

Dependency length minimization vs. word order constraints: an empirical study on 55 treebanks

Xiang Yu, Agnieszka Falenska and Jonas Kuhn

Institut für Maschinelle Sprachverarbeitung

University of Stuttgart

firstname.surname@ims.uni-stuttgart.de

Abstract

This paper expands on recent studies of very large treebank collections aiming to find empirical evidence for language universals, specifically for the functionally motivated Dependency Length Minimization (DLM) hypothesis. According to DLM grammars are set up to support the expression of utterances in a way that minimizes the distance between heads and dependents. We construct several incremental baselines that lead from the random free order linearization to the real language by adding various word order constraints. We conduct detailed analyses on 55 treebanks and find that all of the constraints contribute to DLM. We show that DLM on the one hand shapes the regularity and on the other motivates the attested exceptions from canonical word order. The findings contribute to a more fine-grained, differentiated picture of the role of DLM in the interaction of competing constraints on grammar and language use.

1 Motivation and Background

The recent development of comparable dependency treebanks for a considerable number of languages across the typological spectrum (Nivre et al., 2016) has made it possible to address some long-standing hypotheses regarding a functional explanation of linguistic universals.

A number of recent papers (Liu, 2008; Futrell et al., 2015, a.o.) have used evidence from treebanks across languages to address what is arguably the most prominent hypothesis of a functionally motivated universal constraint, the Dependency Length Minimization (DLM) hypothesis, which can be traced back to (Behaghel, 1932).

Phrased as a language typological universal, the DLM hypothesis states that the evolution of languages is driven by the constraint that grammars should allow dependents to be realized as closely as possible to their heads – which is known to reduce the cognitive burden in processing (Gibson, 1998; Gibson, 2000). The Dependency Length (DL) of a sentence is defined as the sum of the distance between the head and dependent of all the dependency arcs in the sentence (see the example in Figure 1).

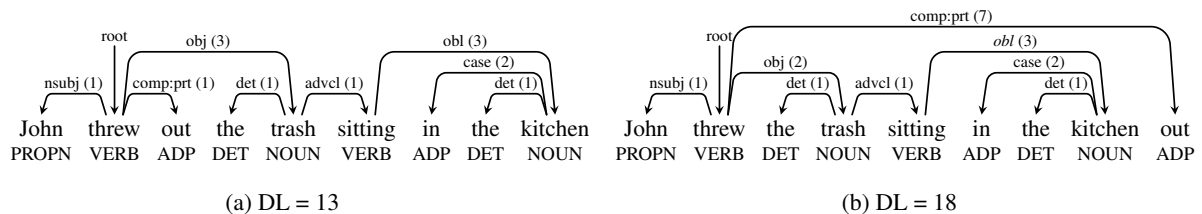


Figure 1: Example dependency trees and their dependency lengths, adapted from Futrell et al. (2015). The tree on the left is preferred since it has shorter DL and lower cognitive burden.

Liu (2008) perform the first cross-treebank study to compare the actual length of dependencies, for 20 languages, against the length that results when the dependency structures are linearized in random ways. The results indicate that languages indeed tend to minimize the dependency distance. Futrell et al. (2015) present a recent expansion of this type of treebank study to a set of 37 languages (which they argue to be the first comprehensive analysis that covers a broad range of typologically diverse

languages), presenting comparisons of the real dependency trees from the treebank with random re-orderings of the same dependency structures. The analysis shows that indeed across all 37 analyzed languages the real DL is significantly shorter than chance. This result corroborates findings from a broad range of empirical studies that are typologically less comprehensive (Gildea and Temperley, 2010; Gulordava and Merlo, 2015; Gulordava et al., 2015). This type of cross-treebank study has prompted a fair number of expansions and discussion regarding the typological implications (e.g. Jiang and Liu (2015), Chen and Gerdes (2018), Temperley and Gildea (2018)). In the present contribution, we go into some detail regarding a question that Futrell et al. (2015) have touched on, namely the relation between (the objective of minimizing) dependency length and language-specific word order constraints (which can also contribute to minimizing the cognitive load in parsing – but may conflict with the DLM objective).

Futrell et al. (2015) are careful to point out their awareness that the type of corpus study they performed makes it hard to distinguish the language typological aspect of the DLM hypothesis on the one hand (which would explain the exclusion of certain logically possible grammatical systems, which go against functional constraints/cognitive processing preferences) from facts about language use, relative to the respective grammatical constraints of a language, on the other.¹ The latter aspect, which is purely a matter of language processing, has also been discussed extensively under the DLM hypothesis (see e.g. Wasow (2002)). Futrell et al. (2015) “do not distinguish between DLM as manifested in grammars and DLM as manifested in language users’ choice of utterances; the task of distinguishing grammar and use in a corpus study is a major outstanding problem in linguistics, which we do not attempt to solve here.”

Besides the methodological question of how one could separate effects from corpus observations, it is worthwhile noting that in the logic of the typological DLM hypothesis, the aspect of language use cannot be completely ignored: if language evolution is indeed driven by these functional constraints, it should favor languages that permit *variation* – so speakers can react to the relative heaviness of constituents in the specific content they want to realize.

A comparison of the real DL with a random re-ordering baseline will necessarily conflate the effect from strict grammatical constraints (which the baseline re-orderings may break arbitrarily) and the relative freedom that any given grammar will leave open within the space of its constraints – and which speakers can exploit to optimize their utterances. Since they are aware of the effects of strict word order constraints that many languages impose, Futrell et al. (2015) present an additional comparison with a *Fixed Word Order Baseline* that is not fully random, but enforces consistent ordering constraints within each sentence. We note that the fact that this baseline chooses a new dependency relation ordering scheme for each sentence does not make it a good candidate for getting closer to the separation of globally fixed word order constraints dictated by the grammar and remaining spaces of free variation – which does seem to play a crucial role for approaching more fine-grained typological generalizations (and which ultimately needs to be clarified before one can claim to have empirical evidence for the DLM hypothesis as manifested in grammars rather than as a cognitive processing preference).

In this paper, we propose alternative baseline realizations of dependency trees that allow us to look more closely at the effect of specific relative ordering constraints in the comparison between random re-ordering realizations and the real treebank sequences along with their DL. We can thus study the manifestations of DLM relative to specific ordering phenomena in isolation. With a differentiated set of baselines, we can identify the DLM effect (1) in the distribution of dependents to both sides of the head; (2) in the direction of each single dependent to its head, and (3) in the ordering of the siblings on the same side of the head. These three phenomena can be seen as different types of word order constraints, where the first one concerns the quantity and balance of dependents, the latter two involve the order of individual dependency patterns, i.e., the combination of part-of-speech and dependency relation of the involved tokens. These word order constraints have been studied in various work (Ferrer-i Cancho, 2008; Ferrer-i Cancho, 2015; Liu, 2010; Gulordava et al., 2015). In this work, we identify all of the constraints

¹Other effects that cannot be captured (as Futrell et al. (2015) note) are differences in inflectional marking (in languages with strong case marking, much of the functional burden from word order is lifted) and the availability of non-projective dependency structures, which are a way of achieving shorter overall dependency length without violating certain hard word order constraints (Ferrer-i Cancho and Gómez-Rodríguez, 2016).

together in a large collection of treebanks, and show that each constraint contributes to the DLM in its unique way. Furthermore, we study their interaction by experimenting alternative word orders deviating from the original data and observe the impact on the dependency length.

We experiment with data instances following the regularities of word order and instances that manifest exceptions separately, and show that the word orders in the real data have shorter DL in both cases. This suggests that DLM is likely not just a result of the fixed word order, since it also happens in the non-canonical word order. Rather, it supports the hypothesis that the DLM influences both the regularities and the exceptions of word order.

2 Experiments on Dependency Length

2.1 Data

We perform our experiments on a selection of 55 treebanks from Universal Dependencies v2.4 (Nivre et al., 2019). The selection consists of training sets from all treebanks with at least 500 sentences (we take sentences with maximum 50 words). Where there are multiple treebanks for one language, we use the largest one to ensure stable and consistent estimation. We remove punctuation from trees and do not consider it when calculating DL, since punctuation biases statistics by introducing many long-distance right dependencies, and such dependencies do not contribute to the meaning of the head.

We consider only projective trees, since the non-projectivity could be interpreted as another way to minimize DL (Ferrer-i Cancho and Gómez-Rodríguez, 2016), which is out of the scope of this paper. Focusing on projective trees allows us to (1) analyze the DL on the subtree level, since the internal ordering of each subtree does not influence other subtrees; and (2) efficiently find the optimal ordering in terms of DL (Gildea and Temperley, 2007).

2.2 Baselines

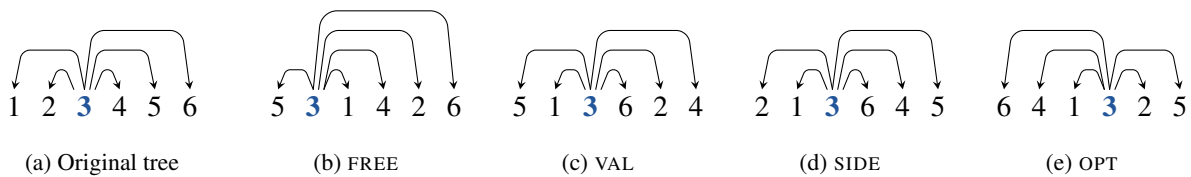


Figure 2: An example tree and results of four baseline linearization methods.

We compare the DL of the observed sentences in the treebanks (**OBS**) with four baselines that generate random linearizations with incremental constraints that leads to the real data:

FREE: free word order baseline, which does not impose any constraints except projectivity to the linearization. It is also used in Futrell et al. (2015).

VAL: same-valency baseline, which ensures that the numbers of left and right dependents of a head in the random subtree are the same as in the real data². In other words, we shuffle the left and right dependents in the original tree separately.

SIDE: same-side baseline, which ensures that the dependents in the random tree stay on the same side of the head as in the real data. This also satisfies the constraints in VAL, and differs from OBS only in the linear arrangement of dependents on each side.³

OPT: optimal baseline, which minimizes the dependency length, as in Gildea and Temperley (2007).

Figure 2 illustrates an example for each baseline, where 2a is the original ordering; 2b shuffles all the tokens in the tree; 2c ensures that there are 2 left dependents and 3 right dependents as in 2a; 2d shuffles the left and right dependents of 2a separately; and 2e is the optimal ordering of the tokens, assuming the label of each dependent also signifies the size of its subtree.

²We use the term *valency* only to describe the number of dependents, not their types.

³SIDE baseline is mentioned but not analyzed in the appendix of Futrell et al. (2015).

2.3 Results

Figure 3 illustrates the average DL of the five described baselines with respect to the sentence length.⁴ From the longest to the shortest are: FREE, VAL, SIDE, OBS, and OPT, and each adjacent pair is clearly separated. We follow with systematic analyses to explain the differences between the baselines. Section 3.2 explains the influence of balanced left and right dependents to the DL (FREE vs. VAL); Section 3.3 illustrates the influence of the head direction as a word order constraint (VAL vs. SIDE); and 3.4 shows the ordering of same-side siblings as another word order constraint (SIDE vs. OBS). In each of the aspects, the constraints from the observed data show clear preference for shorter DL, but does not reach the shortest possible DL due to other constraints, which partly explains the difference between OBS and OPT.

3 Analysis on Word Order Constraints

3.1 Definition and Statistics

Apart from the quantitative distribution of dependents, we study two types of word order constraints conditioned on the dependency pattern, namely **head direction** and **sibling ordering**.

The head direction constraint describes on which side of the head lies the dependent for each dependency pattern. Its dependency pattern is a triple of the universal part-of-speech (UPOS) tag of the head, the UPOS tag of the dependent, and the dependency label. An example dependency pattern from Figure 1 is $\langle \text{VERB}, \text{ADP}, \text{comp:prt} \rangle$, and its value is *right*.

The sibling ordering constraint describes the pairwise precedence relation of the dependents on the same side of the head. It is an approximation of the total order of the siblings, but much simpler and resistant to the data scarcity problem (Gulordava, 2018). Its dependency pattern is a sextuple of the UPOS of the head, the side of the involved dependents, and the UPOS and label of the two dependents. An example is $\langle \text{VERB}, \text{right}, \text{ADP}, \text{comp:prt}, \text{NOUN}, \text{obj} \rangle$, and the value is *left* for Figure 1a and *right* for Figure 1b.

For both dependency patterns, we count the frequency of their values, and use the entropy to measure the **freedom** of that pattern. For example, if a noun appears 50 times on the left of a verb as the subject, and 10 times on the right, then the entropy of $\langle \text{VERB}, \text{NOUN}, \text{obj} \rangle$ is 0.65. We then measure the freedom of the overall word order constraint by taking the average entropy of each single dependency pattern in that constraint type weighted by its frequency in the data. Figure 4 shows all treebanks characterized by the two types of word order freedom. We mark four most common language families and annotate the rest with their language codes. Generally, we do not see correlation of these two types of freedom, which indicates that they characterize different aspects of the word order. Many verb-final languages cluster near the top left corner, since they tend to have strict constraints that all arguments of a verb are uttered before the verb, but the exact ordering of the arguments is very flexible.

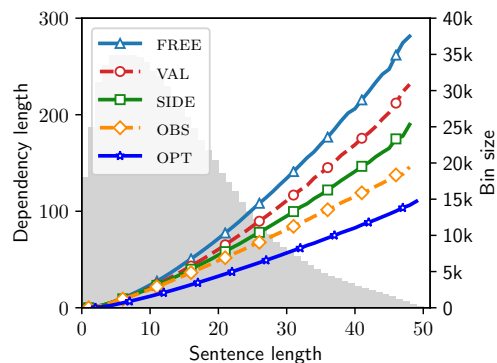


Figure 3: Average DL vs. sentence length across 55 treebanks; grey bars indicate frequency of the sentence length.

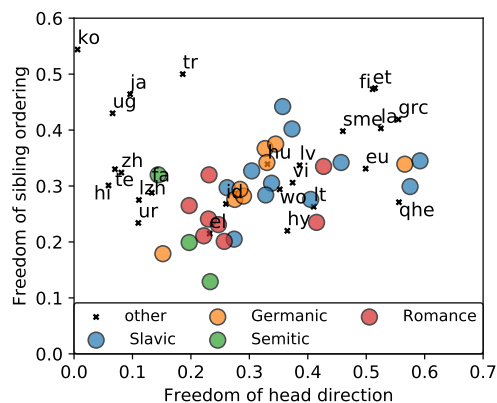


Figure 4: All treebanks characterized by the two types of word order freedom.

⁴The general trend and ranking of the baselines in each individual treebank are consistent with the average.

3.2 Dependent Distribution: FREE vs. VAL

Imposing the valency constraints on the FREE baseline makes the first reduction of DL. It can be easily explained by the fact that DL is shorter when the left and right dependents are more balanced, or in other words, when the head is positioned in the middle of its dependents, as shown in Ferrer-i Cancho (2015). In the FREE baseline, the head is equally likely to be placed in any position of the dependents, while in the VAL baseline, where the number of left and right dependents of each head is the same as from the real tree, it is more likely than chance to have balanced number of dependents on both sides.

To demonstrate this fact, we measure the **imbalance** of a head by the difference of numbers of dependents on both sides divided by the number of dependents. The more balanced a tree is, the smaller the value. We take the averaged imbalance value of all heads with more than one dependent as the imbalance measurement of the whole treebank. We also calculate the expected imbalance of FREE. For a head with n dependents, there are $n + 1$ possible location to insert the head, and the difference of dependents at each location would be: $\{n, n - 2, \dots, 2, 0, 2, \dots, n - 2, n\}$ if n is even, or $\{n, n - 2, \dots, 1, 1, \dots, n - 2, n\}$ if n is odd. The sum of these values is $\left\lfloor \frac{(n+1)^2}{2} \right\rfloor$, normalized by the length n and averaged by the equiprobable locations $n + 1$, thus the expected imbalance value is: $imb(n) = \left\lfloor \frac{(n+1)^2}{2} \right\rfloor / n(n + 1)$.

The measured average imbalance value for VAL is 0.47 (same as OBS) and 0.67 for FREE (also very close to the expectation from the formula), which means that VAL distributes the dependents in a more balanced way. There are only three languages (Telugu (te), Uyghur (ug), and Korean (ko)) where VAL has higher imbalance, mainly because they are verb-final languages, which have very unbalanced verb dependents. This general trend indicates that the real languages tend to have more balanced dependents (in other words, position the head more central) than chance level, thus the reduction of length from FREE to VAL.

Note that in this scenario, we only consider the *number* of dependents as the measurement of balance, which is a simplistic heuristics, while the *subtree size* of the dependents is the more accurate measurement. We explore this factor in Section 3.4.

3.3 Head Direction: VAL vs. SIDE

Next, we consider the length reduction from VAL to SIDE, which puts additional constraints on the head direction based on the real data.

We illustrate the relation between the head direction preference and its effect on DL through two scenarios. The first scenario studies the regularity of head direction: we flip the dependent to the other side of the head if it is on the **majority** side of its dependency pattern based on the statistics of the real data. The second scenario studies the exception of head direction: we flip the dependent if it is on the **minority side**. While flipping the dependent, we keep the order of all other dependents unchanged, and insert the flipped dependent into a position that minimizes the DL. This way, we make sure that the hypothetical flipping is optimal and the comparison to the original order is fair.

All the results are shown in Figure 5, where the y-axis shows the percentage of cases where the original order has shorter DL, i.e., flipping would increase the length, and the x-axis show the freedom of head direction for each language. If a point lies above the 50% line (red line on the plot), it means that the real data in that treebank has shorter DL than by chance.

Figure 5a shows the overall trend of the flipping experiment, where the real data is more likely to have shorter DL than the flipped ones, except for a few verb-final languages (Korean (ko), Japanese (ja), Telugu (te), and Uyghur (ug)), similar to the exceptions in Section 3.2. Since they tend to utter all the arguments before the verb, flipping any arguments would balance the head thus reduce the DL.

Note that by flipping one dependent to the other side, the valency of the head is also changed, therefore it is possible that the reduction of DL is caused by the effect described in Section 3.2. However, we record the imbalance before and after the flipping, and it is almost not changed on average, in other words, the influence from the change of valency is very small for this experiment.

Figure 5b shows the majority flipping scenario, which is very similar to the overall picture, since this scenario has more test cases (by definition), therefore dominates the statistics. The plot shows that the

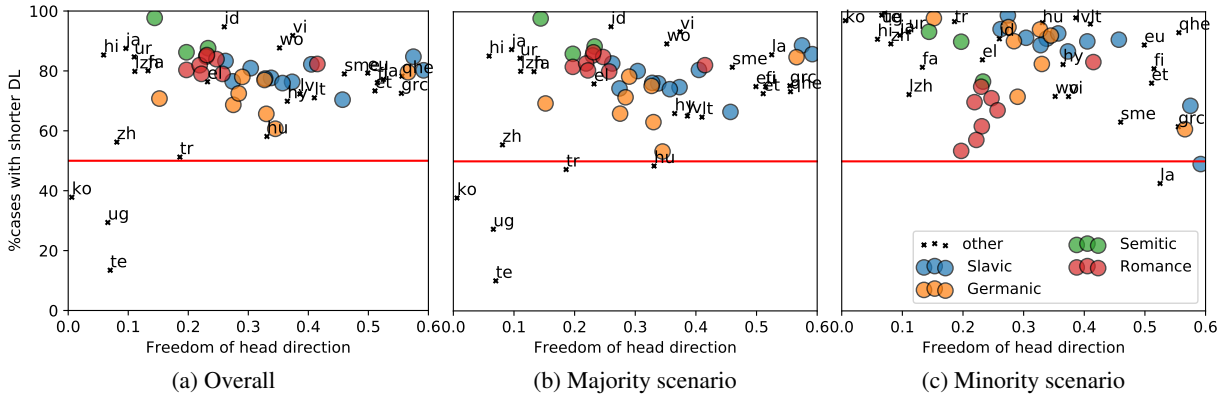


Figure 5: Percentage of trees where the real order has shorter DL than flipping one dependent.

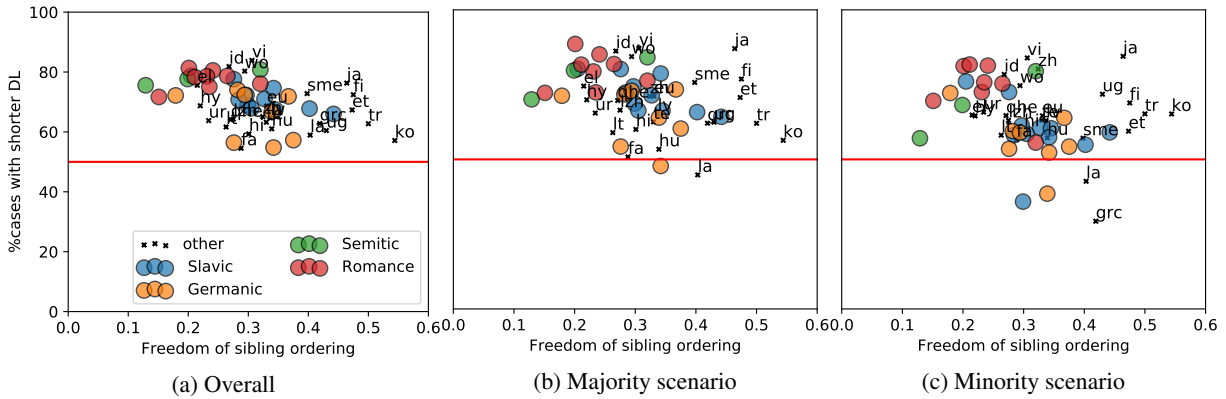


Figure 6: Percentage of trees where the real order has shorter DL than swapping two siblings.

canonical order leads to shorter DL for most of the languages, thus supports the hypothesis that DLM drives the evolution of word order. This then invites the question of why we still use non-canonical order in utterance. We can explain it in terms of DLM by observing the minority scenario, where we “correct” the head direction in the real data if it is not the majority choice. As shown in Figure 5c, for almost all languages, the original minority order has shorter DL than the flipped majority order. More generally, the minority scenario results resemble the mirror image of the majority case, namely the languages with shorter DL in the majority case would have less shorter DL in the minority case.

Assuming the head direction in each instance is realized without influence of DL, then either the majority or the minority scenario should have less than 50% cases with shorter DL. However, the fact that both scenarios have more than 50% positive cases indicates that DLM indeed influences the shaping of the head direction preferences as well as motivating the deviation from such preferences.

3.4 Sibling Ordering: SIDE vs. OBS

Finally, we look at the length reduction from SIDE to OBS, which only differ in the ordering of siblings on the same side. Their difference in DL indicates that the ordering of the same-side siblings is also influenced by DLM, as in the example in Figure 1. According to the explanation of DLM, we prefer 1a over 1b because keeping smaller subtrees closer to the head helps reduce DL and cognitive effort.

To verify whether the data supports the claim, we conduct similar experiments as in Section 3.3, where we compare every subtree in the original data to the modification where we swap one pair of dependents on the same side. We also compare two scenarios, majority and minority, where the swapped pair belongs to the regularity and exception of its dependency pattern, respectively.

The results are shown in Figure 6. We notice, that overall as well as in both majority and minority scenarios, the original subtrees has shorter DL than swapping two dependents. This again indicates that given the head direction, the real language tends to arrange the dependents in a way to minimize the

DL, regardless of whether the order is predominant or not. It also supports the similar conclusion as in Section 3.3 that DLM motivates both the regularity and exceptions in the ordering of siblings.

3.5 Technical Notes

It is worth noting that the dependency relation may be an artifact of the treebank design, which does not necessarily reflect the nature of human cognition. Many previous works have analyzed the syntactic and semantic treatment of the UD annotation scheme, cf. de Lhoneux and Nivre (2016), Wisniewski and Lacroix (2017), Osborne and Gerdes (2019). For example, UD tend to make the content word (NOUN) as the head of the function word (ADP), thus the common word order of placing a prepositional phrase *after* the noun it modifies would have longer DL than the opposite case, which is against the DLM hypothesis. In another annotation scheme, e.g. the Penn Treebank (Marcus et al., 1993), where the preposition is the head of the noun, the same word order would support the DLM hypothesis.

Another example of the influence of annotation scheme is the contrast between Korean(ko) and Japanese(ja) in our experiments. These two languages have very similar typological features, but stay rather far away in the plots. One major reason is that the Japanese treebank splits case markers as individual tokens, while the Korean treebank treats them as part of the noun. These idiosyncrasies could significantly change the statistics of the treebanks in terms of DL.

In this work, we do not consider other annotation schemes, e.g. the HamleDT collection (Zeman et al., 2014), nor do we deal with annotation idiosyncrasies within UD. However, we acknowledge that they have certain influence on the analysis of DL. Furthermore, which annotation scheme is closer to human cognition is still an open question.

4 Conclusion

In this paper we have broken down the effect of dependency length minimization step by step, and analyzed its relation to the dependent distribution, head direction, and sibling ordering. The systematic breakdown indicates that natural languages universally show clear preference for shorter dependency length in all three aspects, both in the regularities and in the exceptions. Our findings provide more detailed evidence for the hypothesis that DLM is a universal phenomenon in natural language.

One very interesting direction for future work is the interaction of multiple word order constraints. We have shown that most constraints are locally optimal, i.e., reversing a single constraint would likely increase the dependency length, since it might not be compatible with some other constraints. The grouping effect of constraints with respect to DL might provide explanations to some observations in Greenberg’s linguistic universal (Greenberg, 1963).

In this work, we use dependency patterns extract from the treebank to characterize the word order constraints, which is transparent but maybe not nuanced enough. An alternative way could be to use a statistical linearizer (Bohnet et al., 2010) to model the word order constraints (but the features implicitly related to DL should be carefully disabled), which could serve as a even stronger baseline.

Acknowledgements

This work was in part supported by funding from the Ministry of Science, Research and the Arts of the State of Baden-Württemberg (MWK), within the CLARIN-D research project. We also thank the anonymous reviewers for their feedbacks and ideas for future work.

References

- Otto Behaghel. 1932. *Deutsche Syntax: eine geschichtliche Darstellung. Wortstellung, Periodenbau*. Winter.
- Bernd Bohnet, Leo Wanner, Simon Mille, and Alicia Burga. 2010. Broad coverage multilingual deep sentence generation with a stochastic multi-level realizer. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 98–106. Association for Computational Linguistics.

- Xinying Chen and Kim Gerdes. 2018. How do universal dependencies distinguish language groups? In Jingyang Jiang and Haitao Liu, editors, *Quantitative Analysis of Dependency Structures*, pages 277–293, Berlin/Boston. De Gruyter Mouton.
- Miryam de Lhoneux and Joakim Nivre. 2016. Should have, would have, could have. investigating verb group representations for parsing with universal dependencies. In *Proceedings of the Workshop on Multilingual and Cross-lingual Methods in NLP*, pages 10–19.
- Ramon Ferrer-i Cancho and Carlos Gómez-Rodríguez. 2016. Crossings as a side effect of dependency lengths. *Complexity*, 21(S2):320–328.
- Ramon Ferrer-i Cancho. 2008. Some word order biases from limited brain resources: A mathematical approach. *Advances in Complex Systems*, 11(03):393–414.
- Ramon Ferrer-i Cancho. 2015. The placement of the head that minimizes online memory: a complex systems approach. *Language Dynamics and Change*, 5(1):114–137.
- Richard Futrell, Kyle Mahowald, and Edward Gibson. 2015. Large-scale evidence of 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. *Image, language, brain*, 2000:95–126.
- Daniel Gildea and David Temperley. 2007. Optimizing grammars for minimum dependency length. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 184–191.
- Daniel Gildea and David Temperley. 2010. Do grammars minimize dependency length? *Cognitive Science*, 34(2):286–310.
- Joseph H Greenberg. 1963. Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language*, 2:73–113.
- Kristina Gulordava and Paola Merlo. 2015. Diachronic trends in word order freedom and dependency length in dependency-annotated corpora of latin and ancient greek. In *Proceedings of the third international conference on dependency linguistics (Depling 2015)*, pages 121–130.
- Kristina Gulordava, Paola Merlo, and Benoit Crabbé. 2015. Dependency length minimisation effects in short spans: a large-scale analysis of adjective placement in complex noun phrases. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, volume 2, pages 477–482.
- Kristina Gulordava. 2018. *Word order variation and dependency length minimisation: a cross-linguistic computational approach*. Ph.D. thesis, University of Geneva.
- 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.
- Haitao Liu. 2008. Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, 9(2):159–191.
- Haitao Liu. 2010. Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua*, 120(6):1567–1578.
- Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330.
- Joakim Nivre, Marie-Catherine De Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D Manning, Ryan T McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, et al. 2016. Universal Dependencies v1: A Multilingual Treebank Collection. In *LREC*.
- Joakim Nivre, Mitchell Abrams, Željko Agić, et al. 2019. Universal dependencies 2.4. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.
- Timothy Osborne and Kim Gerdes. 2019. The status of function words in dependency grammar: A critique of universal dependencies (ud). *Glossa: a journal of general linguistics*, 4(1).

David Temperley and Daniel Gildea. 2018. Minimizing syntactic dependency lengths: typological/cognitive universal? *Annual Review of Linguistics*, 4:67–80.

Thomas Wasow. 2002. *Postverbal Behavior*. CSLI lecture notes. CSLI Publ.

Guillaume Wisniewski and Ophélie Lacroix. 2017. A systematic comparison of syntactic representations of dependency parsing. In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*, pages 146–152.

Daniel Zeman, Ondřej Dušek, David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Zdeněk Žabokrtský, and Jan Hajič. 2014. HamleDT: Harmonized multi-language dependency treebank. *Language Resources and Evaluation*, 48(4):601–637.