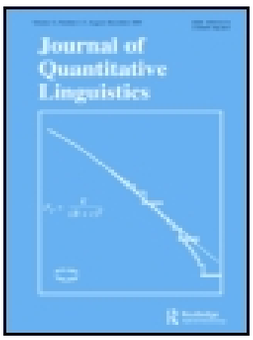

Journal of Quantitative Linguistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/njql20

# Dependency Distances and Their Frequencies in Indo-European Language

Xinying Chen & Kim Gerdes

**To cite this article:** Xinying Chen & Kim Gerdes (2020): Dependency Distances and Their Frequencies in Indo-European Language, Journal of Quantitative Linguistics, DOI: 10.1080/09296174.2020.1771135

**To link to this article:** https://doi.org/10.1080/09296174.2020.1771135

Published online: 18 Jun 2020.

JOURNAL OF QUANTITATIVE LINGUISTICS
https://doi.org/10.1080/09296174.2020.1771135

Routledge
Taylor & Francis Group



# Dependency Distances and Their Frequencies in Indo-European Language

Xinying Chen[a,b] and Kim Gerdes[c,d,e]

[a]Department of Czech Language, University of Ostrava, Ostrava, Czech Republic; [b]School of Foreign Studies, Xi'an Jiaotong University, China; [c]LPP (CNRS); [d]Institute of General and Applied Linguistics and Phonetics, Sorbonne Nouvelle, France; [e]Almanach (Inria)

**ABSTRACT**

The present study investigates the relationship between two features of dependencies, namely, dependency distances and dependency frequencies. The study is based on the analysis of a parallel dependency treebank that includes 10 Indo-European languages. Two corresponding random dependency treebanks are generated as baselines for comparison. After computing the values of dependency distances and their frequencies in these treebanks, for each lan-guage, we fit four functions, namely quadratic, exponent, logarithm, and power-law func-tions, to its original and random datasets. The preliminary result shows that there is a rela-tion between the two dependency features for all 10 Indo-European languages. The relation can be further formalized as a power-law function which can distinguish the observed data from randomly generated datasets.

## 1. Introduction

Dependency distance is the linear distance between a word and its head (Futrell et al., 2015; Hudson, 1995; Liu et al., 2017; Temperley, 2007).[1] It becomes a popular topic in quantitative linguistics because it has a rather clear definition and it is very easy to quantify. Dependency distance can be calculated as the difference of the word ID and its governor's ID in a CoNLL style dependency treebank (Buchholz & Marsi, 2006). Furthermore, the development of large-scale dependency treebanks provides more and more datasets for dependency distance analysis. Therefore, the interest of it has constantly grown.

One of the most important and interesting outcomes is probably the dependency distance minimization phenomena. Liu (2008) and Futrell et al. (2015) investigated the distributions of dependency distance in different languages and they discovered that there is a universal trend of minimizing the dependency distance in human languages. A well-accepted interpretation of this phenomenon is that dependency distance is somehow connected to the short-term memory (or working memory) of human beings and the least

**CONTACT** Xinying Chen cici13306@gmail.com

effort principle (Zipf, 1949) as well. More specifically, a long dependency distance is more difficult or inefficient to process because it requires more memory storage. Consequently, in order to lower the processing difficulty and improving communication efficiency, short dependencies are in favoured according to the least effort principle.

Despite the significant success of dependency distance minimization studies, there are still some gaps to be filled. One of them is the neglected value of individual dependencies. Current dependency minimization studies or associated studies (Futrell et al., 2015; Liu, 2008; Liu et al., 2017) are mainly focusing on the mean dependency distance per sentence rather than every single dependency in a dataset. No doubt that mean dependency distance is a valuable sentence-related (or tree-related) feature to observe. Especially when the studies discuss other tree-related features such as tree heights and widths, mean dependency distance is a more easily comparable feature than a group of individual dependency distances. However, studies also show that individual dependency distances (Liu, 2010; Chen & Gerdes, 2017, 2018) provide more details of the fluctuation than the average, which would level-up differences of dependencies in a sentence. It deserves more attention and discussions.

Therefore, our study here is trying to pick up the missing detail of previous studies by investigating two features of individual dependencies. Distance and frequency are two of the most objective features of each dependency link and they are both easy to quantify. Since we know that natural languages have the tendency to minimize the dependency distance, we would expect much more short distances rather than long distances in dependency links. In other words, the two features of individual frequencies, distance, and frequency, should be inversely related and probably can be formulated as a function. Chen and Gerdes (2019) tested four functions on an English treebank and their results show that the relationship between these two features for English can be formulated as either an exponential or a power-law function. In this study, we would like to investigate that whether such a claim can hold on for other Indo-European languages. To be more specific, we would like to find out for other Indo-European languages, whether the relationship between distance and frequency of individual dependencies can also be formulated as functions. Hereby, our hypothesis is:

- *For Indo-European languages, the relation between distance and frequency can be formulated as a function.*

The paper is structured as follows. Section 2 describes the data set, the Parallel Universal Dependencies (PUD) treebank of Surface-syntactic Universal Dependencies (SUD) treebanks (Gerdes et al., 2018, 2019), and Section 3 presents our method of analysis and introduces our random

baselines. Section 4 is dedicated to the empirical results and discussions. Finally, Section 5 presents our conclusions.

## 2. Material

Universal Dependencies (UD) is a project of developing a cross-linguistically consistent treebank annotation scheme for many languages, with the goal of facilitating multilingual parser development, cross-lingual learning, and parsing research from a language typology perspective. This collaborative project involves dozens of research groups around the world, proposing treebanks based on a universal scheme for more than 80 languages from different language families (De Marneffe & Nivre, 2019; Nivre et al., 2019). It is the largest treebank collection with the best language coverage at present. Furthermore, all UD Treebank data is fully open and accessible to everyone. These features made UD one of the most popular data resources for empirical linguistics studies.

For the current paper, we choose to use the Surface-syntactic Universal Dependencies (SUD) treebanks. SUD is an annotation scheme for syntactic dependency treebanks, near isomorphic to UD (Gerdes et al., 2018). Contrary to UD, it is based on syntactic criteria (favouring functional heads) and the relations are defined on distributional and functional bases. As a result, SUD corrects the artificially long dependency distances of UD into a more standard syntactic analysis based on distributional criteria (Osborne & Gerdes, 2019; Yan & Liu, 2019). For a study on dependency distances, it is thus important to use SUD treebanks instead of UD, so as not to measure idiosyncrasies of the dependency scheme. Meanwhile, just as UD, SUD treebanks are also fully open which allows easy replication and validation of the experiments.

Specifically, the material we use for the present paper is the Indo-European language datasets of Parallel Universal Dependency (PUD) treebank in SUD (transformed from UD 2.3). PUD is a parallel treebank created for the CoNLL 2017 shared task on multilingual parsing (Zeman et al., 2017). It covers a wide range of languages and includes 10 different Indo-European language datasets, namely, Czech, English, French, German, Hindi, Italian, Portuguese, Russian, Spanish, and Swedish. For each language dataset, there are 1000 aligned sentences that are randomly taken from the news and from Wikipedia.[2] Since PUD is a parallel treebank with good language coverage, it is a suitable data resource for the present study.

## 3. Method

We first compute the dependency distance for every single dependency in the treebank except the root relation. The dependency distance is computed as the absolute difference between the word ID and its head's word ID. For instance, in Figure 1, there are 4 dependencies. We would take three of them

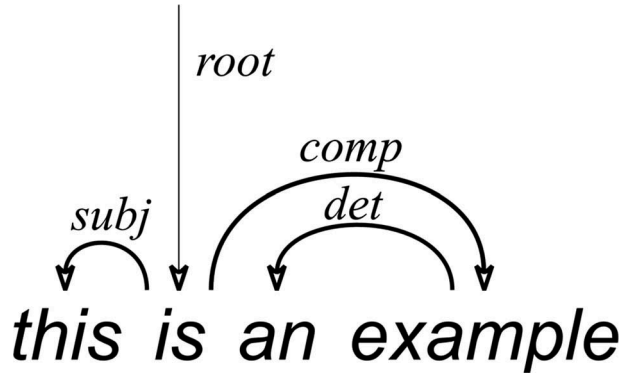


**Figure 1.** Example dependency tree in SUD analysis.

into account, leaving aside the root dependency. The dependency distances of these three dependencies are: *abs (1–2) = 1* (for *subj), abs (4–2) = 2* (for *comp*), and *abs (3–4) = 1* (for *det*).

After computing all the dependency distances of the treebank, we then count the frequencies of each distance, i.e. we count how many dependencies with distance 1 occurred in the treebank, how many dependencies with distance 2 occurred, and so on. We then try to formulate the relation into a function. We will fit different functions to see which one describes the empirical data best. In other words, we try to see whether our data can be fitted by functions.

For each language dataset, we also introduce two random baselines to see whether we can observe similar phenomenon in random dependency trees. There are different ways to make random dependency trees. What we applied in this study is a simple random algorithm that keeps the links and labels (such as subject, object, and etc.) between words but just randomly reorder the linear positions of words for each sentence (Courtin & Yan, 2019). By doing so, we can get random trees that share the same tree structure with the original tree but have different linear distances between words. For the first random baseline, we generate a random dataset without any constraints and then repeat the analysis for distances and frequencies. As for the second baseline, we first generate a *projective* (Lecerf, 1960) random dataset which means we randomly reorder the words in a way that keeps the sentence's dependencies projective (non-crossing). Projectivity is a key feature of dependency trees that holds true for most dependencies in natural languages with well-known exceptions such as *wh*-movement in English. We add projectivity as a constraint to our random trees in order to produce a more controlled baseline. We then repeat the same analysis for this projective random dataset. We expect the projectivity random dataset (PRD) to have greater similarity with the original dataset than the purely random dataset (RD).

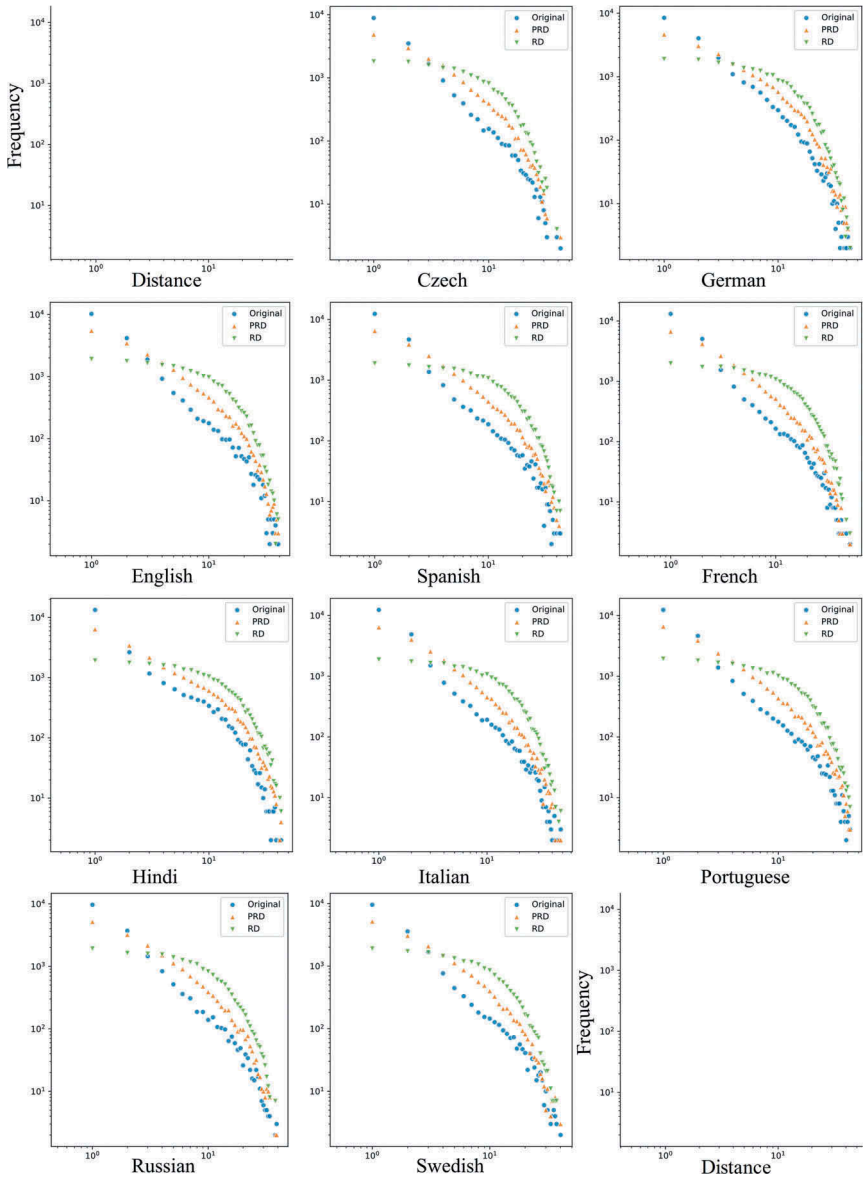


**Figure 2.** Scatter plot of dependency distance and frequency of 10 Indo-European language datasets.

## 4. Results and Discussion

Figure 2 presents on a logarithmic scale the relationship between dependency distance and frequency for the 10 Indo-European language datasets in PUD as well as their baselines.

For a more accurate analysis, we then tried to fit four functions, namely quadratic, exponent, logarithm, and power-law functions to the origianl datasets, RD, and PRD. These functions are commonly seen in empirical studies. Especially, exponent and power-law functions are frequently applied in linguistics (Baixeries et al., 2013; Mačutek et al., 2017; Wimmer & Altmann, 1999; Chen & Gerdes, 2019). The corresponding specific equations for the four functions are

$$y = ax^2 + bx + c \tag{1}$$

$$y = ab^x \tag{2}$$

$$y = aln(x) + b \tag{3}$$

$$y = ax^{-b} \tag{4}$$

With $y$ being the observed frequency, $x$ being the observed distance, $a$, $b$ and $c$ are empirical parameters.

Although there are different ways of measuring the goodness-of-fit (Mačutek & Wimmer, 2013), we choose to use the most common Pearson chi-square goodness-of-fit test to evaluate the fitting results in this study. The formula of the test is defined as

$$R^2 = \sum_{i=1}^{n} \frac{(f_i - NP_i)^2}{NP_i} \tag{5}$$

with $f_i$ being the observed frequency of the value $i$, $P_i$ being the expected probability of the value $i$, $n$ being the number of different data values and $N$ being the sample size. The obtained empirical parameters and $R^2$ are presented in Table 1.[3]

The $R^2$ results of original datasets show that quadratic and logarithm models do not provide satisfying results. The observed language samples can indeed be formulated as a power-law function. However, it seems that an exponent regression also fits to the data well. We conduct a Pair T-test for mean to the $R^2$ values of these two models. The result show that there is no significant difference between two groups ($p = 0.28, > 0.01$), which means that we cannot decide which model is better for our samples. This is a very common issue in quantitative linguistic studies (Baixeries et al., 2013). In many situations, both exponent and power-law models can describe the data fluctuation reasonably well. Our provided solution here is to introduce baselines for comparison. Random treebanks (or random baselines) are less human-like comparing to the original datasets. If one model fits the random treebanks as well as the original datasets, it means the model may represent something very general rather than human language specific features. In reverse, if one model

**Table 1.** Empirical parameters and $R^2$ of four models (values rounded to 2 decimal places).

| | Quadratic | | | | Exponent | | | Logarithm | | | Power Law | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Language | a | b | c | $R^2$ | a | b | $R^2$ | a | b | $R^2$ | a | b | $R^2$ |
| Czech | 4.76 | −260.48 | 3102.75 | 0.40 | 20353.74 | 0.43 | 0.99 | −1231.01 | 3846.22 | 0.53 | 8992.73 | 1.57 | 1.00 |
| English | 3.10 | −206.00 | 2963.43 | 0.34 | 23652.67 | 0.43 | 0.99 | −1261.96 | 4100.77 | 0.49 | 10384.65 | 1.58 | 0.99 |
| French | 3.94 | −256.60 | 3601.92 | 0.31 | 34788.68 | 0.38 | 1.00 | −1467.16 | 4823.51 | 0.44 | 13278.28 | 1.72 | 0.99 |
| German | 3.40 | −215.21 | 2983.36 | 0.44 | 16282.84 | 0.51 | 0.99 | −1159.07 | 3812.98 | 0.57 | 8650.27 | 1.37 | 0.99 |
| Hindi | 4.21 | −259.27 | 3491.08 | 0.29 | 58861.70 | 0.23 | 0.99 | −1443.33 | 4660.11 | 0.41 | 13294.55 | 2.14 | 1.00 |
| Italian | 2.89 | −203.84 | 3129.49 | 0.29 | 32004.28 | 0.39 | 1.00 | −1390.57 | 4578.25 | 0.44 | 12514.00 | 1.70 | 0.99 |
| Portuguese | 2.99 | −208.77 | 3151.19 | 0.29 | 33465.11 | 0.37 | 1.00 | −1387.76 | 4550.21 | 0.44 | 12540.07 | 1.74 | 0.99 |
| Russian | 4.61 | −259.49 | 3170.71 | 0.38 | 23982.57 | 0.40 | 1.00 | −1228.13 | 3916.42 | 0.49 | 9778.89 | 1.65 | 0.99 |
| Spanish | 3.01 | −209.76 | 3157.17 | 0.29 | 33230.04 | 0.37 | 1.00 | −1380.72 | 4528.99 | 0.43 | 12494.79 | 1.73 | 0.99 |
| Swedish | 3.83 | −228.09 | 2949.29 | 0.35 | 23368.23 | 0.41 | 1.00 | −1245.40 | 3923.03 | 0.49 | 9658.14 | 1.64 | 1.00 |

fits the original datasets better than the random treebanks, then the model catches up some human language features which do not exist in the random datasets. Therefore, we can cautiously claim that such a model is better suited to represent the relation between dependency distance and frequency.

For the two random treebank variations, RD and PRD, we replicate the same computation for the frequency and dependency distance, see Appendix. Similarly, we fit both exponent and power-law models to the data points, see Table 2 for results.

There are several interesting points in the obtained results. First, for all 10 languages, the exponent model fits to RD well and the power-law model does not. This means that we could distinguish an original natural language treebank from its corresponding random treebank according to how well the power-law model fits to the relationship of dependency distances and their frequencies. In other words, the power-law model is a better option to represent the relation between dependency distance and frequency in natural language, although the exponential model could also describe the original datasets well.

Second, when we add the projectivity restriction, the fitting results of PRD, a controlled baseline, seems more 'human language' like. Similar to the original datasets, both exponent and power-law models fit to PRD well and there are no significant differences between the $R^2$ values of these two models (Pair T-test for mean, $p = 0.06, > 0.01$). However, if we compare the empirical parameters a, b, and $R^2$ of PRD to their corresponding values of the original datasets, we can find differences. The result of the Pair T-test for mean shows that the differences between the $R^2$ of PRD and the original datasets for exponent model fitting are significant ($p = 0, < 0.01$). There are also significant differences between the empirical parameters *a* and *b* of PRD and the original datasets (both *a* and *b*, $p = 0, < 0.01$). The same holds for the $R^2$ power-law fitting results (for all *a, b*, and $R^2$, $p = 0, < 0.01$). This means that we can distinguish an original treebank from its corresponding projective random treebank according to the exact values of the parameters a, b or $R^2$ of the exponent model or the power-law model.

## 5. Conclusion

Our results are coherent with our hypothesis that there is indeed a relation between dependency distance and frequency. However, it is not very clear which model, namely exponent or power-law model, is better suited for describing the observed datasets.

The results in Table 1 show that for the original nature datasets, this relation can be formulated as either a power-law function or an exponent function. There are no significant differences between the goodness-of-fit results, $R^2$, of these two models for 10 different Indo-European languages.

**Table 2.** Empirical parameters and $R^2$ results of RD and PRD.

| | | Exponent | | | Power-Law | | |
|---|---|---|---|---|---|---|---|
| Language | | a | b | $R^2$ | a | b | $R^2$ |
| Czech | RD | 2279.81 | 0.89 | 0.99 | 2506.20 | 0.67 | 0.76 |
| | PRD | 6287.99 | 0.72 | 0.99 | 5095.23 | 1.04 | 0.96 |
| | Original | 20353.74 | 0.43 | 0.99 | 8992.73 | 1.57 | 1.00 |
| English | RD | 2294.10 | 0.90 | 0.99 | 2605.29 | 0.65 | 0.74 |
| | PRD | 7248.28 | 0.71 | 0.98 | 5831.16 | 1.05 | 0.97 |
| | Original | 23652.67 | 0.43 | 0.99 | 10384.65 | 1.58 | 0.99 |
| French | RD | 2251.75 | 0.92 | 0.99 | 2664.68 | 0.62 | 0.73 |
| | PRD | 9500.36 | 0.68 | 0.99 | 7164.91 | 1.09 | 0.97 |
| | Original | 34788.68 | 0.38 | 1.00 | 13278.28 | 1.72 | 0.99 |
| German | RD | 2248.37 | 0.91 | 0.99 | 2560.59 | 0.64 | 0.74 |
| | PRD | 5689.41 | 0.76 | 0.98 | 5119.34 | 0.97 | 0.96 |
| | Original | 16282.84 | 0.51 | 0.99 | 8650.27 | 1.37 | 0.99 |
| Hindi | RD | 2243.07 | 0.92 | 0.99 | 2614.80 | 0.61 | 0.73 |
| | PRD | 8793.52 | 0.67 | 0.95 | 6570.67 | 1.09 | 0.99 |
| | Original | 58861.70 | 0.23 | 0.99 | 13294.55 | 2.14 | 1.00 |
| Italian | RD | 2227.01 | 0.92 | 0.99 | 2615.84 | 0.63 | 0.73 |
| | PRD | 8820.33 | 0.69 | 0.99 | 6777.78 | 1.08 | 0.97 |
| | Original | 32004.28 | 0.39 | 1.00 | 12514.00 | 1.70 | 0.99 |
| Portuguese | RD | 2253.94 | 0.91 | 0.99 | 2654.80 | 0.64 | 0.74 |
| | PRD | 9031.46 | 0.68 | 0.99 | 6819.61 | 1.09 | 0.97 |
| | Original | 33465.11 | 0.37 | 1.00 | 12540.07 | 1.74 | 0.99 |
| Russian | RD | 2265.73 | 0.90 | 0.99 | 2531.92 | 0.66 | 0.77 |
| | PRD | 6856.70 | 0.70 | 0.99 | 5426.07 | 1.06 | 0.97 |
| | Original | 23982.57 | 0.40 | 1.00 | 9778.89 | 1.65 | 0.99 |
| Spanish | RD | 2264.42 | 0.91 | 0.99 | 2649.77 | 0.64 | 0.74 |
| | PRD | 9053.56 | 0.68 | 0.98 | 6813.64 | 1.10 | 0.97 |
| | Original | 33230.04 | 0.37 | 1.00 | 12494.79 | 1.73 | 0.99 |
| Swedish | RD | 2265.03 | 0.89 | 0.99 | 2496.53 | 0.66 | 0.75 |
| | PRD | 6919.60 | 0.70 | 0.99 | 5441.11 | 1.07 | 0.97 |
| | Original | 23368.23 | 0.41 | 1.00 | 9658.14 | 1.64 | 1.00 |

For figuring out which model is better for representing the relation, we introduce two random baselines. By randomly reordering the words in a sentence, while preserving the words' dependencies, we generate random datasets for corresponding language treebanks: PRD with and RD without the projectivity restriction, in which PRD possesses a more 'natural' structure reproducing more closely the rarity of non-projective relations. We replicate the exponent and power-law fitting analysis on these twenty random treebanks and compare the results with the obtained results of original datasets. We found that we can distinguish the original language treebanks from their corresponding RD by looking at the results of goodness-of-fit. To be more specific, the power-law model fits to the original datasets well while it does not fit to RD. Meanwhile, both the original datasets and the RD fit exponent models well. In other words, a power-law model can be used to tell apart a natural language treebanks from their corresponding RD and it performs better than the exponent model in this specific task.

As for the analysis of the more 'natural' baseline PRD, the situation is different from that of RD. We can no longer distinguish the original language treebanks from their corresponding PRD by simply looking at how good the fitting results are. The $R^2$ values from Table 2 indicate that both models can describe the original datasets and the PRD well. However, we can find significant differences between these two groups of datasets concerning the values of parameters and the exact $R^2$ values both from power-law and exponent fitting results. Meanwhile, we cannot decide which model is better suited for describing the original language treebanks as in the previous experiment, since these two models have a similar performance in this task. It is important to mention that, our result here is different from the pilot study of Chen and Gerdes (2019). Their analysis of English shows that the power-law model fits to the original language datasets better than the PRD. The differences might be caused by different fitting methods and tools. Chen and Gerdes (2019) transformed the functions into linear functions and carried out the fitting analysis with R. We do not make such transformations in the current study and we directly conduct the regressions with NLREG. We believe that this might by the reason why we obtain better fitting results for power-law model when we analyse PRD.

In summary, we would like to cautiously draw the conclusion here that the power-law model is possibly a better choice for representing the relation between dependency distance and frequency for Indo-European languages. This conclusion is coherent with that of the English pilot study (Chen & Gerdes, 2019). In the near future, we would like to strengthen the current study in two directions. First, we would like to replicate this study for non-Indo-European languages and see whether we can make further generalizations. Second, we would like to introduce more controls, such as dependency types, into the analysis.

## Notes

1. Hudson's original measures takes two adjacent words to have distance zero. We prefer the alternative definition where $x = y \Leftrightarrow d(x,y) = 0$, i.e. a word has distance zero with itself, making the measure a metric in the mathematical sense.
2. For more details,https://github.com/UniversalDependencies/UD_English-PUD.
3. All parameter values in the models were obtained by *NLREG* (version 6.3). The same below.

## Disclosure Statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the National Social Science Fund of China [2018CYY031].

## References

Baixeries, J., Elvevåg, B., & Ferrer-i-Cancho, R. (2013). The evolution of the exponent of Zipf's law in language ontogeny. *PloS One*, *8*(3), e53227. https://doi.org/10.1371/journal.pone.0053227

Buchholz, S., & Marsi, E. (2006, June). Conll-x shared task on multilingual dependency parsing. In *Proceedings of the tenth conference on computational natural language learning (CoNLL-X)*, New York City (pp. 149–164).

Chen, X., & Gerdes, K. (2017, September). Classifying languages by dependency structure: Typologies of delexicalized universal dependency treebanks. In *Proceedings of the fourth international conference on dependency linguistics (Depling 2017)*, Pisa. Linköping University Electronic Press.

Chen, X., & Gerdes, K. (2018). How do universal dependencies distinguish language groups? In J. Jiang, & H. Liu, (Eds.), *Quantitative Analysis of Dependency Structures* (pp. 277–294). Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110573565-014

Chen, X., & Gerdes, K. (2019, August). The relationship between dependency distance and frequency. In *Proceedings of the first workshop on quantitative syntax (Quasy, SyntaxFest 2019)*, Paris. Association for Computational Linguistics.

Courtin, M., & Yan, C. (2019, August). What can we learn from natural and artificial dependency trees. In *Proceedings of the first workshop on quantitative syntax (Quasy, SyntaxFest 2019)*, Paris. Association for Computational Linguistics.

De Marneffe, M.-C., & Nivre, J. (2019). Dependency grammar. *Annual Review of Linguistics*, *5*(1), 197–218. https://doi.org/10.1146/annurev-linguistics-011718-011842

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336–10341. https://doi.org/10.1073/pnas.1502134112

Gerdes, K., Guillaume, B., Kahane, S., & Perrier, G. (2018, November). SUD or surface-syntactic universal dependencies: An annotation scheme near-isomorphic to UD. In *Universal dependencies workshop 2018*. Brussels.

Gerdes, K., Guillaume, B., Kahane, S., & Perrier, G. (2019, August). Improving Surface-syntactic Universal Dependencies (SUD): Surface-syntactic relations and deep syntactic features. In *Proceedings of the 18th international workshop on treebanks and linguistic theories (TLT, SyntaxFest 2019)*, Paris. Association for Computational Linguistics.

Hudson, R. (1995). *Measuring syntactic difficulty* (Draft of manuscript). http://dickhudson.com/wp-content/uploads/2013/07/Difficulty.pdf

Lecerf, Y. (1960). Programme des conflits, modèle des conflits. *Bulletin bimestriel de l'ATALA*, *1*(4), 11–18.

Liu, H. (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, *9*(2), 159–191. https://doi.org/10.17791/jcs.2008.9.2.159

Liu, H. (2010). Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua*, *120*(6), 1567–1578. https://doi.org/10.1016/j.lingua.2009.10.001

Liu, H., Xu, C., & Liang, J. (2017). Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, *21*, 171–193. https://doi.org/10.1016/j.plrev.2017.03.002

Mačutek, J., Radek, Č., & Jiří, M. (2017). Menzerath-altmann law in syntactic dependency structure. In S. Montemagni & J. Nivre (eds.). *Proceedings of the fourth international conference on dependency linguistics (Depling 2017). Linköping electronic conference proceedings, No. 139* (pp. 100–107). Pisa, Italy.

Mačutek, J., & Wimmer, G. (2013). Evaluating goodness-of-fit of discrete distribution models in quantitative linguistics. *Journal of Quantitative Linguistics*, *20*(3), 227–240. https://doi.org/10.1080/09296174.2013.799912

Nivre, J., Abrams, M., & Agić, Ž. 2019. Universal dependencies 2.4. *LINDAT/CLARIN digital library* at the Institute of Formal and Applied Linguistics (UFAL), Faculty of Mathematics and Physics, Charles University.

Osborne, T., & Gerdes, K. (2019). The status of function words in dependency grammar: A critique of Universal Dependencies (UD). *Glossa: A Journal of General Linguistics*, *4*(1), 17. 1–28. https://doi.org/10.5334/gjgl.537

Temperley, D. (2007). Minimization of dependency length in written English. *Cognition*, *105*(2), 300–333. https://doi.org/10.1016/j.cognition.2006.09.011

Wimmer, G., & Altmann, G. (1999). *Thesaurus of univariate discrete probability distributions*. Stamm.

Yan, J., & Liu, H. (2019, August). Which annotation scheme is more expedient to measure syntactic difficulty and cognitive demand? In *Proceedings of the first workshop on quantitative syntax (Quasy, SyntaxFest 2019)*, Paris. Association for Computational Linguistics.

Zeman, D., Popel, M., Straka, M., Hajic, J., Nivre, J., Ginter, F., Luotolahti, J., Pyysalo, S., Petrov, S., Potthast, M., Tyers, F., Badmaeva, E., Gokirmak, M., Nedoluzhko, A., Cinkova, S., Hajic, J. jr., Hlavacova, J., Kettnerová, V., Uresova, Z., Kanerva, J., ... Li, J. (2017). CoNLL 2017 shared task: Multilingual parsing from raw text to universal dependencies. In *CoNLL 2017*, Vancouver, Canada.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley.

## Appendix

The table shows the complete dependency distance frequency data from the SUD version of 10 Indo-European PUD treebanks.

| Language | Distance | Original | PRD | RD | Language | Distance | Original | PRD | RD |
|---|---|---|---|---|---|---|---|---|---|
| **Czech** | 1 | 8882 | 4732 | 1877 | **English** | 1 | 10236 | 5451 | 1893 |
| | 2 | 3505 | 3079 | 1714 | | 2 | 4157 | 3422 | 1800 |
| | 3 | 1673 | 2027 | 1621 | | 3 | 1887 | 2311 | 1610 |
| | 4 | 912 | 1493 | 1500 | | 4 | 924 | 1653 | 1566 |
| | 5 | 529 | 1162 | 1389 | | 5 | 544 | 1261 | 1470 |
| | 6 | 395 | 867 | 1205 | | 6 | 412 | 970 | 1350 |
| | 7 | 260 | 719 | 1114 | | 7 | 292 | 797 | 1245 |
| | 8 | 221 | 573 | 936 | | 8 | 209 | 679 | 1071 |
| | 9 | 147 | 470 | 921 | | 9 | 192 | 545 | 1051 |
| | 10 | 156 | 374 | 775 | | 10 | 177 | 435 | 946 |
| | 11 | 137 | 315 | 719 | | 11 | 139 | 396 | 830 |
| | 12 | 112 | 270 | 590 | | 12 | 133 | 282 | 740 |
| | 13 | 90 | 236 | 498 | | 13 | 98 | 263 | 640 |
| | 14 | 86 | 189 | 471 | | 14 | 96 | 269 | 577 |
| | 15 | 85 | 171 | 377 | | 15 | 97 | 189 | 524 |
| | 16 | 59 | 142 | 301 | | 16 | 72 | 206 | 461 |
| | 17 | 59 | 132 | 258 | | 17 | 52 | 143 | 387 |
| | 18 | 50 | 83 | 214 | | 18 | 71 | 139 | 305 |
| | 19 | 34 | 95 | 200 | | 19 | 52 | 108 | 291 |
| | 20 | 31 | 69 | 164 | | 20 | 47 | 90 | 259 |
| | 21 | 29 | 73 | 133 | | 21 | 43 | 87 | 195 |
| | 22 | 25 | 59 | 130 | | 22 | 50 | 81 | 174 |
| | 23 | 24 | 44 | 108 | | 23 | 27 | 67 | 128 |
| | 24 | 22 | 46 | 87 | | 24 | 18 | 50 | 108 |
| | 25 | 13 | 31 | 49 | | 25 | 26 | 48 | 88 |
| | 26 | 17 | 21 | 66 | | 26 | 24 | 47 | 79 |
| | 27 | 6 | 28 | 40 | | 27 | 22 | 43 | 74 |
| | 28 | 13 | 28 | 32 | | 28 | 11 | 27 | 55 |
| | 29 | 11 | 11 | 21 | | 29 | 18 | 26 | 49 |
| | 30 | 8 | 15 | 20 | | 30 | 12 | 18 | 46 |
| | 31 | 5 | 9 | 16 | | 31 | 3 | 13 | 27 |
| | 32 | 3 | 12 | 15 | | 32 | 5 | 11 | 27 |
| | 33 | 1 | 7 | 15 | | 33 | 2 | 9 | 25 |
| | 34 | 1 | 7 | 8 | | 34 | 5 | 12 | 15 |
| | 35 | 0 | 7 | 8 | | 35 | 3 | 5 | 14 |
| | 36 | 1 | 3 | 5 | | 36 | 5 | 6 | 11 |
| | 37 | 1 | 4 | 3 | | 37 | 4 | 3 | 8 |
| | 38 | 0 | 0 | 3 | | 38 | 2 | 4 | 8 |
| | 39 | 3 | 2 | 1 | | 39 | 2 | 1 | 5 |
| | 40 | 1 | 3 | 1 | | 40 | 1 | 2 | 3 |
| | 41 | 0 | 0 | 1 | | 41 | 1 | 1 | 3 |

(*Continued*)

(Continued).

| Language | Distance | Original | PRD | RD | Language | Distance | Original | PRD | RD |
|---|---|---|---|---|---|---|---|---|---|
| | 42 | 2 | 1 | 0 | | 42 | 0 | 1 | 3 |
| | 43 | 0 | 0 | 1 | | 43 | 1 | 0 | 3 |
| | 44 | 0 | 1 | 2 | | 44 | 0 | 0 | 2 |
| | 45 | 1 | 0 | 0 | | 45 | 1 | 0 | 1 |
| | 46 | 0 | 0 | 1 | | 46 | 0 | 0 | 2 |
| | | | | | | 47 | 0 | 1 | 0 |
| | | | | | | 48 | 0 | 2 | 2 |
| | | | | | | 49 | 1 | 0 | 2 |
| | | | | | | 50 | 0 | 0 | 2 |
| | | | | | | 51 | 0 | 1 | 1 |
| | | | | | | 53 | 1 | 0 | 0 |
| | | | | | | 55 | 0 | 1 | 0 |
| | | | | | | 56 | 1 | 0 | 0 |
| **Language** | **Distance** | **Original** | **PRD** | **RD** | **Language** | **Distance** | **Original** | **PRD** | **RD** |
| **French** | 1 | 13120 | 6722 | 1889 | **German** | 1 | 8461 | 4726 | 1857 |
| | 2 | 5044 | 4224 | 1801 | | 2 | 4029 | 3065 | 1768 |
| | 3 | 1547 | 2669 | 1656 | | 3 | 1997 | 2197 | 1657 |
| | 4 | 817 | 1821 | 1681 | | 4 | 1096 | 1711 | 1566 |
| | 5 | 500 | 1303 | 1440 | | 5 | 813 | 1337 | 1393 |
| | 6 | 403 | 1069 | 1421 | | 6 | 692 | 1051 | 1297 |
| | 7 | 311 | 855 | 1326 | | 7 | 562 | 904 | 1232 |
| | 8 | 241 | 673 | 1209 | | 8 | 431 | 767 | 1115 |
| | 9 | 210 | 532 | 1198 | | 9 | 333 | 644 | 1023 |
| | 10 | 164 | 464 | 1076 | | 10 | 297 | 519 | 938 |
| | 11 | 133 | 418 | 970 | | 11 | 230 | 470 | 772 |
| | 12 | 134 | 348 | 858 | | 12 | 202 | 411 | 698 |
| | 13 | 126 | 355 | 809 | | 13 | 173 | 337 | 735 |
| | 14 | 110 | 282 | 736 | | 14 | 162 | 335 | 600 |
| | 15 | 102 | 243 | 656 | | 15 | 123 | 271 | 517 |
| | 16 | 85 | 192 | 566 | | 16 | 95 | 230 | 488 |
| | 17 | 80 | 178 | 522 | | 17 | 92 | 200 | 422 |
| | 18 | 87 | 179 | 473 | | 18 | 89 | 174 | 322 |
| | 19 | 65 | 150 | 440 | | 19 | 66 | 160 | 304 |
| | 20 | 54 | 132 | 391 | | 20 | 52 | 135 | 255 |
| | 21 | 45 | 118 | 379 | | 21 | 42 | 102 | 232 |
| | 22 | 37 | 106 | 277 | | 22 | 33 | 97 | 188 |
| | 23 | 43 | 98 | 273 | | 23 | 42 | 66 | 162 |
| | 24 | 30 | 85 | 208 | | 24 | 29 | 70 | 142 |
| | 25 | 27 | 54 | 215 | | 25 | 23 | 52 | 117 |
| | 26 | 26 | 59 | 174 | | 26 | 26 | 41 | 99 |
| | 27 | 25 | 55 | 157 | | 27 | 30 | 49 | 93 |
| | 28 | 19 | 54 | 146 | | 28 | 20 | 34 | 54 |
| | 29 | 30 | 48 | 102 | | 29 | 19 | 32 | 54 |
| | 30 | 17 | 29 | 81 | | 30 | 10 | 31 | 50 |

(*Continued*)

(Continued).

| Language | Distance | Original | PRD | RD | Language | Distance | Original | PRD | RD |
|---|---|---|---|---|---|---|---|---|---|
| | 31 | 8 | 28 | 96 | | 31 | 11 | 19 | 34 |
| | 32 | 16 | 27 | 79 | | 32 | 4 | 23 | 34 |
| | 33 | 9 | 18 | 71 | | 33 | 10 | 14 | 16 |
| | 34 | 12 | 26 | 58 | | 34 | 5 | 8 | 15 |
| | 35 | 8 | 21 | 46 | | 35 | 2 | 4 | 14 |
| | 36 | 8 | 14 | 52 | | 36 | 3 | 6 | 6 |
| | 37 | 8 | 9 | 35 | | 37 | 5 | 8 | 13 |
| | 38 | 5 | 11 | 28 | | 38 | 2 | 4 | 12 |
| | 39 | 5 | 12 | 21 | | 39 | 2 | 4 | 5 |
| | 40 | 3 | 9 | 24 | | 40 | 2 | 1 | 3 |
| | 41 | 3 | 11 | 20 | | 41 | 3 | 2 | 5 |
| | 42 | 5 | 4 | 9 | | 42 | 0 | 2 | 5 |
| | 43 | 3 | 6 | 11 | | 43 | 2 | 2 | 5 |
| | 44 | 0 | 3 | 11 | | 44 | 1 | 1 | 4 |
| | 45 | 0 | 2 | 6 | | 45 | 1 | 1 | 0 |
| | 46 | 1 | 0 | 11 | | 46 | 0 | 3 | 1 |
| | 47 | 3 | 5 | 7 | | 47 | 1 | 0 | 1 |
| | 48 | 0 | 2 | 4 | | 49 | 0 | 3 | 0 |
| | 49 | 1 | 1 | 1 | | 50 | 1 | 0 | 1 |
| | 50 | 1 | 1 | 1 | | 51 | 1 | 1 | 1 |
| | 51 | 2 | 3 | 4 | | 53 | 0 | 1 | 0 |
| | 52 | 0 | 2 | 1 | | | | | |
| | 53 | 1 | 3 | 3 | | | | | |
| | 54 | 0 | 1 | 1 | | | | | |
| | 56 | 0 | 0 | 2 | | | | | |
| | 57 | 0 | 0 | 2 | | | | | |
| **Language** | **Distance** | **Original** | **PRD** | **RD** | **Language** | **Distance** | **Original** | **PRD** | **RD** |
| **Hindi** | 1 | 13350 | 6395 | 1869 | **Italian** | 1 | 12351 | 6336 | 1838 |
| | 2 | 2646 | 3390 | 1755 | | 2 | 4889 | 4032 | 1760 |
| | 3 | 1172 | 2079 | 1695 | | 3 | 1510 | 2556 | 1691 |
| | 4 | 814 | 1582 | 1583 | | 4 | 782 | 1756 | 1592 |
| | 5 | 642 | 1171 | 1533 | | 5 | 517 | 1294 | 1489 |
| | 6 | 515 | 1034 | 1393 | | 6 | 386 | 1069 | 1459 |
| | 7 | 469 | 913 | 1296 | | 7 | 325 | 820 | 1223 |
| | 8 | 426 | 712 | 1142 | | 8 | 236 | 658 | 1200 |
| | 9 | 401 | 678 | 1124 | | 9 | 188 | 531 | 1028 |
| | 10 | 340 | 584 | 1056 | | 10 | 192 | 506 | 1003 |
| | 11 | 269 | 560 | 934 | | 11 | 159 | 398 | 915 |
| | 12 | 297 | 485 | 889 | | 12 | 142 | 355 | 901 |
| | 13 | 207 | 422 | 758 | | 13 | 133 | 335 | 800 |
| | 14 | 201 | 371 | 707 | | 14 | 107 | 264 | 698 |
| | 15 | 156 | 316 | 604 | | 15 | 86 | 223 | 605 |
| | 16 | 145 | 303 | 609 | | 16 | 78 | 214 | 566 |
| | 17 | 122 | 262 | 509 | | 17 | 84 | 172 | 536 |

(*Continued*)

(Continued).

| Language | Distance | Original | PRD | RD | Language | Distance | Original | PRD | RD |
|---|---|---|---|---|---|---|---|---|---|
| | 18 | 93 | 217 | 474 | | 18 | 64 | 161 | 439 |
| | 19 | 83 | 194 | 374 | | 19 | 60 | 139 | 419 |
| | 20 | 77 | 160 | 352 | | 20 | 59 | 128 | 351 |
| | 21 | 78 | 156 | 304 | | 21 | 39 | 108 | 308 |
| | 22 | 44 | 126 | 279 | | 22 | 39 | 93 | 288 |
| | 23 | 62 | 104 | 250 | | 23 | 29 | 71 | 222 |
| | 24 | 34 | 105 | 194 | | 24 | 34 | 72 | 196 |
| | 25 | 29 | 75 | 182 | | 25 | 26 | 60 | 168 |
| | 26 | 26 | 66 | 150 | | 26 | 30 | 58 | 138 |
| | 27 | 17 | 53 | 122 | | 27 | 32 | 47 | 137 |
| | 28 | 26 | 59 | 100 | | 28 | 26 | 37 | 113 |
| | 29 | 15 | 33 | 104 | | 29 | 20 | 31 | 94 |
| | 30 | 10 | 40 | 81 | | 30 | 19 | 37 | 87 |
| | 31 | 14 | 31 | 67 | | 31 | 13 | 23 | 69 |
| | 32 | 6 | 22 | 60 | | 32 | 9 | 18 | 63 |
| | 33 | 6 | 25 | 56 | | 33 | 7 | 21 | 54 |
| | 34 | 6 | 18 | 36 | | 34 | 15 | 14 | 40 |
| | 35 | 2 | 10 | 38 | | 35 | 7 | 20 | 43 |
| | 36 | 6 | 10 | 35 | | 36 | 4 | 10 | 32 |
| | 37 | 6 | 6 | 29 | | 37 | 6 | 6 | 36 |
| | 38 | 7 | 8 | 17 | | 38 | 4 | 5 | 21 |
| | 39 | 2 | 14 | 15 | | 39 | 3 | 6 | 18 |
| | 40 | 0 | 6 | 10 | | 40 | 2 | 4 | 21 |
| | 41 | 1 | 7 | 8 | | 41 | 1 | 3 | 12 |
| | 42 | 2 | 4 | 8 | | 42 | 5 | 7 | 9 |
| | 43 | 2 | 5 | 7 | | 43 | 2 | 4 | 11 |
| | 44 | 0 | 2 | 4 | | 44 | 1 | 4 | 4 |
| | 45 | 0 | 5 | 5 | | 45 | 0 | 3 | 5 |
| | 46 | 1 | 1 | 2 | | 46 | 2 | 2 | 6 |
| | 47 | 0 | 4 | 2 | | 47 | 1 | 2 | 7 |
| | 48 | 1 | 1 | 3 | | 48 | 3 | 0 | 2 |
| | 49 | 0 | 2 | 0 | | 49 | 0 | 2 | 0 |
| | 50 | 0 | 1 | 2 | | 50 | 0 | 3 | 0 |
| | 51 | 1 | 0 | 2 | | 51 | 0 | 1 | 2 |
| | 52 | 0 | 2 | 0 | | 52 | 0 | 3 | 2 |
| | 58 | 0 | 0 | 1 | | 53 | 1 | 0 | 1 |
| | | | | | | 54 | 0 | 1 | 1 |
| | | | | | | 55 | 0 | 1 | 0 |
| | | | | | | 56 | 0 | 2 | 1 |
| | | | | | | 57 | 0 | 0 | 1 |
| | | | | | | 58 | 0 | 0 | 2 |
| | | | | | | 61 | 0 | 1 | 1 |

(*Continued*)

(Continued).

| Language | Distance | Original | PRD | RD | Language | Distance | Original | PRD | RD |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 62 | 0 | | 1 |
| | | | | | | 63 | 1 | 1 | 0 |
| | | | | | | 65 | 0 | 1 | 0 |
| **Language** | **Distance** | **Original** | **PRD** | **RD** | **Language** | **Distance** | **Original** | **PRD** | **RD** |
| **Portuguese** | 1 | 12410 | 6419 | 1895 | **Russian** | 1 | 9673 | 5069 | 1949 |
| | 2 | 4634 | 3928 | 1782 | | 2 | 3707 | 3221 | 1671 |
| | 3 | 1396 | 2563 | 1664 | | 3 | 1450 | 2128 | 1526 |
| | 4 | 839 | 1760 | 1587 | | 4 | 835 | 1491 | 1591 |
| | 5 | 514 | 1270 | 1464 | | 5 | 516 | 1127 | 1372 |
| | 6 | 394 | 998 | 1334 | | 6 | 361 | 914 | 1243 |
| | 7 | 286 | 755 | 1307 | | 7 | 308 | 693 | 1173 |
| | 8 | 247 | 666 | 1254 | | 8 | 187 | 575 | 1006 |
| | 9 | 201 | 511 | 1090 | | 9 | 186 | 501 | 886 |
| | 10 | 178 | 478 | 1009 | | 10 | 139 | 411 | 853 |
| | 11 | 155 | 379 | 890 | | 11 | 153 | 336 | 697 |
| | 12 | 127 | 339 | 842 | | 12 | 107 | 289 | 649 |
| | 13 | 113 | 290 | 730 | | 13 | 103 | 230 | 563 |
| | 14 | 84 | 261 | 684 | | 14 | 98 | 198 | 454 |
| | 15 | 91 | 223 | 619 | | 15 | 64 | 184 | 422 |
| | 16 | 84 | 196 | 568 | | 16 | 75 | 165 | 370 |
| | 17 | 74 | 189 | 478 | | 17 | 59 | 135 | 339 |
| | 18 | 62 | 144 | 415 | | 18 | 46 | 112 | 247 |
| | 19 | 70 | 128 | 394 | | 19 | 49 | 94 | 215 |
| | 20 | 46 | 116 | 331 | | 20 | 26 | 74 | 196 |
| | 21 | 43 | 99 | 282 | | 21 | 39 | 70 | 163 |
| | 22 | 48 | 81 | 271 | | 22 | 34 | 50 | 123 |
| | 23 | 33 | 81 | 211 | | 23 | 22 | 60 | 123 |
| | 24 | 25 | 72 | 193 | | 24 | 16 | 36 | 113 |
| | 25 | 25 | 68 | 151 | | 25 | 15 | 46 | 74 |
| | 26 | 24 | 50 | 134 | | 26 | 22 | 30 | 64 |
| | 27 | 34 | 49 | 146 | | 27 | 17 | 25 | 52 |
| | 28 | 22 | 33 | 112 | | 28 | 11 | 23 | 52 |
| | 29 | 13 | 41 | 102 | | 29 | 7 | 13 | 35 |
| | 30 | 13 | 17 | 68 | | 30 | 6 | 11 | 38 |
| | 31 | 11 | 26 | 54 | | 31 | 5 | 8 | 25 |
| | 32 | 8 | 20 | 41 | | 32 | 5 | 10 | 17 |
| | 33 | 8 | 16 | 47 | | 33 | 4 | 7 | 15 |
| | 34 | 8 | 18 | 28 | | 34 | 4 | 6 | 15 |
| | 35 | 4 | 13 | 32 | | 35 | 1 | 7 | 7 |
| | 36 | 11 | 12 | 19 | | 36 | 1 | 5 | 4 |
| | 37 | 6 | 9 | 33 | | 37 | 1 | 2 | 5 |
| | 38 | 4 | 5 | 27 | | 38 | 2 | 0 | 4 |
| | 39 | 2 | 8 | 11 | | 39 | 3 | 1 | 5 |
| | 40 | 4 | 7 | 12 | | 40 | 1 | 1 | 0 |

(*Continued*)

(Continued).

| Language | Distance | Original | PRD | RD | Language | Distance | Original | PRD | RD |
|---|---|---|---|---|---|---|---|---|---|
| | 41 | 5 | 5 | 9 | | 41 | 1 | 1 | 2 |
| | 42 | 3 | 1 | 7 | | 42 | 1 | 2 | 0 |
| | 43 | 0 | 5 | 8 | | 43 | 1 | 1 | 0 |
| | 44 | 1 | 2 | 3 | | 44 | 0 | 0 | 2 |
| | 45 | 0 | 3 | 3 | | 46 | 1 | 0 | 1 |
| | 46 | 1 | 2 | 5 | | 47 | 0 | 0 | 1 |
| | 47 | 1 | 4 | 2 | | | | | |
| | 48 | 0 | 0 | 1 | | | | | |
| | 49 | 0 | 0 | 2 | | | | | |
| | 50 | 0 | 1 | 4 | | | | | |
| | 51 | 0 | 0 | 1 | | | | | |
| | 52 | 1 | 1 | 0 | | | | | |
| | 53 | 1 | 0 | 1 | | | | | |
| | 54 | 0 | 0 | 1 | | | | | |
| | 55 | 0 | 0 | 2 | | | | | |
| | 56 | 0 | 0 | 1 | | | | | |
| | 57 | 0 | 0 | 1 | | | | | |
| | 58 | 0 | 0 | 1 | | | | | |
| | 59 | 0 | 1 | 0 | | | | | |
| | 60 | 0 | 1 | 1 | | | | | |
| | 61 | 1 | 0 | 1 | | | | | |
| | 63 | 0 | 1 | 0 | | | | | |
| **Language** | **Distance** | **Original** | **PRD** | **RD** | **Language** | **Distance** | **Original** | **PRD** | **RD** |
| **Spanish** | 1 | 12358 | 6407 | 1927 | **Swedish** | 1 | 9555 | 5122 | 1821 |
| | 2 | 4673 | 3990 | 1722 | | 2 | 3587 | 3082 | 1768 |
| | 3 | 1369 | 2442 | 1648 | | 3 | 1679 | 2140 | 1608 |
| | 4 | 827 | 1725 | 1610 | | 4 | 767 | 1521 | 1488 |
| | 5 | 482 | 1296 | 1515 | | 5 | 446 | 1149 | 1399 |
| | 6 | 364 | 970 | 1392 | | 6 | 331 | 874 | 1240 |
| | 7 | 318 | 782 | 1344 | | 7 | 241 | 667 | 1138 |
| | 8 | 234 | 711 | 1195 | | 8 | 181 | 541 | 1040 |
| | 9 | 216 | 555 | 1050 | | 9 | 154 | 466 | 936 |
| | 10 | 188 | 448 | 974 | | 10 | 145 | 405 | 822 |
| | 11 | 144 | 384 | 934 | | 11 | 127 | 334 | 719 |
| | 12 | 124 | 317 | 825 | | 12 | 115 | 265 | 649 |
| | 13 | 109 | 258 | 783 | | 13 | 94 | 247 | 549 |
| | 14 | 105 | 272 | 651 | | 14 | 83 | 210 | 485 |
| | 15 | 93 | 210 | 607 | | 15 | 71 | 160 | 383 |
| | 16 | 75 | 190 | 559 | | 16 | 73 | 127 | 340 |
| | 17 | 70 | 162 | 463 | | 17 | 48 | 125 | 292 |
| | 18 | 58 | 151 | 399 | | 18 | 56 | 100 | 255 |
| | 19 | 56 | 123 | 365 | | 19 | 47 | 89 | 196 |
| | 20 | 58 | 113 | 333 | | 20 | 41 | 74 | 190 |
| | 21 | 35 | 116 | 318 | | 21 | 22 | 57 | 142 |

(*Continued*)

(Continued).

| Language | Distance | Original | PRD | RD | Language | Distance | Original | PRD | RD |
|---|---|---|---|---|---|---|---|---|---|
| | 22 | 40 | 85 | 250 | | 22 | 40 | 49 | 124 |
| | 23 | 38 | 81 | 212 | | 23 | 33 | 46 | 77 |
| | 24 | 46 | 83 | 181 | | 24 | 24 | 45 | 90 |
| | 25 | 24 | 68 | 172 | | 25 | 15 | 35 | 65 |
| | 26 | 41 | 53 | 145 | | 26 | 18 | 25 | 49 |
| | 27 | 17 | 46 | 113 | | 27 | 20 | 17 | 37 |
| | 28 | 17 | 49 | 92 | | 28 | 15 | 22 | 33 |
| | 29 | 20 | 36 | 77 | | 29 | 6 | 14 | 22 |
| | 30 | 16 | 23 | 75 | | 30 | 10 | 11 | 21 |
| | 31 | 4 | 16 | 50 | | 31 | 5 | 14 | 18 |
| | 32 | 17 | 18 | 40 | | 32 | 0 | 7 | 20 |
| | 33 | 9 | 15 | 39 | | 33 | 3 | 5 | 12 |
| | 34 | 9 | 15 | 38 | | 34 | 4 | 8 | 7 |
| | 35 | 7 | 19 | 27 | | 35 | 5 | 6 | 10 |
| | 36 | 2 | 11 | 24 | | 36 | 4 | 8 | 5 |
| | 37 | 5 | 8 | 25 | | 37 | 3 | 1 | 6 |
| | 38 | 3 | 3 | 12 | | 38 | 1 | 0 | 4 |
| | 39 | 1 | 3 | 14 | | 39 | 1 | 2 | 3 |
| | 40 | 3 | 3 | 11 | | 40 | 2 | 1 | 3 |
| | 41 | 1 | 4 | 14 | | 41 | 0 | 1 | 3 |
| | 42 | 3 | 4 | 14 | | 42 | 0 | 0 | 2 |
| | 43 | 3 | 5 | 14 | | 44 | 0 | 0 | 1 |
| | 44 | 0 | 2 | 5 | | 45 | 0 | 1 | 0 |
| | 45 | 1 | 4 | 2 | | 47 | 0 | 0 | 1 |
| | 46 | 0 | 1 | 7 | | 48 | 1 | 0 | 1 |
| | 47 | 0 | 2 | 6 | | 49 | 0 | 1 | 0 |
| | 48 | 0 | 1 | 5 | | 50 | 1 | 0 | 0 |
| | 49 | 0 | 1 | 1 | | | | | |
| | 50 | 1 | 1 | 0 | | | | | |
| | 51 | 0 | 2 | 1 | | | | | |
| | 52 | 1 | 0 | 1 | | | | | |
| | 54 | 0 | 0 | 2 | | | | | |
| | 56 | 1 | 0 | 2 | | | | | |
| | 58 | 0 | 1 | 0 | | | | | |
| | 59 | 1 | 0 | 1 | | | | | |
| | 61 | 0 | 1 | 1 | | | | | |
| | 62 | 0 | 1 | 0 | | | | | |