[Proceedings of the Society for Computation in Linguistics](https://scholarworks.umass.edu/scil)

Volume 1 Article 23

2018

Dependency Length Minimization and Lexical Frequency in Prepositional Phrase Ordering in English

Zoey Liu
University of California, Davis, yiliu@ucdavis.edu

Kenji Sagae
University of California, Davis, ks@sagae.org

Recommended Citation

Liu, Zoey and Sagae, Kenji (2018) "Dependency Length Minimization and Lexical Frequency in Prepositional Phrase Ordering in English," *Proceedings of the Society for Computation in Linguistics*: Vol. 1 , Article 23.
DOI: https://doi.org/10.7275/R5N014QW
Available at: https://scholarworks.umass.edu/scil/vol1/iss1/23

Dependency Length Minimization and Lexical Frequency in Prepositional Phrase Ordering in English

Zoey Liu and **Kenji Sagae**
Department of Linguistics
University of California, Davis
One Shields Avenue, Davis, CA
{yiliu, sagae}@ucdavis.edu

1 Introduction

Finding and explaining the structural variation among languages has served as one of the primary goals in the study of linguistics and cognitive science. The availability of multilingual corpora now allows for empirical investigation of issues in linguistic typology using computational methods (Merlo, 2015; Cotterell and Eisner, 2017), shedding new light on cross-linguistic regularities and the underlying mechanisms behind structural patterns in natural language.

Recent research has shown that the human language parser prefers constituent orders that minimize the distance between syntactic heads and their dependents (Futrell et al., 2015). This preference, known as Dependency Length Minimization (DLM), predicts that if alternative orderings exist for the constituents within a sentence, constituents of shorter length tend to be placed closer to their heads and thus shorten overall dependency distance in the sentence. Focusing on prepositional phrases in English, which in some contexts can be moved within a sentence resulting in alternatives with comparable grammaticality and meaning, we investigate to what extent DLM explains prepositional phrase orderings observed in corpora of spoken and written language, and the role played by lexical frequency. If DLM preferences are due to processing efficiency (Gibson, 1998; Temperley, 2007), it is possible that other factors related to processing efficiency, such as the cost of lexical retrieval, may also influence prepositional phrase ordering.

As an illustration of how DLM applies to the ordering of prepositional phrases, consider the following sentences:

(a) He drove [with friends] [to the park nearby]

(b) He drove [to the park nearby] [with friends]

Both (a) and (b) have two prepositional phrases (PPs), shown within square brackets: *with friends* and *to the park nearby*. Switching the order of the two PPs does not affect the grammaticality or the meaning of the sentences, that is, (a) and (b) are both grammatical and convey the same meaning. As indicated by the syntactic dependency arcs, we consider the prepositions in both PPs as dependents of the verb *drove*, and it is the dependency arc between a preposition and *drove* that attaches the corresponding PP to the head of the verb phrase. The length of the dependency that attaches each PP is then the linear distance between the head of the dependency (the verb *drove*) and the preposition, which serves as the dependent. In both (a) and (b), the dependency length between *drove* and its closest PP is the same; however, the distance between *drove* and the farther PP is shorter in (a), where the PP of shorter length

Proceedings of the Society for Computation in Linguistics (SCiL) 2018, pages 189-192.
Salt Lake City, Utah, January 4-7, 2018

is placed closer to the verb. From this example, we can see that DLM predicts that in cases where there are two PPs to one side of the same head that can be expressed in either order, there is a preference for placing the shorter PP closer to its head.

The effect of DLM on syntactic preferences has been examined in various ways (Temperley, 2007; Gildea and Temperley, 2007; Futrell et al., 2015; Hawkins, 1999), and although strong evidence for DLM has been found, it is clear that it is not the only factor in determining preferred word orders. The interaction between DLM and other preferences and constraints in different contexts is currently under investigation (Gulordava et al., 2015; Wiechmann and Lohmann, 2013). If the preference for DLM shown cross-linguistically is indeed driven by ease of processing and the goal for efficient communication, we expect other factors that have been found to facilitate processing efficiency to exert an effect on word ordering as well. One factor that has been shown to be relevant to the ease of processing is lexical frequency (Hawkins, 2014).

2 PP ordering and lexical frequency

A traditional ordering rule for PPs and adverbials in postverbal position in English appears to be semantic, namely Manner before Place before Time (MPT), as in *Zoey danced* $[_{manner}$ *elegantly*$]$ $[_{place}$ *on the dance floor*$]$ $[_{time}$ *at night*$]$. Using 394 relevant sentences, Hawkins (1999) found that dependency length serves as the primary factor governing the order of PPs with the same head, with no substantial influence from MPT. Wiechmann and Lohmann (2013) found similar results with 1,256 sentences from both the written and spoken sections of International Corpus of English.

We go beyond these studies by examining the role of lexical frequency. The correlation between lexical frequency and structural complexity can be traced back to the markedness hierarchies proposed by Greenberg (1966). For instance, in languages with rich morphology, the markedness hierarchy of case (Nom>Acc>Dat>Other) reflects the frequency ranking of the different cases; in other words, as the formal marking goes from *Nom* to *Other*, the frequency of occurrence of each case marking declines. The markedness hierarchy of number (Sing>Plur>Dual>Trial/Paucal) shows the same correspondence to frequency (e.g. in English singular form *dog* is less morphologically complex than the plural form *dogs*, and thus occurs more often than plural form). As suggested by Keenan and Comrie (1977) in their Accessibility Hierarchy, the underlying cause for such pattern is attributed to ease of processing which declines for each position down the hierarchy. The processing load of different hierarchy positions is shaped by their complexity and frequency of occurrence. More frequent categories are associated with greater processing ease, accessibility and predictability, whereas less frequent items are harder to access; they require more effort for activation and processing, and thus more explicit coding is needed down the hierarchies.

In our experiments, we leverage larger corpora annotated with syntactic information and commonly used in computational linguistics to explore the role of DLM, MPT and lexical frequency in PP ordering with a comparative analysis using a larger number of PPs and different language genres than in the previous work by Hawkins (1999) and Wiechmann and Lohmann (2013). We hypothesize that both DLM and lexical frequency play a role in PP ordering, and specifically that PPs with more frequent words, which are presumably more easily processed and retrieved on average, tend to appear first.

3 Experiments & Results

3.1 Data

We use the Penn Treebank (Marcus et al., 1993), which includes syntactic structures for approximately one million words of text from each of: the Wall Street Journal (WSJ), transcriptions of spontaneous spoken conversations from the Switchboard corpus (Godfrey et al., 1992), and the Brown corpus (Kučera and Francis, 1967). We search for sentences in the Penn Treebank with verb phrases containing exactly two PPs that are attached to the same verb, where the order of the PPs can be switched without affecting the grammaticality or the meaning of the sentence.[1]

[1] A precise estimate of how many of these sentences with two PPs fit our criteria of retained grammaticality and meaning when the PPs are swapped is pending, but preliminary manual inspection of a sample from each corpora suggests a large ma-

3.2 Effect of DLM on PP ordering

To estimate the effect of DLM on PP ordering, we follow a similar procedure as Hawkins (1999). We measure the lengths of the PP closer to the verb and of the PP farther from the verb as the number of tokens in each PP,[2] then calculate the proportion of cases where the shorter PP occurs closer to the head verb (V), the longer PP appears closer to V, or when the two PPs are of equal length, for each corpora separately. As shown in Table 1, the order predicted by DLM is strongly preferred, as the number of sentences that have the shorter PP closer to the verb is 1.8 to 3.5 times larger than the number of sentences that have the longer PP closer to the verb. However, in roughly 20% of all sentences, DLM makes no prediction, since the two PPs have the same number of tokens. Although these numbers suggest that the effect of DLM is not as strong in spoken language as it is in written text, the preference for shorter dependencies is substantial across all three corpora.

Corpus	Shorter PP closer to V	Longer PP closer to V	Equal length	Total
WSJ	**54.7%**	22.4%	22.7%	3596
Brown	**62.2%**	17.9%	19.9%	3033
SWB	**48.3%**	27.3%	24.4%	1187

Table 1: Effect of DLM in PP ordering preference

3.3 Effect of MPT on PP ordering

While Hawkins (1999) found that MPT plays no significant role in PP ordering, and Wiechmann and Lohmann (2013) found it plays only a weak role, it is possible these results may be due to the use of smaller language samples. In our dataset, MPT is very effective in predicting PP order when each of the two PPs are annotated in the treebank with function tags that reflect manner (PP-MNR), place (PP-LOC) or time (PP-TMP). Once we restrict our analysis to sentences that have both PPs annotated with these function tags, we are left with 6% of all sentences with two PPs with the same head. Within this set where MPT can be applied, it correctly accounts for the order of 89.3% of sentences in WSJ, 100% in Switchboard, and 100% in Brown. However, because it applies so infrequently, its overall impact is much smaller than that of DLM.

3.4 Effect of lexical frequency on PP ordering

Our investigation of the role of lexical frequency as a factor in PP ordering attempts to address the hypothesis that words that occur more frequently tend to appear first, as they are easier to retrieve. We first estimate a unigram language model from a large corpus with no overlap with the Penn Treebank. The product of the unigram probabilities for each token in the PP is clearly sensitive to length, and to separate the contributions of PP length and lexical frequency in PP ordering, we look instead at the perplexity of each PP. According to our hypothesis that more frequent items are produced first, the PP with lowest perplexity should appear closer to the verb. Table 2 shows how often the PP with lowest perplexity is closer to the verb in each of the three corpora. To examine the effect of domain match and mismatch between the language model and the target text, we use two unigram models: one estimated from 20 million words from the Wall Street Journal taken from the BLLIP corpus (Charniak et al., 2000), and one estimated from the unigram counts for the 1/3 million most frequent words from the Google Web Trillion Word Corpus provided by Peter Norvig[3].

Model	WSJ	Brown	Switchboard
BLLIP	54.5%	56.1%	51.2%
Norvig	53.1%	58.9%	55.2%

Table 2: Effect of lexical frequency on PP ordering preference

Although the effect of lexical frequency shown in Table 2 is not quite as pronounced as the effect of DLM, there does seem to be a general preference for PPs with more frequent words to appear first. We expected this preference to appear most clearly in the WSJ corpus using the WSJ language model (BLLIP), following the intuition that some of the lexical frequency information is dependent on domain. However, there is little difference between the results in the WSJ corpus using the two different language models. Domain mismatch could be behind the lack of preference on Switchboard based on lexical frequency distribution estimated from the

jority of sentences fit our criteria.

[2]We approximate phrase length using the number of tokens following the Penn Treebank tokenization.

[3]http://norvig.com/ngrams/

BLLIP corpus. The preference for higher frequency lexical items to appear first is most pronounced in the Brown corpus, with either language model.

We now return to the idea that DLM and frequency operate at the same time, and each contributes to PP ordering preference. While the precise way in which they interact is unknown, based on the observations in Tables 1 and 2 that DLM is a stronger predictor of PP ordering than lexical frequency, we combine DLM and frequency in a simplistic way, using primarily DLM to predict PP ordering and falling back to lexical frequency when the PPs are of equal length. Predictions made according to this simple combination account for roughly 70% of the observed PP orders in the two corpora of written text (WSJ and Brown), and 60% of the observed PP orders in the corpus of transcribed spontaneous spoken language (Switchboard), regardless of the unigram language model used (Table 3).

Model	WSJ	Brown	Switchboard
BLLIP	70.2%	73.4%	60.7%
Norvig	70.4%	72.8%	61.5%

Table 3: Simple combination of DLM and lexical frequency

Although these predictions may appear more accurate than those in Table 1, they are close to what one would obtain by using DLM and breaking ties randomly assuming a uniform distribution. In other words, the contribution of lexical frequency under our simple fall-back scheme is very limited. Since Table 2 suggests frequency may be useful in modeling PP order, it is possible that a combination of DLM and frequency using weights estimated using logistic regression might yield better predictions. This is left as future work, as is an in depth analysis of the apparent difference between spoken and written corpora, other ways to take frequency effects into account (e.g. word co-occurrence), and a cross-lingual investigation of DLM and lexical frequency in PP ordering.

References

Eugene Charniak, Don Blaheta, Niyu Ge, Keith Hall, John Hale, and Mark Johnson. 2000. BLLIP 1987-89 WSJ Corpus Release 1.

Ryan Cotterell and Jason Eisner. 2017. Probabilistic typology: Deep generative models of vowel inventories. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1182–1192. Association for Computational Linguistics.

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.

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. Association for Computational Linguistics.

John J. Godfrey, Edward C. Holliman, and Jane McDaniel. 1992. Switchboard: Telephone speech corpus for research and development. In *1992 IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP*, pages 517–520. IEEE.

Joseph H. Greenberg. 1966. *Universals of language*. MIT press.

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 (Short Papers)*, pages 477–482.

John A. Hawkins. 1999. The relative order of prepositional phrases in english: Going beyond manner–place–time. *Language variation and change*, 11(3):231–266.

John A. Hawkins. 2014. *Cross-linguistic variation and efficiency*. OUP Oxford.

Edward L. Keenan and Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic inquiry*, 8(1):63–99.

Henry Kučera and Winthrop Nelson Francis. 1967. *Computational analysis of present-day American English*. Dartmouth Publishing Group.

Mitchell P. Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. Building a large annotated corpus of english: The penn treebank. *Computational linguistics*, 19(2):313–330.

Paola Merlo. 2015. Predicting word order universals. *Journal of Language Modelling*, 3(2):317–344.

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

Daniel Wiechmann and Arne Lohmann. 2013. Domain minimization and beyond: Modeling prepositional phrase ordering. *Language Variation and Change*, 25(1):65–88.