A Normalized Wiener Index for the Cross-Linguistic Comparison of Dependency Tree Shape

Anonymous ACL submission

Abstract

This paper proposes W', a normalized form of the Wiener index as a measure of a syntactic property of sentences which focuses on the overall shape of the dependency tree. It is a score of the shape of dependency trees on a $[0, 1]$ scale, in which the star tree is the lowest ($W' = 0$) and the path tree is the largest ($W' = 1$). W' is designed to have weaker empirical length dependence than other shape indices. This index is applied to all 21 languages of the Parallel Universal Dependencies treebank, and it is revealed that every mean W' is below 0.5, meaning that no language is mean-path-like, that so-called head-final languages belong to two groups (a path-leaning group and a star-leaning group), and that W' correlates negatively with mean dependency distance (MDD) (Pearson $r = -0.68$; Spearman $r = -0.79$). W' is also compared to normalized tree depth and to Freeman's centralization index of degree, showing that W' correlates strongly with the former, and that W' has substantially weaker empirical length dependence compared to the latter.

1 Introduction

The framework of Dependency Grammar (Tesnière, 1959), along with its operationalization (The Universal Dependencies (de Marneffe et al., 2021; Nivre et al., 2020; Zeman et al., 2017)), has allowed researchers in the field of quantitative syntax to focus on a cross-linguistic phenomenon: mean dependency distance (MDD).

MDD is the mean of the distances between a word and its head within a sentence. MDD is shorter than four across typologically diverse languages (Liu, 2008; Futrell et al., 2015; Liu et al., 2017). This phenomenon has been argued to be a result of the limit on working memory (Gibson, 1998, 2000; Gildea and Temperley, 2010; Temperley, 2007, 2008), and it has been found across different genres (Wang and Liu, 2017), second language proficiency levels (Ouyang and Jiang, 2018; Yan and Li, 2019; Li and Yan, 2021), and assessment contexts (Ouyang et al., 2022).

While MDD measures the *linear extent* of dependencies, i.e., how far the head and its dependent are located, it does not measure the *shape* of the dependency tree. For example, two sentences with identical MDD can have dependency trees in different shapes.

As such, the fact that MDD does not measure tree shapes is of significance; MDD may capture processing pressures in terms of linear distance between a word and its head, yet it does not capture them in terms of the hierarchical depth of dependency trees, which have already been discussed in prior research (Gibson, 1998; Grodner and Gibson, 2005).

In addition to this, the distribution of frequencies of dependency distances (DD) has been found to fit Zipf-Alekseev Distribution (Liu, 2007, 2009; Jiang and Liu, 2015; Niu et al., 2023). This is a significant research program made possible by focusing on DD for understanding universal properties of human languages. However, it also does not focus on the shapes of dependency trees, which are expected to show certain aspects of universality of human language.

In this context, we need to have a linguistically well-motivated and mathematically well-defined index to measure the shape of dependency trees so that we can deepen our understanding on universality of human language.

This study proposes W', a normalized form of the Wiener index (Wiener, 1947), as a measure for the shape of dependency trees. This index has the following three properties: (i) it is bounded in the range of $[0, 1]$, with extremes of the star tree ($W' = 0$), in which one word is the head of all the other words, and the path tree ($W' = 1$), in which one word depends on another which depends on

another, and so on; (ii) its bounds are not strongly influenced by sentence length, so its dependence on sentence lengths is substantially weaker than that of other shape indices; and (iii) it is computable in $O(g)$ time per sentence which has g tokens.

2 The Wiener Index

The Wiener index of a graph G (Wiener, 1947) is a topological index of a molecule, represented by the sum of geodesic distances over all unordered pairs of vertices in the chemical graph:

$$W(G) = \sum_{i<j} d(i,j). \tag{1}$$

For trees on g vertices, $W(G)$ is bounded between the star (minimum) and the path (maximum) (Entringer et al., 1976):

$$W_{\min}(g) = (g-1)^2, \quad W_{\max}(g) = \frac{g(g^2-1)}{6}. \tag{2}$$

For trees, $W(T)$ can be computed in $O(g)$ time without constructing the full distance matrix (Wiener, 1947; Mohar and Pisanski, 1988). Removing an edge e from T splits the tree into two components of sizes $n_1(e)$ and $n_2(e) = g - n_1(e)$, and that edge lies on exactly $n_1 \cdot n_2$ shortest paths; therefore:

$$W(T) = \sum_{e \in E(T)} n_1(e) \cdot n_2(e). \tag{3}$$

We compute $n_1(e)$ for each edge via a single post-order traversal that aggregates subtree sizes.

Following Tian and Choi (2013), the normalized Wiener index (represented as W' in this study) is defined as:

$$W' = \frac{W(T) - W_{\min}(g)}{W_{\max}(g) - W_{\min}(g)}. \tag{4}$$

As mentioned above, $W' \in [0,1]$, meaning that $W' = 0$ corresponds to a pure star tree, and $W' = 1$ corresponds to a pure path tree. For $g \geq 4$, W' is a scalar representation of tree shape whose bounds are invariant under sentence length, i.e., the same $[0,1]$ scale, with the same achievable extremes, applies at every g. For $g \leq 3$, the index is undefined because the star and path coincide.

The Wiener index has been developed in the tradition of chemical graph theory. This study attempts to apply the normalized Wiener index into the investigation of cross-linguistic quantitative comparison of the shapes of dependency trees for natural languages on the assumption that dependency trees are networks of words dependent on each other through dependency relations.

3 Related Work

3.1 Universal Dependencies and the PUD

Cross-linguistic comparisons of the present study are made possible on the Universal Dependencies framework (UD) and the Parallel Universal Dependencies treebank (de Marneffe et al., 2021; Nivre et al., 2020; Zeman et al., 2017) (PUD). The PUD contains 1,000 sentences which are further translated and annotated according to UD in parallel across 21 languages. The design of the PUD is expected to control the semantic content of these sentence pairs across languages, so if we observe any structural differences in the PUD, we can attribute them not to the topic or register, but to the structural variance of the languages in the PUD.

3.2 Dependency Distance and its Minimization

Since the seminal work by Liu (Liu, 2007, 2008), DD has been explored as one of the measures of syntactic complexity in the field of dependency grammar. The Dependency Distance Minimization (DDM) hypothesis (Gibson, 2000; Gildea and Temperley, 2010; Temperley, 2007, 2008) states that the syntax of human languages is organized to keep MDD short, possibly due to the limited capacity of working memory (Gibson, 1998; Grodner and Gibson, 2005). Cross-linguistic investigation (Futrell et al., 2015) and within-language variation across different genres and proficiency levels (Wang and Liu, 2017; Ouyang and Jiang, 2018; Yan and Li, 2019; Li and Yan, 2021) result in the finding that MDDs are consistently shorter than chance baselines.

3.3 Tree Depth

The concept of *tree depth* was proposed as a measure of syntactic complexity with respect to the memory load (Yngve, 1960, 1961), stating that there is a maximum depth of a left-branching structure that is considered to be bounded by working-memory capacity. The concept of tree depth has been developed in the field of quantitative linguistics; Köhler and Altmann (2000) investigated the distributional properties of depth in syntactic structures, and Yadav et al. (2020) compared SOV and

SVO languages in UD-based corpora with respect to tree depth as a measure of syntactic complexity. From the viewpoint of investigating indices for the shape of dependency trees, the tree depth of a given dependency tree can be considered as one of such indices, and what we need to know in this context is how related it is to other indices for the shape of dependency trees.

3.4 Graph Centrality and Dependency Trees

The present study advocates the assumption that dependency trees can be treated as graphs in the sense of social network analysis (Freeman, 1978; Wasserman and Faust, 1994). Anonymous Author (2010, 2013, 2014) applied centrality measures (Freeman's centralization index of degree) to quantify structural properties of sentences. For the definition of Freeman's centralization index of degree, see Appendix. Anonymous Author (2025) is a refinement of how to interpret this index; it argues that syntactic structures with high degree centralization are structurally *fragile*, less robust trees rather than more *complex* ones, pointing out that a star-like tree loses its global structure through removal of the central node, while a path-like tree preserves local interpretability through removal of any node. Provided that these studies (Anonymous Author, 2010, 2013, 2014, 2025) on graph-centrality measures may supplement dependency-distance-based measures for the quantification of structural properties of sentences, they have not yet explicitly attempted to investigate whether graph centrality measures are influenced by the difference of sentence length. In this context, the present study attempts to examine the length invariance of graph-centrality measures by comparing them to the Wiener index. If it is empirically shown that the Wiener index is more length-invariant than graph-centrality measures, then we can argue for the Wiener index over graph-centrality measures as far as length-invariance is concerned.

4 This Study

In this study, the index W' is applied to all 21 languages of the PUD, to answer the following questions: (1) are the languages in the PUD mean-path-like or mean-star-like? (2) Does the ranking of languages by W' differ from their ranking by MDD? (3) Does W' validate against normalized tree-depth measures (Yngve, 1960; Yadav et al., 2020)? and (4) is W' substantially less dependent on sentence length than Freeman's centralization index of degree?

Questions (1) and (2) are the main questions to be asked in the present study. Questions (3) and (4) are intended to examine the extent to which W' actually measures the shape of dependency trees by comparing it to other tree-shape measures.

4.1 Data

The PUD provides 21,000 sentences with UD annotations in parallel across 21 languages: Arabic, Chinese, Czech, English, Finnish, French, Galician, German, Hindi, Icelandic, Indonesian, Italian, Japanese, Korean, Polish, Portuguese, Russian, Spanish, Swedish, Thai, and Turkish. Of the 1,000 sentences per language, the first 750 are originally English (sentence-identifier prefix 01); the remaining 250 are originally German (02), French (03), Italian (04), or Spanish (05), and were translated to other languages via English. See `https://universaldependencies.org/conll17/` and `https://universaldependencies.org/` for details of the PUD.

4.2 Method

An original Python script is created by the author to calculate the normalized Wiener index, the MDD, the normalized tree-depth measure and Freeman centralization of each language in the PUD. Then, these figures are compared to answer the research questions (1) through (4). The code and processed data will be released upon publication.

The Wiener Index of each language is calculated in the following procedure: For the dependency tree for each sentence in a language in the PUD T with $g \geq 4$ tokens,

1. Multi-word tokens (IDs like "1-2") and empty nodes (IDs like "1.1") are excluded before building the graph.

2. Compute the Wiener index $W(T)$ of the current sentence using Equation 3, where $n_1(e)$ is the number of nodes on one side of edge e and $n_2(e) = g - n_1(e)$ is the number on the other side in the dependency tree. This is done in one post-order traversal of the tree.

3. Compute $W_{\min}(g)$ and $W_{\max}(g)$ using Equation 2.

4. Set W' using Equation 4.

The result is a single value in $[0, 1]$ for that sentence. Sentences with $g \leq 3$ are dropped because the denominator is zero or the formula is undefined.

Each language has its normalized Wiener index $\bar{W}_L$, which is the arithmetic mean of per-sentence W' values across the sentences of that language, after disregarding sentences with $g \leq 3$; for language L with sentences $T_1, ..., T_n$ (where n is 997–1000 depending on the language, since a small number of sentences are too short to admit W'):

$$\bar{W}_L = \frac{1}{n_L} \sum_{i=1}^{n_L} W(T_i). \quad (5)$$

Averaging per-sentence W' interprets each sentence as its own point in $[0, 1]$. Therefore, the per-sentence W' values of each language are treated as a sample from the distribution of dependency-tree shapes of that language.

5 Results

5.1 Sentence-level W' Distributions in the PUD

Table 1 summarizes the $\bar{W}_L$ and the MDD of each language in the PUD. The range of $\bar{W}_L$ across the 21 languages is from 0.248 (Japanese) to 0.435 (Korean). Note that no language has mean $\bar{W}_L$ above 0.5. This means that the sentences in these languages in the PUD tend to be more star-like than path-like on average, as far as the PUD is concerned. The left columns of Table 1 show the full ranking from most star-like to most path-like.

We found substantial within-language variation of W'. Standard deviations range from 0.067 (Japanese) to 0.128 (Korean), and the empirical range of W' across all sentences in the PUD is approximately $[0.08, 0.73]$. The highest W' observed for a single sentence is 0.73, well below the path-tree limit of 1.

The split between classically head-final languages (Korean and Turkish at the path-like end and Japanese and Hindi at the star-like end) illustrates that head-finality alone does not determine dependency-tree shape.

5.2 $\bar{W}_L$ versus MDD

The relationship between $\bar{W}_L$ and MDD across the 21 languages is strongly negative: Pearson $r = -0.68$ ($p = 7 \times 10^{-4}$), Spearman $r = -0.79$ ($p = 2 \times 10^{-5}$). Their negative correlation is expected: a pure path tree has minimum MDD (every edge has linear length 1) and $W' = 1$, while a star tree with the head centrally positioned has high MDD and $W' = 0$. Therefore, the empirical correlation is not only geometrically expected but also substantively informative. This result reflects how the dependencies are distributed in linear order as a function of tree shape.

The MDD ranking shown in the right columns of Table 1 is not an exact reverse of that of the $\bar{W}_L$ ranking. For example, Japanese has the lowest $\bar{W}_L$ (0.25) but only the eighth-lowest MDD (2.89) rather than the longest (it would be the longest if the ranking was completely reversed). This suggests that Japanese tends to have flat branching while its linear dispersion is moderate. Such cross-linguistic deviations in terms of the $\bar{W}_L$ and MDD of individual languages will be one of the topics for future research.

5.3 Validation of W' Against Tree Depth

In order to examine the extent to which W' actually measures the shape of dependency trees, it is compared to two indices for dependency trees with regard to their depths (Yngve, 1960): *max_depth* (the longest root-to-leaf path) and *mean_depth* (the mean root-to-token distance). Both of them are calculated from the corpus data in raw and normalized form (divided by $g - 1$ and $(g - 1)/2$ respectively). Table 2 shows the values of $\bar{W}_L$ and tree depths in the languages in the PUD. Note that Japanese has the second-largest raw mean depth (2.50) despite being the most star-like language by $\bar{W}_L$ (0.248), while Finnish has the smallest raw mean depth despite being ranked only 15th-most-star-like by $\bar{W}_L$.

The sentence level shows weak and inconsistent Spearman correlations between W' and raw depth measures ($+0.09$ for *max_depth* and $+0.02$ for *mean_depth*) (Figure 1). By contrast, strong and uniform correlations are found between W' and normalized depth measures: mean Spearman of $+0.75$ for normalized *max_depth* and $+0.71$ for normalized *mean_depth* (Figure 2). This pattern seems to validate W' as a shape index: If sentence length is factored out, W' correlates strongly with the depth-based shape measures. If not, on the other hand, it barely correlates with raw depth measures.

The strong correlation between W' and normalized depth offers us support for their validity as measures for dependency trees. They can be considered as complementary operationalizations of the underlying property of dependency trees. While normalized depth is computationally simple, W'

Rank	Lang (by $\bar{W}_L$)	$\bar{W}_L$	Lang (by MDD)	MDD
1	ja	0.248	th	2.370
2	de	0.271	ko	2.541
3	fr	0.276	tr	2.697
4	hi	0.276	fi	2.832
5	it	0.280	id	2.838
6	en	0.280	pl	2.858
7	pt	0.287	is	2.861
8	gl	0.290	ja	2.895
9	zh	0.292	ru	2.914
10	es	0.292	ar	3.008
11	sv	0.294	cs	3.010
12	is	0.310	sv	3.028
13	id	0.322	es	3.063
14	cs	0.323	pt	3.068
15	fi	0.324	gl	3.070
16	ru	0.330	fr	3.080
17	pl	0.334	it	3.086
18	ar	0.339	en	3.139
19	th	0.344	zh	3.290
20	tr	0.372	hi	3.533
21	ko	0.435	de	3.705

Table 1: The 21 PUD languages ordered by increasing $\bar{W}_L$ (left columns) and by increasing MDD (right columns). Rank 1 corresponds to the most star-like (lowest $\bar{W}_L$) and to the shortest MDD respectively. Abbreviations: ar; Arabic: cs; Czech: de; German: en; English: es; Spanish: fi; Finnish: fr; French: gl; Galician: hi; Hindi, id; Indonesian: is; Icelandic: it; Italian: ja; Japanese; ko; Korean: pl; Polish: pt; Portuguese: ru; Russian: sv; Swedish: th; Thai: tr; Turkish: zh; Mandarin Chinese.

Rank	Lang	$\bar{W}_L$	*max*	*mean*	*norm_max*	*norm_mean*
1	ja	0.248	4.89	2.50	0.190	0.197
2	de	0.271	4.21	2.21	0.226	0.240
3	fr	0.276	4.67	2.44	0.215	0.227
4	hi	0.276	4.63	2.37	0.222	0.229
5	it	0.280	4.57	2.41	0.221	0.236
6	en	0.280	4.33	2.26	0.232	0.244
7	pt	0.287	4.52	2.41	0.222	0.239
8	gl	0.290	4.64	2.46	0.227	0.243
9	zh	0.292	4.30	2.21	0.234	0.242
10	es	0.292	4.59	2.43	0.227	0.242
11	sv	0.294	4.13	2.15	0.248	0.261
12	is	0.310	4.34	2.22	0.262	0.271
13	id	0.322	4.35	2.27	0.257	0.270
14	cs	0.323	4.39	2.29	0.271	0.285
15	fi	0.324	3.90	1.99	0.286	0.296
16	ru	0.330	4.47	2.31	0.266	0.277
17	pl	0.334	4.41	2.26	0.276	0.285
18	ar	0.339	4.85	2.48	0.270	0.278
19	th	0.344	5.18	2.66	0.266	0.276
20	tr	0.372	4.33	2.19	0.295	0.301
21	ko	0.435	4.92	2.40	0.341	0.334

Table 2: The 21 PUD languages ordered by increasing $\bar{W}_L$ (left columns), along with *max_depth*, *mean_depth*, *normalized max_depth* and *normalized mean_depth*. Rank 1 corresponds to the most star-like (lowest $\bar{W}_L$). Abbreviations: ar; Arabic: cs; Czech: de; German: en; English: es; Spanish: fi; Finnish: fr; French: gl; Galician: hi; Hindi, id; Indonesian: is; Icelandic: it; Italian: ja; Japanese; ko; Korean: pl; Polish: pt; Portuguese: ru; Russian: sv; Swedish: th; Thai: tr; Turkish: zh; Mandarin Chinese.

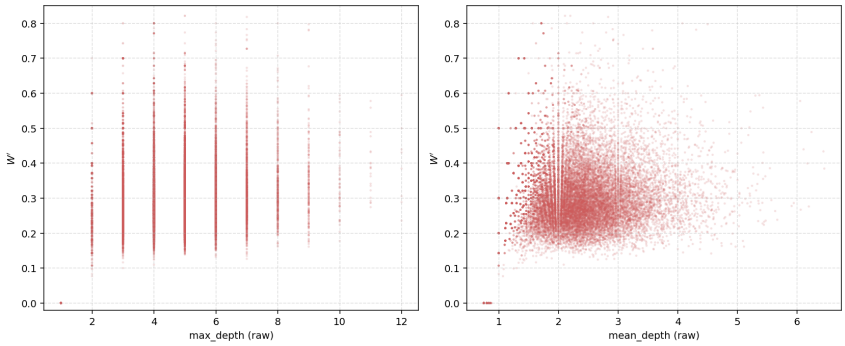

Figure 1: Sentence-level W' versus raw depth measures (21,000 sentences in the PUD)

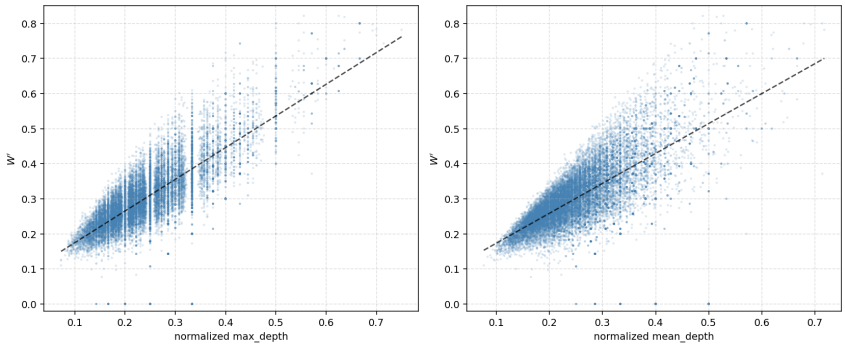

Figure 2: Sentence-level W' versus normalized depth measures (21,000 sentences in the PUD)

offers a more principled theoretical foundation (applied in the field of chemical graph theory) with length-invariant bounds represented by the $[0, 1]$ range, which normalized depth lacks. Researchers may choose either index according to their objectives.

5.4 Length Dependence of W' and Comparison with Freeman Centralization

The previous section showed that W' correlates with depth-based shape measures only when those measures are normalized with respect to sentence length. A related question is how independent W' itself is from sentence length. The bounds of W' are invariant under g: every tree on at least four nodes admits a value in $[0, 1]$, and the extremes 0 and 1 are achievable at every length. However, how W' values are distributed within those bounds is another issue.

Across the 21,000 PUD sentences (excluding those with $g \leq 3$), W' correlates with sentence length at Pearson $r = -0.45$ and Spearman $r = -0.49$. Within each language, the Spearman correlation with g ranges from -0.30 (Finnish) to -0.56 (Spanish), with no language giving a value near zero. That is, longer sentences in the PUD tend toward more star-like values of W', even though the $[0, 1]$ range is technically available to them at every length. This is a substantive empirical observation about natural-language structure rather than an artifact of the index: longer sentences in the PUD favor flatter, more star-like organization, plausibly in response to processing constraints on deeply nested structure (Gibson, 1998).

On the other hand, Freeman's centralization index of degree gives a Spearman of -0.70 with sentence length on the same PUD sentences, and within-language values range from -0.61 to -0.77. Figure 3 shows the contrast: the Freeman C_D scatters in the hyperbolic envelope, while the W' scatter is modestly downward-trending, inside the constant $[0, 1]$ envelope.

This comparison clarifies what W' achieves and what it does not. The bounds of W' are length-invariant by construction, and its empirical length dependence observed in the PUD is substantially weaker than Freeman C_D's. It must be pointed out that W' is not completely independent from sentence length: the residual -0.49 correlation indicates that the conditional distribution of shape given length shifts with length in the PUD. For applications that compare tree shapes across populations with different length distributions, controlling for sentence length may be necessary in addition to using W' rather than Freeman C_D, for example,

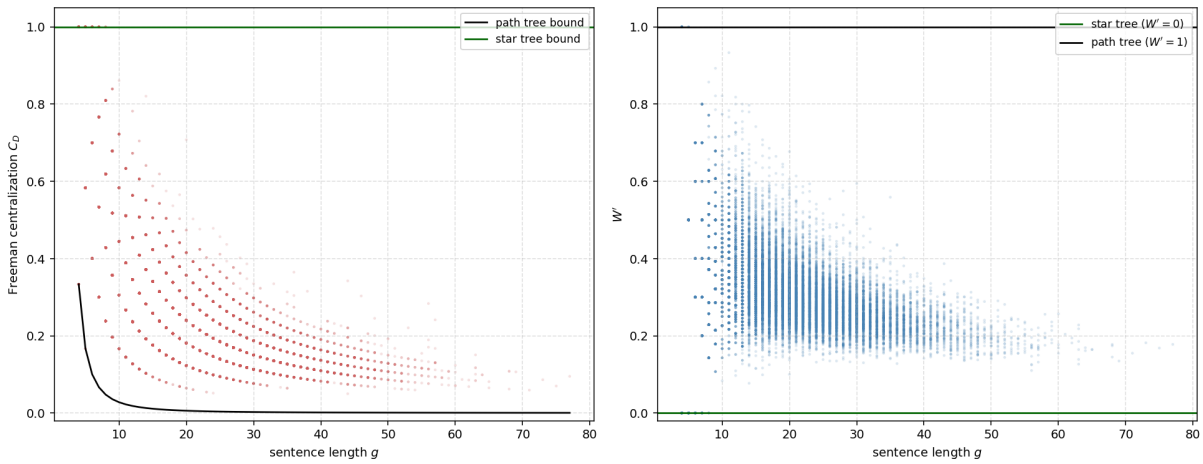

Figure 3: Length dependence of two shape indices on 21,000 PUD sentences. *Left:* Freeman centralization C_D. *Right:* the normalized Wiener index W' with constant bounds at the star ($W' = 0$) and path ($W' = 1$).

by analyzing residuals of W' after regressing on g.

6 Discussion

6.1 A connection to DDM

The result that W' is strongly negatively correlated to MDD suggests that tree-shape pressures are not independent from DDM (Gibson, 2000; Gildea and Temperley, 2010; Futrell et al., 2015, etc.). More path-like trees tend to be produced in languages with shorter linear edge length only to some extent, as every PUD language has mean W' below 0.5, a preference for more star-like dependency trees. Provided that shorter DDs (hence path-like shapes) are preferred due to processing pressures such as working memory, there are also structural pressures which function against deep nesting, thus more star-like dependency trees are produced on average. In this regard, we need to investigate the background of structural pressures against deep nesting, which will be a question for future study.

6.2 A typological observation

The result of comparing $\bar{W}_L$ and MDD in the PUD strongly suggests that classical typological labels such as branch or word order leave their variations unexplained. The most conspicuous case is the contrast between Korean and Japanese: Both are head-final SOV languages whose native speakers are closely located geographically, yet their positions on the W' scale are opposite. While Korean's high W' and low MDD are consistent with deep and linearly compact subordination, Japanese's low W' at moderate MDD reflects more star-like shapes of its dependency trees, in which a clause-final predicate is the head of many other words. Similar contrasts can be found in other language pairs where their same head positions (head-final or head-initial) and their same word order (SOV, SVO, etc.) do not directly result in their overall dependency-tree shapes.

6.3 A methodological recommendation

As shown in Section 5.3, raw tree depth relates to W' quite differently from normalized tree depth does. Section 5.4 also highlights this issue of the effect of sentence length: While Freeman's centralization of degree exhibits a strong negative correlation with sentence length, the bounds of W' are length-invariant by definition, and its length dependence is empirically weaker. These results suggest that typological interpretation of research results like the present study requires normalization of shape measures with respect to sentence length.

7 Conclusion

This study introduced W', a normalized Wiener index for the analysis of dependency-tree shape. Across the 21 languages of the PUD, W' reveals that no language is mean-path-like, that head-final languages split into Korean-Turkish (path-like) and Japanese-Hindi (star-like) groups, and that W' and MDD together describe a two-dimensional space orthogonal to standard typological categories. Validation against normalized depth measures and Freeman centralization shows that W' behaves as expected: a shape index with length-invariant bounds, computable in linear time, with achievable extremes at the star and path. Its empirical length dependence is substantially weaker than that

of Freeman centralization, though not zero, as far as the PUD is concerned. It is recommended that cross-population comparisons of dependency-tree shape control for sentence length even when W' is used. W' is expected to complement existing dependency-distance measures in cross-linguistic studies of syntactic complexity.

Limitations

Several limitations should be noted. First, the PUD provides only 1,000 sentences per language, the majority of which (750) originate in English, with the remaining 250 originating in four Western European languages. The translation pivot may introduce structural biases — including effects on length and clausal nesting — that constrain how strongly typological conclusions can be drawn. Second, the choice of root in UD annotation conventions affects depth-based measures (though not W' itself, which is root-invariant by construction), introducing some annotation-driven noise into the depth comparisons. Third, W' is invariant under the linear order of tokens by construction, and therefore cannot distinguish between projective and non-projective dependency structures with identical underlying trees; capturing the projectivity dimension would require a separate measure. Fourth, W' measures elongation but not left-right balance: two trees that are equally elongated but differently balanced (e.g., a left-heavy vs. a right-heavy subordination chain) receive identical W' values. Future work should validate W' against balance-based shape indices.

Acknowledgments

anonymized

References

Anonymous Author. 2010. Title and venue redacted for blind review.

Anonymous Author. 2013. Title and venue redacted for blind review.

Anonymous Author. 2014. Title and venue redacted for blind review.

Anonymous Author. 2025. Title and venue redacted for blind review.

Marie-Catherine de Marneffe, Christopher Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2):255–308.

Roger C. Entringer, Douglas E. Jackson, and D. A. Snyder. 1976. Distance in graphs. *Czechoslovak Mathematical Journal*, 26(2):283–296.

Linton C. Freeman. 1978. Centrality in social networks: Conceptual clarification. *Social Networks*, 1(3):215–239.

Richard Futrell, Kyle Mahowald, and Edward Gibson. 2015. Large-scale evidence for dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112(33):10336–10341.

Edward Gibson. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1):1–76.

Edward Gibson. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In Alec P. Marantz, Y. Miyashita, and W. O'Neil, editors, *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*, pages 95–126. MIT Press, Massachusetts.

Daniel Gildea and David Temperley. 2010. Do grammars minimize dependency length? *Cognitive Science*, 34(2):286–310.

Daniel Grodner and Edward Gibson. 2005. Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, 29(2):261–290.

Jingyang Jiang and Haitao Liu. 2015. The effects of sentence length on dependency distance, dependency direction and the implications — based on a parallel English–Chinese dependency treebank. *Language Sciences*, 50:93–104.

Reinhard Köhler and Gabriel Altmann. 2000. Probability distributions of syntactic units and properties. *Journal of Quantitative Linguistics*, 7(3):189–200.

Wenping Li and Jianwei Yan. 2021. Probability distribution of dependency distance based on a treebank of Japanese EFL learners' interlanguage. *Journal of Quantitative Linguistics*, 28(2):172–186.

Haitao Liu. 2007. Probability distribution of dependency distance. *Glottometrics*, 15(1):1–12.

Haitao Liu. 2008. Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, 9(2):159–191.

Haitao Liu. 2009. Probability distribution of dependencies based on Chinese dependency treebank. *Journal of Quantitative Linguistics*, 16(3):256–273.

Haitao Liu, Chunshan Xu, and Junyin Liang. 2017. Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, 21:171–193.

Bojan Mohar and Tomaž Pisanski. 1988. How to compute the Wiener index of a graph. *Journal of Mathematical Chemistry*, 2(3):267–277.

Ruochen Niu, Yaqin Wang, and Haitao Liu. 2023. The cross-linguistic variations in Dependency Distance Minimization and its potential explanations. In *Proceedings of the 37th Pacific Asia Conference on Language, Information and Computing (PACLIC)*.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. 2020. Universal Dependencies v2: An evergrowing multilingual treebank collection. In *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)*, pages 4034–4043, Marseille, France.

Jinghui Ouyang and Jingyang Jiang. 2018. Can the probability distribution of dependency distance measure language proficiency of second language learners? *Journal of Quantitative Linguistics*, 25(4):295–313.

Jinghui Ouyang, Jingyang Jiang, and Haitao Liu. 2022. Dependency distance measures in assessing L2 writing proficiency. *Assessing Writing*, 51:100603.

David Temperley. 2007. Minimization of dependency length in written English. *Cognition*, 105(2):300–333.

David Temperley. 2008. Dependency length minimization in natural and artificial languages. *Journal of Quantitative Linguistics*, 15(3):256–282.

Lucien Tesnière. 1959. *Éléments de Syntaxe Structurale*. Klincksieck, Paris.

Dechao Tian and Kwok Pui Choi. 2013. Sharp bounds and normalization of Wiener-type indices. *PLOS ONE*, 8(11):e78448.

Yaqin Wang and Haitao Liu. 2017. The effects of genre on dependency distance and dependency direction. *Language Sciences*, 59:135–147.

Stanley Wasserman and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge.

Harry Wiener. 1947. Structural determination of paraffin boiling points. *Journal of the American Chemical Society*, 69(1):17–20.

Himanshu Yadav, Ashwini Vaidya, Vishakha Shukla, and Samar Husain. 2020. Word order typology interacts with linguistic complexity: A cross-linguistic corpus study. *Cognitive Science*, 44(4):e12822.

Hengbin Yan and Yanghui Li. 2019. Beyond length: Investigating dependency distance across L2 modalities and proficiency levels. *Open Linguistics*, 5(1):601–614.

Victor H. Yngve. 1960. A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, 104(5):444–466.

Victor H. Yngve. 1961. The depth hypothesis. In Roman Jakobson, editor, *Structure of Language and Its Mathematical Aspects*, pages 130–138. American Mathematical Society, Providence, RI.

Daniel Zeman, Martin Popel, Milan Straka, Jan Hajič, Joakim Nivre, Filip Ginter, Juhani Luotolahti, Sampo Pyysalo, Slav Petrov, Martin Potthast, Francis Tyers, Elena Badmaeva, Memduh Gökırmak, Anna Nedoluzhko, Silvie Cinková, Jan Hajič jr., Jaroslava Hlaváčová, Václava Kettnerová, Zdeňka Urešová, and 43 others. 2017. CoNLL 2017 shared task: Multilingual parsing from raw text to Universal Dependencies. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1–19, Vancouver.

A Appendix

For a dependency tree T treated as an undirected graph on g nodes, the per-node degree centrality C'_D and closeness centrality C'_C can be calculated in their normalized social-network forms (Freeman, 1978; Wasserman and Faust, 1994), where $d(x)$ denotes graph distance:

$$C'_D(n_i) = \frac{d(n_i)}{g-1}, \quad C'_C(n_i) = \frac{g-1}{\sum_{j \neq i} d(n_i, n_j)}, \tag{6}$$

Per-node centralities can be aggregated into a single sentence-level value in several ways. One of them, used throughout the prior centrality-based dependency-grammatical literature (Anonymous Author, 2010, 2013, 2014), is Freeman's centralization index:

$$C_D(T) = \frac{\sum_i [\Delta(T) - d(n_i)]}{(g-1)(g-2)}, \tag{7}$$

where $\Delta(T)$ is the maximum degree present in T. The numerator measures how concentrated the degree distribution is around its maximum; the denominator $(g-1)(g-2)$ is the value the numerator attains for a star graph on g nodes (S_g), so that $C_D(S_g) = 1$.

Two facts about the behavior of C_D under variation of sentence length are worth making explicit. For the star tree on g nodes, $C_D(S_g) = 1$ regardless of g. For the path tree, where $\Delta = 2$ and only two nodes (the endpoints) have degree 1, the numerator equals 2 for every $g \geq 3$, so

$$C_D(P_g) = \frac{2}{(g-1)(g-2)}, \tag{8}$$

which shrinks toward zero as g grows: $C_D(P_5) \approx 0.167$, $C_D(P_{10}) \approx 0.028$, $C_D(P_{50}) \approx 0.0009$. More generally, any tree whose maximum degree

remains bounded as g grows has $C_D = \Theta(1/g)$ asymptotically. This means that the bounds of C_D depend on g and that the same value of C_D carries different shape information at different sentence lengths. A short path-shaped sentence and a long path-shaped sentence can differ in C_D by orders of magnitude. As a result, C_D thus has a fundamental limitation as a length-comparable scalar representation of dependency-tree shape.