all macro-areas are represented (with a low number in the case of Australia, Papuan, and North America). Compared to WALS (Dryer and Haspelmath, 2013), there is a fairly even proportion of dominant word order types, despite a greater weighting of the SVO type (approximately +10%). The vast majority of languages are from Eurasia, and the Indo-European family is overrepresented compared to others. Therefore, given the exploratory nature of our study, we consider the language sample to be a good starting point that will reveal relevant information about the languages analyzed.

The types of UD (unified annotation) corpora are varied. Therefore, the corpora include a variety of oral and written texts (although the latter are in the majority), with very diverse origins: wikis, news, reviews, fiction, non-fiction, conversations, etc. This richness will allow us to reduce possible textual typology biases (when there are multiple corpora for a language) and to better represent different communicative situations that represent that language. The lack of restriction by corpus type also allows us to obtain the maximum number of languages available, which is especially relevant in the case of minority and minoritized languages, with much more limited availability (as is also the case with these same languages in the case of the grammar-based approach). Furthermore, we consider that the linguistic element we analyze is much more resistant to textual biases than other linguistic elements such as word order.

Another element of disparity is the size of the corpora. We can find corpora with less than 1K tokens (from languages with few resources) and other corpora with 3500K tokens (for example, in German). This disparity is taken into account and, therefore, neutralized with a simple binary calculation of compliance or non-compliance with the universal or alternative checks. However, exact quantitative results are also presented, allowing occurrences to be viewed. This type of data allows us to check the representativeness of the occurrences obtained, which we consider to be much better data than simply filtering by corpus size. The reason is that, given the linguistic structure sought, a small corpus can offer sufficient and reliable occurrences. Furthermore, we chose not to exclude languages with representativeness issues from the analysis, but rather to mark them, as we believe it is interesting to look at the insignificant occurrences obtained by experts in those languages and validate the possible annotation error detected (or

confirm the exception discovered).

3.2. Formalization

While Grew-Match (Guillaume, 2021) was not employed to run the queries, its formal query language, combined with UD annotation conventions, was adopted to develop a preliminary formal specification. This specification ensured the precise identification of the structures under investigation. In this case, in order to convert Greenberg's universal from natural language to formal language, we chose to present formalizations 1 and 2:

```
%45-PRONpl
pattern {PRON[upos=PRON, Gender, Number=Plur]} (1)
```

```
%45-PRONsg
pattern {PRON[upos=PRON, Gender, Number=Sing]} (2)
```

This step is considered important in order to clearly identify the conflict points in the query. If we limit ourselves strictly to Greenberg's universal, he speaks, on the one hand, of "gender distinctions." Therefore, we did not consider it appropriate to formalize with different genders, as the Gender feature encompasses any possible gender value and simply requires that such a distinction be made, something that is activated if any "gender" value exists. On the other hand, Greenberg referred to "in the plural" and "in the singular." Therefore, in this case, we refined the formalizations so that they capture these two orders, respectively, which are explicitly affected.

To extend the universal and analyze the behavior of other parts of speech, we take advantage of the formalization made and apply it to different "upos." For example, in the case of adjectives, this would be included in formalizations 3 and 4:

```
%45-ADJpl
pattern {PRON[upos=ADJ, Gender, Number=Plur]} (3)
```

```
%45-ADJsg
pattern {PRON[upos=ADJ, Gender, Number=Sing]} (4)
```

We repeated this type of formalization with other parts of speech until we reached 26 different formalizations. These included the formalization of gender distinction in the singular and plural forms of parts of speech that are considered invariable in their most canonical definitions (e.g., adpositions or adverbs). The parts of speech analyzed are: adjective (ADJ), adposition (ADP), adverb (ADV), auxiliary (AUX), coordinating conjunction (CCONJ), determiner (DET), interjection (INTJ), noun (NOUN), numeral (NUM), pronoun (PRON),

proper name (PROPN), subordinating conjunction (SCONJ), verb (VERB).

The inclusion of traditionally invariable categories (ADP, ADV, CCONJ, INTJ, SCONJ) serves as an explicit annotation quality control measure. Since these categories are not expected to bear gender or number features in any language, the working hypothesis is that all treebanks should yield zero occurrences for these patterns. Any deviation from this baseline is therefore treated as a potential annotation error rather than a genuine typological finding, and is analyzed accordingly.

The inclusion of nouns (NOUN) in the analysis deserves a brief justification. Unlike adjectives and pronouns, gender in nouns is typically considered a lexical rather than an inflectional feature, and in most languages it is not overtly marked on the noun itself. However, we include this category because languages may exhibit overt gender alternations on animate nouns (e.g. morphological masculine/feminine pairs), which some authors treat as genuine nominal gender inflection (Spencer, 2002). It allows us to examine whether this subset of cases follows the same "singular > plural" implicational hierarchy, while acknowledging that the results must be interpreted with caution given the fundamentally different nature of nominal gender.

3.3. Tools & Outputs

JupyterNotebook with Python language was used to carry out the analysis. To ensure the traceability and replicability of the study and to avoid extending the article with programming technicalities, all the steps followed are provided in an open-access repository¹.

In general, the analysis was structured in several interconnected phases: first, the 26 formalizations of universal 45 were loaded from a CSV file, specifying search patterns that combine part-of-speech (UPOS) with grammatical number (singular/plural) and gender. Next, the 339 treebanks from Universal Dependencies v2.17 were systematically processed, extracting each token from the CoNLL-U files along with its UPOS and relevant morphological features (gender and number). For each combination of treebank and UPOS, the presence of gender distinction in plural and singular was verified, identifying as "violations" those cases where a UPOS had gender in plural but not in singular, thus contradicting the implicational hierarchy proposed by Greenberg. Subsequently, the exact occurrences of gendered tokens were quantified to assess the statistical robustness of the violations detected, and finally the data was aggregated by language, merging the different treebanks of the same language to obtain a more general typological

¹<https://doi.org/10.34810/data3130>

perspective that would facilitate the interpretation of the patterns observed. Once this search has been carried out, various outputs that may be of interest to the research are requested, which are available in the repository previously mentioned in a detailed manner.

4. Results & Discussion

The results of the analysis provide empirically grounded insights that are pertinent both to typological research and to UD treebank development practices. First, we highlight that of the 339 treebanks, only 2 are exceptions to Greenberg’s original universal, which represents a violation rate of 0.01 in the plural/singular trade-off. The two notable corpora are “Hausa-WesternAutogramm” (Caron Bernard, 2026) and “Maghrebi_Arabic_french-Arabizi” (Riabi Arij, Farah Essaidi, Amal Fethi, Menel Mahamdi, and Djame Seddah, 2026). In the case of the first treebank, there is a double occurrence of the word “ta:sù” as a third-person plural feminine possessive pronoun. In the case of the second treebank, one occurrence of the word “t” is found as a second-person masculine plural pronoun. The size of the corpora, although not very large, is not small either, with 13K and 19K tokens, respectively. This, correlated with the low occurrence of the highlighted elements, may facilitate the interpretation that we are dealing with a possible annotation error. However, given our total lack of knowledge of the two languages involved, we recommend that experts in those languages or annotators of those corpora certify whether we are dealing with a surprising exception or a significant annotation error.

Be that as it may, whether the examples of these languages are eliminated as errors or the detected exceptions are confirmed, the data provide clear confirmation of the robustness of Greenberg’s proposal for the pronominal category. The results show that of the 139 treebanks in which gender distinction was detected in plural pronouns, 137 also showed this distinction in the singular, thus respecting the implicational universal. This proportion (98.6% conformity) constitutes a compelling empirical validation on a massive and multilingual scale.

In the case of checks on other parts of speech, the findings are equally revealing, as shown in Table 1. The analysis detected a total of 36 violations distributed across 13 different UPOS categories, albeit with very uneven patterns. The aggregated data reveal that the NUM (numerals) category has the highest absolute number of violations, with 11 cases detected in treebanks such as Latin-LLCT (629 plural occurrences), Polish-LFG (833 occurrences), and Breton-KEB (38 occurrences). However, when examining the relative vio-

lation rate—calculated as the number of violations divided by the number of treebanks with plural activation, shown in Figure 1—we observe that SCONJ (subordinating conjunctions) has the highest ratio (0.43), followed by CCONJ (coordinating conjunctions) with 0.40 and ADV (adverbs) with 0.15, as can be seen, all in grammatical categories where we would expect to find 0 occurrences, which invites us to review the data language by language to confirm possible annotation errors.

UPOS	Singular only	Plural only (exception)	Both	None
ADJ	15	0	138	186
ADP	15	1	14	309
ADV	9	3	17	313
AUX	42	1	67	229
CCONJ	4	2	3	330
DET	26	1	125	187
INTJ	4	0	3	332
NOUN	4	5	153	177
NUM	20	11	77	231
PRON	45	2	137	155
PROPN	39	4	99	197
SCONJ	7	3	4	325
VERB	24	3	110	202

Table 1: Number of treebanks (out of 339) showing gender distinction by grammatical number across UPOS categories.

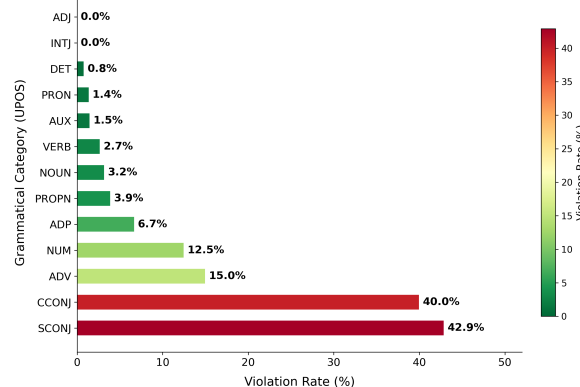


Figure 1: Violation Rate in the singular/plural trade-off by Part of Speech

It is particularly significant that the ADJ category (adjectives), traditionally considered one of the most prone to gender agreement in the world’s languages, shows no violation of the universal. Of the 138 treebanks in which gender distinction was detected in plural adjectives, 100% also showed this distinction in the singular. This behavior contrasts sharply with categories such as NUM, where the violation rate reaches 12.5%. A consistent pattern that emerges from the analysis is that the vast majority of violations detected (31 out of 36) corre-

spond to extremely low occurrences, with frequencies between 1 and 7 tokens, as shown in Figure 2. This data strongly suggests that many of these apparent exceptions could be due to annotation errors. The only significant exceptions from a quantitative point of view are concentrated in specific corpora of Latin, Polish, Hausa, and Breton, where plural occurrences exceed 30 instances.

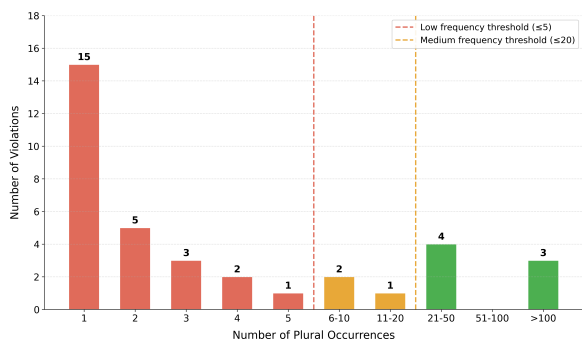


Figure 2: Distribution of Violation Magnitude: Occurrence Frequency

The case of Polish deserves special attention. The 833 occurrences of plural NUM in Polish-LFG without a singular counterpart can be explained by a specific annotation convention: the numeral *jeden* ('one') is systematically labeled as ADJ (adjective), since it can effectively function as such, but the option of labeling it as NUM is never considered, unlike other numerals such as *dwa* ('two') or *cztery* ('four'). Something similar occurs in Latin-LLCT, where *unus* is always considered DET, while *duo* is NUM (something that does not occur in other Latin corpora). This annotation decision creates the apparent mismatch that Polish or Latin lack singular numerals with gender distinction, producing an artifactual violation of the universal that does not reflect linguistic reality but rather a methodological issue in the application of the UPOS guidelines.

Of particular interest is the behavior of five grammatical categories traditionally considered invariable or with little morphological inflection: ADP (adpositions), ADV (adverbs), CCONJ (coordinating conjunctions), INTJ (interjections), and SCONJ (subordinating conjunctions). The presence of gender and number features in these categories in multiple treebanks points to possible systematic annotation problems or disparities in the criteria applied, which hinder the empirical study of linguistic universals. Specifically, 31 treebanks were identified with gender-number occurrences in ADP, 29 in ADV, 9 in CCONJ, 7 in INTJ, and 14 in SCONJ. The full list of affected treebanks for each category is provided in Appendix A.

Qualitative analysis of these cases reveals recurring patterns of annotative inconsistency. In Italian, for example, various corpora label prepositional

contractions such as *della* (de+la) under the ADP category, automatically inheriting the feminine gender and singular number features of the article *la*, while other corpora opt for a much more detailed and less problematic split. This practice generates the artificial appearance of gendered adpositions, when in reality the feature belongs to the joint component of the contraction. In the case of CCONJ in Italian-MarkIT, we observed a probable merge between the part of speech of the coordinating conjunction *e* ('and') and the morphological features of the copulative verb *è* ('is'), possibly due to errors in automatic morphological analysis or in the disambiguation of homophones. In corpora such as Portuguese-Bosque, we found occurrences of number features in ADP that correspond to possible annotation errors, such as the assignment of the singular feature to invariable prepositions such as *sem* ('without').

Therefore, as shown in Figure 2, the vast majority of occurrences are very marginal, possibly induced by a specific annotation error or oversight. These findings suggest that treebanks presenting gender-number features in invariable categories may benefit from a closer review of their annotations, with particular attention to: (a) potential automatic errors in morphological processing, (b) disparities between corpora of the same language applying divergent criteria, (c) alignment with practices in typologically similar languages, and (d) explicit documentation of any genuine phenomena that may justify the assignment of morphological features to traditionally invariable categories, as some of these cases may reflect real linguistic behavior rather than annotation errors.

In summary, despite the specific occurrences mentioned above, which we invite language experts to review in terms of exceptions or errors, the analysis carried out with UD clearly and efficiently allowed us to go beyond the universal 45 proposed by Greenberg and extend its implication/restriction of Singular > Plural in terms of gender distinction beyond the pronoun. This behavior seems to be transversal to all parts of speech that allow inflection, and its use is widely and robustly attested (whether or not there are specific exceptions), allowing for a better understanding of how languages work, as summarized in Figure 3.

5. Conclusions

This study has demonstrated the feasibility and potential of using Universal Dependencies as an empirical infrastructure for the systematic validation of linguistic universals on a large scale. Through the analysis of 339 treebanks corresponding to 186 languages in version 2.17 of UD, we have addressed the three objectives initially set out, with results that

combine robust confirmations, unexpected findings, and methodological caveats of relevance to the computational linguistics and typology community.

With regard to the first objective—to review Greenberg’s universal 45 on the basis of actual textual occurrences in a broad sample of languages—the results are categorical: the universal is overwhelmingly confirmed for the pronominal category. Of the 139 treebanks that show gender distinction in plural pronouns, 137 (98.6%) also exhibit this distinction in the singular, respecting the proposed implicational hierarchy. The only two exceptions detected (Hausa-WesternAutogramm and Maghrebi_Arabic_French-Arabizi) are based on extremely low occurrences (1-2 tokens) in moderately sized corpora, strongly suggesting annotation errors rather than genuine linguistic counterexamples, a fact that would make the universal 100% applicable. This massive empirical validation reinforces the predictive power of Greenberg’s universal six decades after its original formulation, demonstrating that generalizations derived from traditional grammatical analysis remain valid when tested against millions of tokens of real text.

The second objective—to analyze the behavior of languages in relation to gender distinction in different grammatical numbers in UPOS categories not originally considered by Greenberg—has yielded nuanced results that invite reflection on the limits of the extension of universals. While core categories such as ADJ (adjectives) and NOUN (nouns) show compliance rates of 100% and 96.8% respectively, categories such as NUM (numerals), SCONJ (subordinating conjunctions), and CCONJ (coordinating conjunctions) show significantly higher violation rates (12.5%, 42.9%, and 40%). However, qualitative analysis reveals that these apparent exceptions are mostly due to annotation peculiarities rather than genuine typological deviations. The examples shown illustrate how annotation conventions can generate artifacts that distort the evaluation of universals.

The third objective—to detect annotated UD corpora that generate exceptions to the universal for subsequent analysis—has resulted in the identification of systematic problems that transcend Greenberg’s specific universal. The presence of gender-number features corresponding to five traditionally invariable categories (ADP, ADV, CCONJ, INTJ, SCONJ) points to recurring annotation inconsistencies: prepositional contractions that inherit article features, merges between conjunctions and homophonic copulative verbs, and erroneous assignments of morphological features to invariable words. These findings underscore the need for systematic quality audits in UD treebanks, especially those generated through automatic processing without exhaustive manual validation.

More broadly, this work establishes a replicable and scalable method for the empirical verification of linguistic universals, combining formalization of morphosyntactic patterns, systematic corpus search, quantification of occurrences, and qualitative analysis of exceptions.

The availability of the Python code in an open repository facilitates its adaptation to other universals related to word order, agreement, case marking, or any phenomenon codable through UD morphosyntactic features.

6. Ethical considerations and limitations

This paper does not raise any specific ethical considerations.

Despite the findings reported in this study, some methodological and empirical limitations should be acknowledged, as they constrain the scope and generalizability of the conclusions, particularly regarding typological coverage and sampling biases, and the discrepancy between the universality of the annotation scheme and the variability of annotation practices.

6.1. Typological coverage & sampling biases

A fundamental limitation lies in the linguistic coverage of Universal Dependencies compared to traditional typological studies.

Although version 2.17 includes 186 languages, it has significant genealogical biases: Indo-European languages are massively overrepresented, while entire families from Austronesia, Papuan, or North America, for example, have minimal representation. This imbalance contrasts unfavorably with samples designed to maximize genealogical diversity, which ensured better typological balance.

The best-represented languages share relevant typological characteristics: two- or three-value gender systems and extensive morphological agreement.

Languages with radically different systems—nominal classifiers, polysynthetic languages, non-grammaticalized semantic gender marking—are underrepresented.

Additionally, treebanks are mostly built on standard languages written in formal registers, excluding oral varieties, dialects, or colloquial registers where exceptions might emerge.

6.2. From Universal Schema to Heterogeneous Annotation

A critical limitation that creates a false sense of methodological security is precisely the homo-

geneity of the Universal Dependencies annotation scheme. Although UD provides clear and universal general guidelines, in practice the annotation process leaves considerable room for interpretation in borderline phenomena.

As documented in the results, multiple treebanks of the same language exhibit radically different patterns in the assignment of morphological features to peripheral categories, not because of genuine linguistic differences but because of divergent methodological decisions by the annotation teams.

This de facto heterogeneity contrasts paradoxically with the impression of uniformity projected by the universal schema.

While in traditional typological studies based on descriptive grammars the heterogeneity of criteria is explicit and recognized, forcing the researcher to carefully homogenize descriptions, in studies based on UD there is a risk of assuming direct comparability where in reality significant interpretative differences persist. The weight of the individual annotator's criteria, although theoretically minimized by UD guidelines, continues to condition the reliability of the derived generalizations, generating both artifactual exceptions and masking genuine typological patterns.

Finally, the impossibility of manually verifying the 36 violations detected through consultation with specialists or native speakers—especially for minority languages—forces us to rely on statistical heuristics to distinguish genuine exceptions from methodological artifacts.

A related limitation concerns the token-based nature of the approach. In languages with invariable adjectives or other words that lack overt gender marking, annotators may nonetheless assign a gender feature based on syntactic context, since most words of that category in that language do inflect. This could produce apparent violations that reflect annotation practice rather than linguistic reality, as pointed out in [Brosa-Rodríguez and Jiménez-López \(2025\)](#).

7. Acknowledgements

This work was supported by the Universitat Rovira i Virgili *Pla de Foment de la Recerca (PFR)* of the Research Group on Mathematical Linguistics (GRLMC), under which Antoni Brosa-Rodríguez and M. Dolores Jiménez López conducted their research.

Antoni Brosa-Rodríguez's work was made possible through a research stay at the Universidad de Jaén, and he would like to express his sincere gratitude to the ALTYA (Análisis Lingüístico: Teoría y aplicaciones) Research Group for their support.

8. Bibliographical References

- Thomas Berg. 2024. Gender, number and person: a three-way interaction. *Linguistic Typology*, 28(3):537–565.
- Antoni Brosa-Rodríguez and M. Dolores Jiménez-López. 2023. *A Typometrical Study of Greenberg's Linguistic Universal 1*, Lecture Notes in Networks and Systems 741, pages 186–196. Springer.
- Antoni Brosa-Rodríguez and M. Dolores Jiménez-López. 2025. Beyond the data: The impact of annotation inconsistencies in ud treebanks on typological universals and complexity assessment. In *Proceedings of SIGTYP in ACL 2025*, pages 43–51. ACL.
- Antoni Brosa-Rodríguez and Sylvain Kahane. 2024. New proposal of greenberg's universal 14 from typometrics. In *Conference Proceedings of LREC-COLING 2024*, pages 12217–12226. ACL.
- Caron Bernard. 2026. *Hausa-Western Autogramm*. distributed via Universal Dependencies.
- Greville Corbett. 1991. *Gender*. Cambridge University Press.
- Greville Corbett. 2000. *Number*. Cambridge University Press.
- Matthew Dryer and Martin Haspelmath. 2013. *The World Atlas of Language Structures*. Max Planck Institute for Evolutionary Anthropology.
- Kim Gerdes, Sylvain Kahane, and Xinying Chen. 2021. Typometrics: From implicational to quantitative universals in word order typology. *Glossa*, 6(1).
- Joseph Greenberg. 1963. *Universals of Language*, 1st edition, pages 58–90. The M.I.T. Press, Cambridge.
- 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*.
- Matías Guzmán-Naranjo and Laura Becker. 2018. Quantitative word order typology with UD. In *17th International Workshop on Treebanks and Linguistics*.
- Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajic, Christopher Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers and Daniel Zeman. 2020. *Universal Dependencies*

v2: *An Evergrowing Multilingual Treebank Collection*. Proceedings of the 12th International Conference on Language Resources and Evaluation, 2.19.

Natalia Levshina. 2019. Token-based typology and word order entropy: A study based on universal dependencies. *Linguistic Typology*, 23:533–572.

Haitao Liu. 2010. Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua*, 6:1567–1578.

Matti Miestamo, Dik Bakker, and Antti Arppe. 2016. Sampling for variety. *Linguistic Typology*, 20(2):233–296.

Frans Plank. 2025. Typology re-done: Greenberg nos. 37, 45, 34. *Linguistic Typology*, Aop.

Frans Plank and Wolfgang Schellinger. 1997. The uneven distribution of genders over numbers: Greenberg nos. 37 and 45. *Linguistic Typology*, 1:53–101.

Riabi Arij, Farah Essaidi, Amal Fethi, Menel Mahamdi, and Djamé Seddah. 2026. *Maghrebi Arabic French-Arabizi*. distributed via Universal Dependencies.

Anna Siewierska. 2013. Gender distinctions in independent personal pronouns. *The World Atlas of Languages Structures*.

Andrew Spencer. 2002. Gender as an inflectional category. *Linguistics*, 38:279–312.

A. Treebanks with gender-number features in invariable categories

ADP (31 treebanks): Arabic-NYUAD, Bhojpuri-BHTB, Bulgarian-BTB, French-FQB, Greek-GDT, Greek-Messinian, Hindi-HDTB, Hindi-PUD, Icelandic-IcePaHC, Icelandic-PUD, Irish-Cadhan, Irish-IDT, Italian-ISDT, Italian-MarkIT, Italian-PUD, Italian-ParTUT, Italian-ParlaMint, Italian-PoSTWITA, Italian-VIT, Maghrebi_Arabic_French-Arabizi, Marathi-UFAL, Northern_Kurdish-Kurmanji, Old_Irish-DipSGG, Old_Irish-DipWBG, Portuguese-Bosque, Portuguese-DANTEStocks, Portuguese-GSD, Portuguese-PUD, Sindhi-Isra, Urdu-UDTB, Wolof-WTB.

ADV (29 treebanks): Arabic-NYUAD, Belarusian-HSE, Bhojpuri-BHTB, Faroese-OFT, French-ParisStories, German-GSD, German-HDT, Hindi-HDTB, Icelandic-GC, Icelandic-IcePaHC, Icelandic-Modern, Icelandic-PUD, Italian-ISDT,

Italian-KIParlaForest, Italian-MarkIT, Italian-Old, Italian-PUD, Italian-TWITTIRO, Italian-VIT, Portuguese-Bosque, Portuguese-GSD, Portuguese-PetroGold, Romanian-Nonstandard, Sanskrit-UFAL, Sanskrit-Vedic, Sindhi-Isra, Slovak-SNK, Upper_Sorbian-UFAL, Urdu-UDTB.

CCONJ (9 treebanks): Arabic-NYUAD, Belarusian-HSE, Bhojpuri-BHTB, Icelandic-IcePaHC, Italian-KIParlaForest, Italian-MarkIT, Italian-PUD, Italian-VIT, Urdu-UDTB.

INTJ (7 treebanks): Arabic-NYUAD, Beja-Autogramm, Bhojpuri-BHTB, Greek-Cretan, Greek-GUD, Greek-Lesbian, Icelandic-IcePaHC.

SCONJ (14 treebanks): Arabic-NYUAD, Beja-Autogramm, Bhojpuri-BHTB, Greek-GUD, Greek-Lesbian, Greek-Messinian, Hindi-HDTB, Icelandic-IcePaHC, Icelandic-PUD, Italian-MarkIT, Italian-VIT, Old_Irish-DipSGG, Portuguese-Bosque, Urdu-UDTB.