An Argument for Symmetric Coordination from Dependency Length Minimization: A Replication Study

Adam Przepiórkowski[*,†], **Magdalena Borysiak**[†], **Adam Głowacki**[†]
[*]ICS Polish Academy of Sciences, [†]University of Warsaw
adamp@ipipan.waw.pl, {mj.borysiak, am.glowacki}@student.uw.edu.pl

Abstract

It is well known that left conjuncts tend to be shorter in English coordinate structures. On the basis of Penn Treebank, Przepiórkowski and Woźniak (2023) show that this tendency depends on the difference between lengths of conjuncts: the larger the difference, the stronger the tendency for the shorter conjunct to occur on the left. However, this dynamics is observed only when the governor of the coordinate structure is on the left of the coordination (e.g., *Bring apples and oranges!*) or when it is absent (e.g., *Come and sing!*), and not when it is on the right (e.g., *Apples and oranges fell*). Given the principle of Dependency Length Minimization, this turns out to provide an argument for the symmetric structure of coordination. We replicate and sharpen this result on the basis of a much larger dataset: parts of the COCA corpus parsed with Stanza. We also investigate the dependence of this result on the assumed unit of length (word vs. character) and on genre.

Keywords: Coordination, Dependency Length Minimization, Syntax, Stanza, Universal Dependencies

1. Introduction

Since the beginning of the replication crisis in early 2000s (Ioannidis, 2005), there has been a growing recognition of the importance of replication studies. Also within linguistics, a steadily increasing number of researchers put emphasis on repeated testing of claims, rather than just on producing novel claims (Sönning and Werner, 2021).

The aim of this paper is to report on a replication of a novel result reported in a recent ACL paper (Przepiórkowski and Woźniak, 2023), namely, that the dynamics of conjunct lengths in English coordinations provides a linguistic argument for certain views on the structure of coordination (e.g., that utilized in Prague Dependency Treebanks) and against others (e.g., that used in basic Universal Dependencies). Such replication is important, as this claim, if confirmed, is bound to have impact on theoretical linguistics: it provides an argument against some popular generative analyses of coordination (e.g., Munn 1993 and Zhang 2009) and for others (e.g., Neeleman et al. 2023). Based on a much larger and more varied dataset, we do not only confirm and sharpen this claim, but also discover a certain subtlety related to how textual lengths are measured (in words or in characters).

We start in §2 by presenting the claim and reasoning in Przepiórkowski and Woźniak 2023 (PW23, henceforth), then in §3 we describe how our dataset (parts of COCA parsed with Stanza) differs from theirs (manually annotated PTB). In §4 we present basic statistical results, discuss why the results of logistic regression modelling, as utilized in PW23, are misleading, and in §5 we present a more detailed analysis of the findings, which confirms PW23's argument for symmetric coordinations. §6 evaluates the dataset on which these results are based and points out limitations of the current study, while §7 concludes.

2. Przepiórkowski and Woźniak 2023

2.1. Empirical Findings

PW23's empirical findings are based on $PTB_{\&}$, a version of the Penn Treebank (PTB; Marcus et al. 1993) made available by Ficler and Goldberg (2016) and containing various corrections and more explicit representations of coordinate structures. PW23 extract 21,825 binary coordinations from $PTB_{\&}$ and investigate how the proportion of coordinations with the left conjunct shorter than the right conjunct (with respect to all coordinations with unequal lengths of the two conjuncts) changes with the absolute difference between the two lengths, depending on the presence and position of the governor. Their findings are summarized in Figure 1, which presents the result of fitting logistic models to the $PTB_{\&}$ data.

Whether length is measured in words (see the first column in Figure 1) or in characters (see the second column), when the governor is on the left of the coordinate structure, as in *Bring apples and oranges!* (where *bring* is the governor of the coordinate structure *apples and oranges*), the proportion of coordinations with the left conjunct shorter grows with the absolute difference of lengths (see

LREC-COLING 2024, pages 1021–1033
20-25 May, 2024.

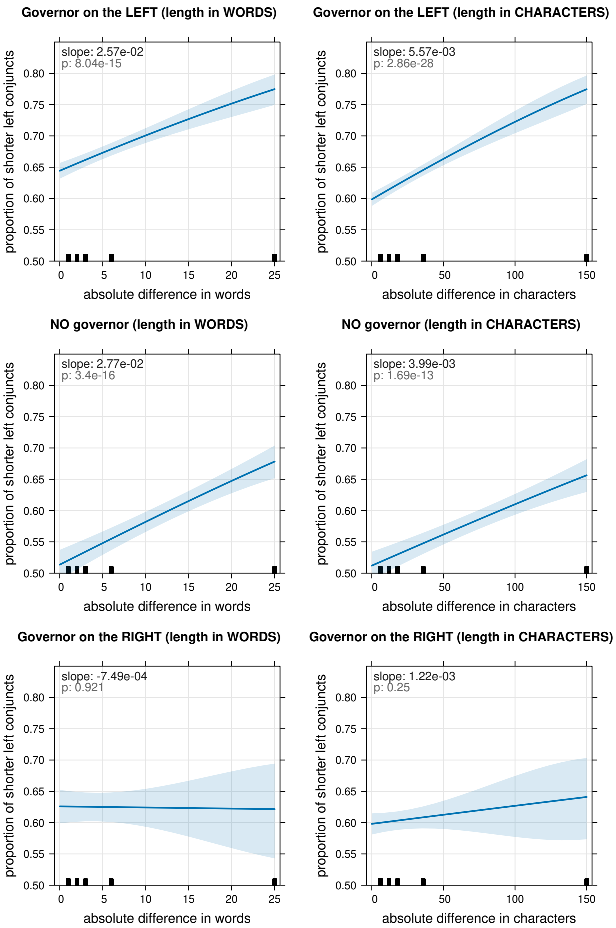


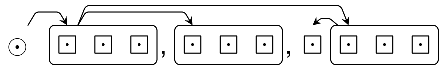
Figure 1: Modelled proportions of coordinations in PTB$_{\&}$ with left conjuncts shorter, depending on the absolute difference of conjunct lengths, with confidence bands (Przepiórkowski and Woźniak, 2023).

the first row). For example, if the two conjuncts are *red apples* (2 words, 10 characters including spaces) and *very large oranges* (3 words, 18 characters), *red apples* is more likely to be realized as the first conjunct (i.e., as in *red apples and very large oranges* rather than as in *very large oranges and red apples*), but when the second conjunct is even longer, e.g., *those very large ripe oranges* (5 words, 29 characters), the likelihood that *red apples* will be the left conjunct is even greater. The likelihood of the shorter conjunct being on the left grows similarly when there is no governor, as in *Come and sing!* (see the second row). In all four plots, the slopes of the curves are significantly positive ($p \ll 0.001$). However, and this is the crucial new observation of PW23, this effect disappears when the governor is on the right, as in *Apples and oranges fell* (see the third row). Here, when the length is counted in words, the slope is insignificantly negative, and when it is counted in characters, it is insignificantly positive.

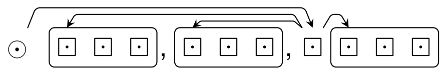
2.2. Argument

PW23 consider 4 dependency approaches to coordination, two asymmetric (which they call Stanford and Moscow) and two symmetric (Prague and London). We illustrate their argument on 2 of those: the asymmetric Stanford approach utilized in Universal Dependencies (UD; `https://universaldependencies.org/`; Nivre et al. 2016; de Marneffe et al. 2021; Zeman et al. 2022), and the symmetric approach utilized in Prague Dependency Treebanks (`https://ufal.mff.cuni.cz/prague-dependency-treebank`; Hajič et al. 2006).[1] These 2 approaches are schematically presented in (1)–(2), where the governor is marked as ⊙, tokens within the coordination as ⊡, and tokens within each conjunct are grouped.

(1) **Stanford ("bouquet", used in UD)**:



(2) **Prague (conjunction-headed)**:



PW23 argue that, given the principle of Dependency Length Minimization (DLM), only symmetric approaches, such as Prague (or London), are compatible with the empirical findings in Figure 1, while asymmetric approaches, such as Stanford (or Moscow), are not. DLM is the robustly demonstrated tendency for speakers to produce structures with maximally local dependencies, i.e., structures that minimize aggregate dependency lengths; see, e.g., Futrell et al. 2020 and references therein. For example, in a binary coordination with the governor on the left, DLM predicts that there should be a tendency for the shorter conjunct to be on the left, whether one assumes an asymmetric or a symmetric approach. If one assumes the asymmetric Stanford approach, placing the shorter conjunct on the left will minimize the aggregate dependency length by the difference of conjunct lengths, as illustrated in (3a–b). Assuming the symmetric Prague approach, the gain is even greater: it is twice the absolute difference between conjunct lengths – see (4a–b).[2] Both approaches also predict the same tendency when there is no governor – see (3c–d) and (4c–d). In both cases the gain is the same: the absolute length difference between the two conjuncts. So

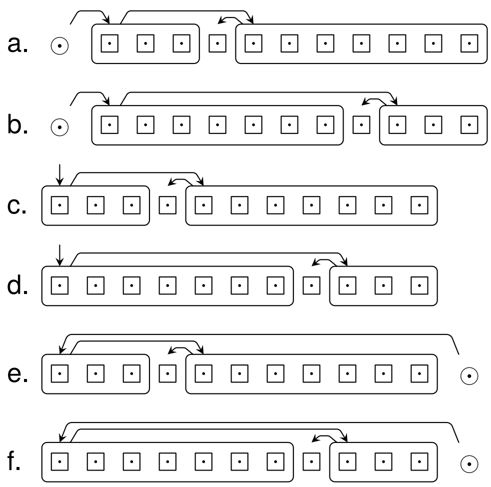
[1]The symmetric London approach is discussed in §5.

[2]This reasoning is based on the observation that, in English, heads of both conjuncts are on average situated the same short distance from the left periphery.

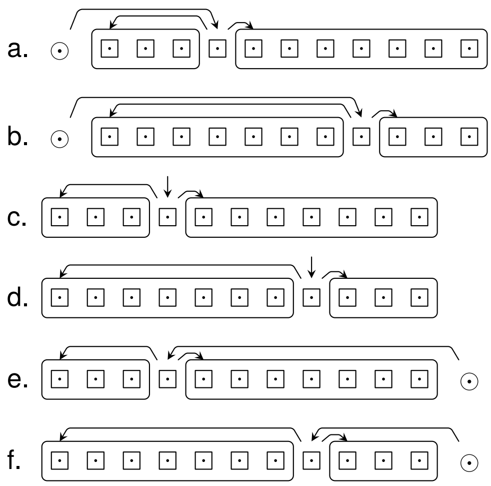
far, both approaches seem to be roughly compatible with Figure 1.

However, the predictions of the two approaches differ when the governor is on the right. On the Stanford approach, the aggregate dependency length is still minimized when the left conjunct is shorter – see (3e–f). Hence, the third row in Figure 1 should not differ considerably from the first two, contrary to fact. On the other hand, on the Prague approach, the aggregate dependency length does not depend on which conjunct is on the left – see (4e–f). This is compatible with the third row of Figure 1. By extending this reasoning to the Moscow and London approaches, and by considering DLM both at the level of use and at the level of grammaticalized conventions (see §5), PW23 argue that generally only symmetric approaches are compatible with the empirical observations.

(3) **Stanford**:



(4) **Prague**:



2.3. Limitations

Among the limitations of PW23, the one that we address here is that it is based on a relatively small and stylistically limited dataset, namely, on 21,825 coordinations extracted from $PTB_{\&}$, i.e., from a single newspaper (*Wall Street Journal*). Moreover, out of these 21,825 coordinations, only 4,719 have the governor on the right, of which only 1,754 have different lengths of conjuncts (in words). Further, only 405 of these 1,754 have absolute length difference greater than 3; e.g., there are only 34 such coordinations for the length difference of 7 and only 4 for the length difference of 15. Proportions calculated on the basis of such small samples would be highly unreliable. For this reason, PW23 grouped all observations of length differences greater than 3 words into only two further buckets (apart from the three buckets for the differences 1, 2, and 3): one for differences from 4 to 6, and the other for differences from 7 to 25 (and similarly for characters) – see the black boxes indicating these 5 buckets at the bottom of each plot in Figure 1. However, the resulting data is still scarce, as evidenced by wide confidence bands when the governor is on the right (see the third row of Figure 1): in this case, it is not clear whether the actual slope is positive, zero, or negative.

3. COCA Parsed with Stanza

In this replication study, instead of using a high-quality but small dataset, as in PW23, we used a large but low-quality dataset, namely, large parts of the Corpus of Contemporary American English (COCA; Davies 2008–2023)[3] automatically parsed with Stanza (Qi et al., 2020).[4] We parsed texts from 6 genres: newspapers, magazines, academic, fiction, blogs, and other web pages; for each genre, we included all texts from 17 years (1990, 1992, …, 2018, as well as 2001 and 2011) or – in the case of blogs and web texts, which have no date indication – 17 batches (numbered 02, 04, …, 34) of sizes slightly smaller than those yearly batches. The sizes of the data after cleaning (mainly removing bits with @ characters inserted by corpus distributors) are as follows:

genre	sentences	words
news	3,631,442	64,336,215
mag	3,758,381	66,715,341
acad	2,907,094	62,278,715
fic	4,973,112	67,487,228
blog	3,204,837	58,608,822
web	3,296,553	60,997,250
total	21,771,419	380,423,571

From dependency trees produced by Stanza, we extracted information about 11,502,053 coordinations, including coordinations with more than

[3] https://www.english-corpora.org/coca/
[4] https://stanfordnlp.github.io/stanza/

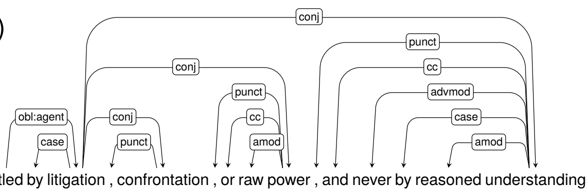
two conjuncts (in such cases, we took into account lengths of the first and the last conjunct).[5]

To this end, all texts were first split into sentences with Trankit (Nguyen et al., 2021),[6] as it seems more reliable than Stanza in this respect, then parsed with Stanza, using the following processors in the pipeline: `tokenize` (with sentence segmentation disabled), `lemma`, `pos`, `depparse`, `ner`.

Stanza returns a dependency tree of a sentence that is supposed to follow Universal Dependencies, although sometimes the resulting tree violates various UD conventions (e.g., it may contain conj dependencies directed to the left). Coordinations were identified in these trees by looking for relations labelled conj. An attempt was made to minimize the number of incorrectly parsed coordinations by detecting and excluding conj edges directed to the left or attached to punctuation marks.

As some UD representations of coordination are ambiguous between flat and nested coordinations (see Przepiórkowski and Patejuk 2019), an attempt was also made to disambiguate such representations: if there were two different conjunctions in one apparent coordination, this coordination was treated as containing a nested coordination, otherwise it was treated as a single flat coordination. For example, in sentence (5), whose partial structure produced by Stanza is given in (6), *litigation* has 3 outgoing conj dependencies: to *confrontation*, to *power*, and to *understanding*. This representation is in principle ambiguous between a flat coordination with 4 conjuncts and various nested coordinations. However, as one of the conjuncts, *power*, has the conjunction *or* attached to it, and another, *understanding*, has a different conjunction, *and*, nested coordinations are detected here: a binary *and*-coordination with a nested ternary *or*-coordination (see the brackets in (5)).

(5) It leaves the sense that all disputes must be settled **[[**by [litigation, confrontation, *or* raw power]**]**, *and* **[**never by reasoned understanding that may lead to honorable compromise or even enlightenment**]]**.

(6)


Once coordinations were identified, the next step was to identify the exact extents of the first and last conjuncts (crucial for counting and comparing conjunct lengths). For example, there is a case dependency from *litigation* to *by* in (6), suggesting that the first conjunct in the *or*-coordination is *by litigation* (2 words, 13 characters), while in reality *by* is a shared dependent of all 3 conjuncts, so the first separate conjunct is actually *litigation* alone (1 word, 10 characters). Hence, the following heuristics were employed for extracting first conjuncts.[7] First, all dependents directly to the right of the conjunct head were considered private to that conjunct, and similarly for compound dependents to the left. However, if a dependency to the left of the head had any other label, it was checked whether any other conjuncts had a dependency with the same label. If so, these dependencies were considered private to the particular conjuncts; e.g., if two or more conjuncts had an nsubj dependency, this was treated as coordination of clauses. On the other hand, if only the first conjunct had a given dependency type, it was treated as shared by all conjuncts; e.g., if this shared dependency was nsubj, this was treated as coordination of subsentential constituents (verb phrases). According to these heuristics, the leftwards case dependency of *litigation* is taken to be shared by all 3 conjuncts of the *or*-coordination, as none of the other 2 conjuncts of this coordinate structure has the same dependency type. On the other hand, *understanding*, the head of the last conjunct of the *and*-coordination, also has a case dependency, so it is taken to be private to the first conjunct in the *and*-coordination (cf. the brackets in bold in (5)).

After identifying all private dependencies belonging to the heads of conjuncts, all of the dependencies of those dependencies were identified recursively in order to find complete conjuncts. Then some punctuation marks were removed from the beginning of the conjuncts – commas, colons, semi-colons, dashes and double dashes. Finally, the text of each conjunct was found by taking the fragment of the sentence from the starting character of the first token in the conjunct to the ending character of the last token of the conjunct.

4. Results

For the statistics based on lengths measured in words, we took into consideration coordinations with absolute length differences between the first and the last conjunct in the range of 1 to 15 words; for lengths measured in characters, we inspected differences in the range of 1 to 60 characters.[8] In

[5]The vast majority (86.4%) of coordinations were binary, and removing non-binary coordinations does not affect the tendencies and conclusions reported below.

[6]`https://trankit.readthedocs.io/`

[7]No such heuristics were needed for the last conjuncts, whose all dependents (apart from cc) were treated as private.

[8]Just like PW23, we also performed all calculations for lengths measured in syllables, but the results were

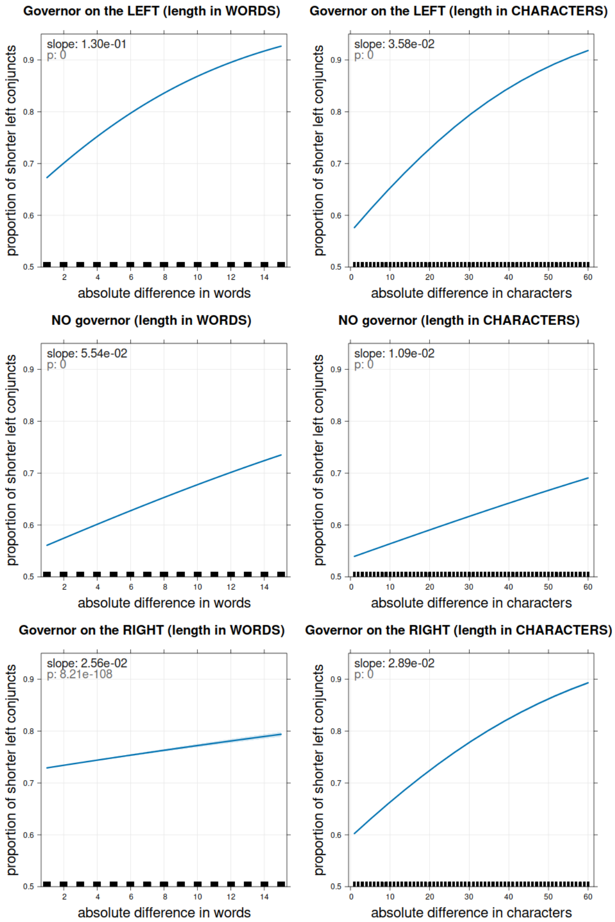


Figure 2: Modelled proportions of coordinations in COCA with left conjuncts shorter, depending on the absolute difference of conjunct lengths, with (extremely thin) confidence bands.

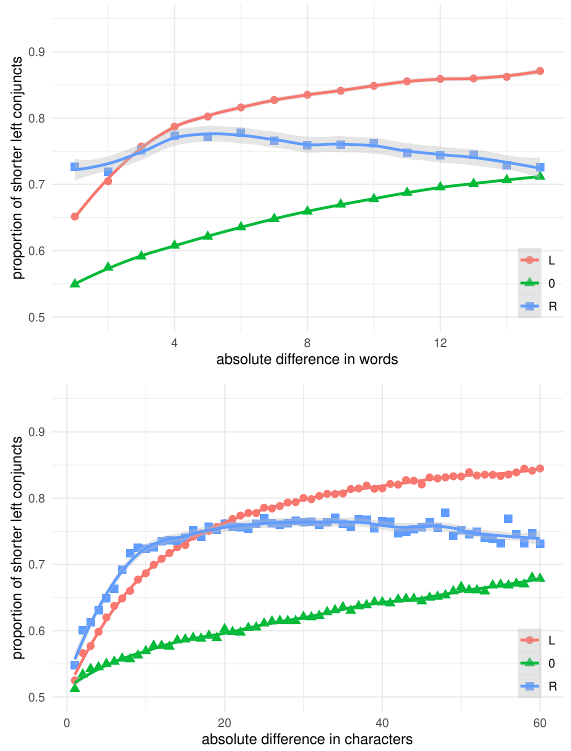


Figure 3: Observed and loess-smoothed proportions of coordinations in COCA with left conjuncts shorter, depending on the absolute difference of conjunct lengths.

these ranges, there were between around 1000 and over 310,000 observations for each combination of governor position (left, absent, right) $\times$ length difference (1–15 words / 1–60 characters). For example, there were 1790 coordinations with the governor on the right and the length difference of exactly 15 words. In 1299 the first conjunct was shorter than the last and in 491 the last conjunct was shorter than the first, resulting in the likelihood of the first conjunct being shorter when the governor is on the right and the difference is 15 words equal to $1299/(1299+491) \approx 0.726$.

Figure 2 summarizes our results in a way analogous to how Figure 1 summarizes PW23, i.e., via logistic regression. The first two rows are analogous to those in Figure 1 and are compatible with both symmetric and asymmetric approaches. However, the third row, corresponding to the governor on the right, differs considerably from the third row of Figure 1: both slopes are highly significantly positive and, moreover, when length difference is measured in characters, the slope (2.89×10^{-2}) is much steeper than when the governor is absent (1.09×10^{-2}). This result seems incompatible with symmetric theories of coordination and, thus, it seems to contradict PW23's findings.

However, a more detailed examination of the data shows that (linear) logistic regression models are not immediately appropriate for the data at hand. This is made conspicuous in Figure 3, which presents the observed proportions of shorter left conjuncts independently for each combination of governor position (L: left, see red circles; 0: absent, green triangles; R: right, blue squares) $\times$ length difference (1–15 words / 1–60 characters). Recall that for each such combination there is a sufficient number of observations (usually many thousands) to reliably calculate such proportions.

As can be seen in Figure 3, the proportion of shorter left conjuncts grows monotonically with length difference measured in words (the upper plot) when the governor is on the left or absent, and similarly for length difference measured in characters (the lower plot). However, when the governor is on the right, the proportion of shorter left conjuncts increases initially, up to 4 words or 20 characters, but then it plateaus and eventually decreases. This initial increase is especially pro-

similar to those for lengths measured in words and characters, so we do not report them here.

nounced when lengths are measured in characters, and it is responsible for the surprisingly steep slope in the corresponding plot in Figure 2. When this initial growth is discarded, the results of logistic regression – presented in Figure 4 – are similar to PW23's results in Figure 1. However, while Figure 1 does not make it clear whether the actual slopes are negative, zero, or positive when the governor is on the right (recall the wide confidence bands), here the slopes are very significantly negative ($p < 0.01$), especially when length is measured in words ($p \ll 0.001$).

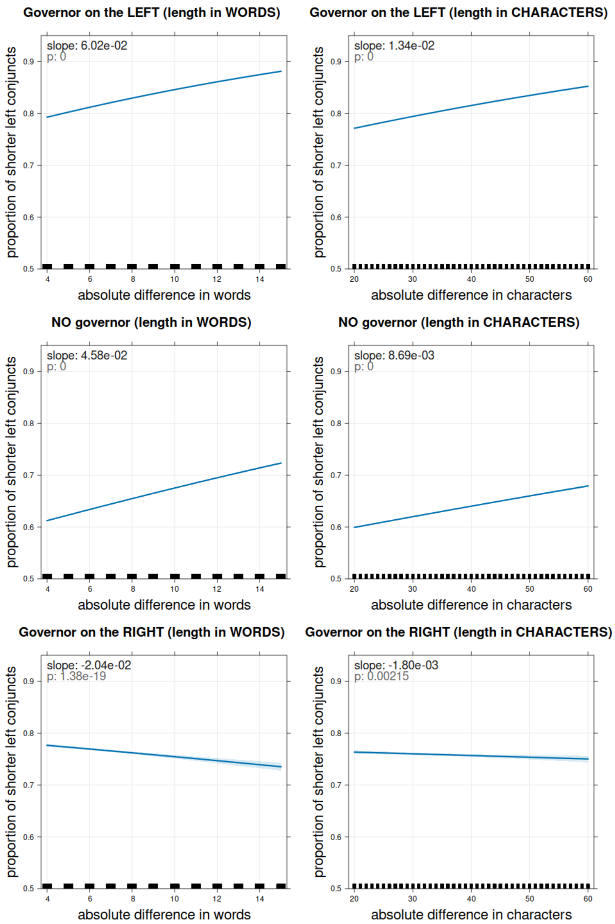


Figure 4: Modelled proportions of coordinations in COCA with left conjuncts shorter, depending on the absolute difference of conjunct lengths, with (very thin) confidence bands, for length differences of at least 4 words or 20 characters.

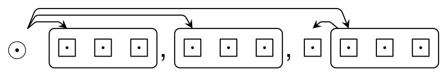
5. Discussion

Let us for a moment ignore the initial growth of proportions of shorter left conjuncts when the governor is on the right; we return to this issue at the end of this section.

As argued in PW23, the fact that the slopes are significantly positive when the governor is on the left or absent, but not when it is on the right, is incompatible with asymmetric theories of coordination: such theories predict that all slopes should be significantly positive regardless of the position of the governor. On the other hand, the Prague symmetric representation predicts that, when the governor is on the right, the slope should be close to zero, as the aggregate dependency length does not depend on which conjunct is shorter (recall (4e–f)). This prediction is compatible with the findings of PW23 (see again the bottom row of Figure 1), but it is not immediately compatible with the empirical findings of the current study, which show that the relevant slopes are significantly negative (disregarding the initial growth; see again the bottom row of Figure 4).

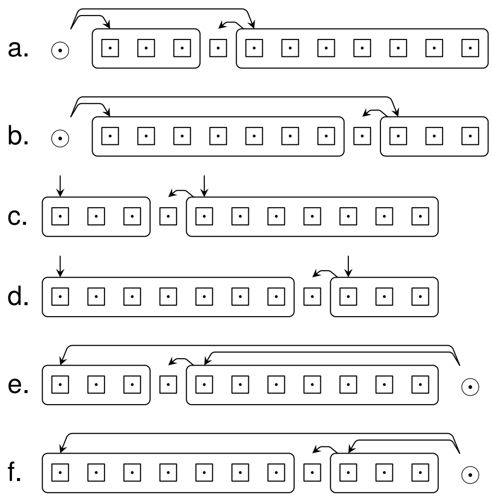
It turns out that these empirical observations are more directly compatible with another symmetric approach, which PW23 dub London, assumed in Word Grammar (Hudson, 1984); see (7).

(7) **London (multi-headed)**:



On this approach, shorter first conjuncts are still preferred when the governor is on the left (see (8a–b)), but shorter *last* conjuncts are preferred when it is on the right – the more so, the larger the length difference (see (8e–f)). This corresponds well to the empirical observations summarized in the top and bottom rows of Figure 4.

(8) **London**:



However, the London approach is not immediately compatible with the behaviour of coordinations with no governor, as it seems to predict no preference as to the placement of the shorter conjunct (see the multi-rooted (8c–d)). That is, it seems that on this approach the middle row of Figure 4 should contain roughly flat lines, instead of lines with highly significantly positive slope.

PW23 claim that their empirical findings, which also demonstrate the growing tendency for shorter conjuncts to occur on the left when there is no governor (see the middle row of Figure 1), can be made compatible with the London approach on the assumption, defended for example in Hawkins 1994 and Futrell et al. 2020, that DLM may work both at the level of use (as assumed above) and at the level of grammar. They illustrate DLM at the level of grammar with verbal constructions involving NP and PP dependents:[9] "when an NP… and a PP are both dependents of a verb V, the [V NP PP] order incurs shorter dependency lengths than the [V PP NP] order on average, given that NPs are on average shorter than PPs. Hawkins (1994, 90) argues that this tendency is conventionalized: present in grammar, not in use. The reason for this claim is that there is a strong preference for this order not only when the NP is shorter than the PP, but also when they are of similar lengths (e.g., *I sold [my mother's ring] [for five dollars]* vs. *I sold [for five dollars] [my mother's ring]*). However, this convention may be overridden in use, when length differences become large (e.g., *I sold [for five dollars] [my mother's silver engagement ring that she got from my father]* is more natural), again in compliance with DLM."

PW23 hypothesize that such a conventionalized rule is also at play in coordinations: given that in most coordinate structures the governor is on the left,[10] the at-use tendency for left conjuncts to be shorter, present in the majority of coordinations, got conventionalized into an at-grammar tendency for left conjuncts to be shorter in coordinations in general. If so, this at-grammar pressure could explain the growing proportions of shorter left conjuncts when there is no governor, even though there is no at-use incentive for left conjuncts to be shorter. Moreover, this could also explain why the growth is even more dynamic when the governor is on the left (then the slope is 6.02×10^{-2} if length is measured in words, vs. 4.58×10^{-2} when there is no governor, and similarly if it is measured in characters), as then both at-grammar and at-use pressures are at play and converge. Finally, in the case of governor on the right, proportions of shorter left conjuncts diminish with length difference, as the at-use working of DLM, which requires the last conjunct to be shorter, trumps the conventionalized rule, which would prefer the first conjunct to be shorter. This is fully analogous to how *I sold [for five dollars] [my mother's silver engagement ring that she got from my father]*, which satisfies the at-use preference for the shorter PP dependent to be closer to the head V, is more natural despite violating the at-grammar convention that the [V NP PP] order is preferred to [V PP NP].

To summarize, the significantly negative slopes witnessed in the bottom row of Figure 4 are dramatically at odds with asymmetric approaches to coordination, which instead predict a significantly positive slope, somewhat at odds with the symmetric Prague approach, which predicts a roughly horizontal line, but they are compatible with the symmetric London approach, which, however, requires invoking the general grammaticalized convention that – other things being equal – shorter conjuncts are preferred to be realized as initial conjuncts. This way the current study sharpens the results of PW23: it not only confirms their argument against asymmetric approaches to coordination, but also favours one symmetric approach (London) over another (Prague).

Let us finally return to the issue of the initial growth of proportions of shorter left conjuncts when the governor is on the right (see again Figure 3). One initially plausible explanation could be based on the common assumption that at-use DLM reflects imperfect working memory: syntactic dependencies are minimized as, over time, the memory of specific words used fades. This in turn makes it more difficult to integrate into syntactic structure new words bearing syntactic relations to such distant words (see, e.g., Futrell and Levy 2017). Arguably, when the length difference between conjuncts is very small, this working memory effect is not visible, so the at-grammar convention of placing shorter conjuncts first wins, even when the governor is on the right. However, as the length difference becomes greater, the at-use DLM pressure becomes more pronounced, winning over the at-grammar convention.

Unfortunately, this explanation is unlikely to be correct. If it were correct, the initial growth in coordinations with no governor, where the at-grammar convention is not counterbalanced by the at-use pressure, should be even stronger, contrary to fact; see, especially, the bottom plot in Figure 3.

The fact that a strong initial growth is observed in both cases where there is a governor (on the left or on the right), but not when there is no governor, suggests another explanation. As noted in PW23, coordinate structures without a governor are categorially relatively homogeneous: almost all are coordinations of (often long) clauses and (often longish) verb phrases. By contrast, over 60% of coordinations with a governor are coordinations of nominal constituents and – when the governor is on the right – further 15% are coordinations of adjectival constituents. Such coordinations are often

[9] NP: nominal phrase, PP: prepositional phrase.

[10] 55% in the COCA dataset processed here, while 30% have no governor, and 15% have the governor on the right. In the dataset of PW23, the corresponding proportions are 60% (left), 18% (absent), 22% (right).

relatively short, e.g., *salt and pepper*, *first and foremost*, etc. Coordinations of short constituents of these kinds, so-called binomials, have been extensively studied and it has been demonstrated that constituent length is one of the most important factors influencing the ordering of conjuncts (see, e.g., Benor and Levy 2006, Lohmann 2014, and references therein).[11] In the case of such short constituents, it is unlikely that the strong preference for shorter left conjuncts is caused by memory limitations; it is more likely that such a preference is yet another grammaticalized convention, applicable to short coordinations, with relatively small length differences. If so, this convention may be responsible for the initial dynamic growth of proportions of shorter left conjuncts in coordinations with a governor (on the left or on the right). As this convention applies only to short conjuncts, with small length differences, its effect vanishes around the difference of 10–20 characters (or 3–4 words), and only the memory-driven at-use DLM is operational for greater length differences, resulting in flatter but still significantly positive slope in the case of coordinations with the governor on the left and in significantly negative scope when the governor is on the right, as predicted by the symmetrical London representation of coordinate structures. (We leave an empirical verification of this hypothesis for future work.)

6. Evaluation and Limitations

The current study overcomes 2 limitations of PW23, namely, the scarcity of data and lack of genre variation: it is based on over 11.5 million coordinations (as opposed to less than 22 thousand in PW23) and on 6 genres (as opposed to a single newspaper in PW23); a more detailed analysis of particular genres is presented in Appendix A. On the other hand, this study shares with PW23 another limitation that should be addressed in subsequent studies: being based solely on English.

An important limitation specific to the current study is that the results reported above are derived from a dataset that is based on automatic – error-prone – parsing of COCA and extraction of coordinations from the resulting parses. There is no reason to believe that the differences in how proportions of shorter left conjuncts change with length differences and the presence and position of the governor reflect systematic errors in these two procedures; the null hypothesis is that the parser makes errors whether the governor is on the left, absent, or on the right, and that perhaps it makes more errors in the case of larger length differences (which are positively correlated with longer coordinations involving more complex conjuncts). What we do not expect is that the parser errs in such a way that the proportions of shorter left conjuncts seem to be decreasing, while in reality they are increasing, and that such errors only distort the statistics when the governor is on the right, and not when it is on the left or absent. However, we cannot – with absolute confidence – rule out such an effect, so we performed an evaluation allowing us to estimate the quality of the dataset.

From the over 11.5 million automatically parsed and extracted coordinations, 15 coordinations were chosen at random for each of the 3 values of governor position (left, absent, right) and for each of 20 values of length difference in words (from 1 to 20), resulting in $15 \times 3 = 45$ coordinations for each length difference, $45 \times 20 = 900$ coordinations altogether. Two raters evaluated each extracted coordination, marking it as "good" when 1) both extracted conjuncts are the first and last conjuncts in a coordination occurring in the pragmatically most likely – not just any – parse of the given sentence, and 2) the automatically established position of the governor is as in this most likely parse; the statistics reported above crucially rely on exactly these data. Out of 900 coordinations, the raters initially provided the same ratings to 739, resulting in substantial agreement ($\kappa = 0.64$). Subsequently, the raters discussed and agreed on the ratings of all coordinations.

Unfortunately, only 451 (50.1%) of these 900 coordinations were rated as "good" in the above sense. As expected, there was a general – but relatively minor – negative correlation between the length difference and the proportion of "good" coordinations. Because of the large proportion of "bad" coordinations, we attempted to estimate whether the dynamics of proportions of shorter left conjuncts was the same in coordinations rated as "good" and in those rated as "bad". However, despite the relatively large number of manually evaluated extracted coordinations, this number was still too small to make such a reliable estimation possible: there were only 15 coordinations for each governor position $\times$ length difference pair, only half of those were "good" on average, and the observed proportion of shorter left conjuncts in a sample of 7 or 8 coordinations cannot be used to estimate the true proportion in the whole population.

In summary, data quality is a major limitation of the current study, with a great many parsed and extracted coordinations rated as "bad", but there is no reason to believe that errors introduced by this procedure 1) make an actually increasing tendency seem decreasing and 2) have this effect only when the governor is on the right. Rather, until demon-

[11] Semantic factors, e.g., preferring more animate conjuncts first (as in *chicken or egg* rather than *egg or chicken*), are stronger, but they are applicable to many fewer coordinations.

strated otherwise, the null hypothesis must be that parsing and extraction errors create random noise which does not considerably affect the observed tendencies: increasing proportions when the governor in on the left or absent, but decreasing (from a point) when it is on the right.

7. Conclusion

The current study replicates and sharpens the findings of Przepiórkowski and Woźniak (2023) and confirms their argument against asymmetric approaches to coordinations, such as used in basic Universal Dependencies[12] and in some prominent theoretical linguistic approaches. Considering symmetric approaches, the current findings are more compatible with the multi-headed London approach than with the conjunction-headed Prague approach.

The advantage of the current study over that in PW23 is data size – 11.5 million vs. less than 22 thousand coordinations – as well as the variety of data. Because of the low quality of the input data, the current study *alone* would not constitute a strong argument against asymmetric and for symmetric approaches to coordination. Nevertheless, even though it is based on a very different dataset than that used in PW23, this study to a large extent replicates the empirical findings of PW23 and supports their general conclusions that only (some of) the symmetric approaches to coordination are compatible with the well-established principle of Dependency Length Minimization. Needless to say, further replication studies, also based on languages other than English, are needed to make these results even more robust.

8. Bibliographical References

Sarah B. Benor and Roger Levy. 2006. The chicken or the egg? A probabilistic analysis of English binomials. *Language*, 82(2):233–278.

Mark Davies. 2008–2023. The Corpus of Contemporary American English (COCA). Available online at https://www.english-corpora.org/coca/.

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

[12]This is not meant to be a criticism of UD, which is known to make some linguistically controversial decisions (see, e.g., Osborne and Gerdes 2019, as well as Przepiórkowski and Patejuk 2018) in the interest of reaching "a very subtle compromise between a number of competing criteria" (de Marneffe et al., 2021, 302), which include typological uniformity, comprehensibility by non-linguists, and suitability for computational tasks.

Jessica Ficler and Yoav Goldberg. 2016. Coordination annotation extension in the Penn Tree Bank. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 834–842, Berlin, Germany.

Richard Futrell and Roger Levy. 2017. Noisy-context surprisal as a human sentence processing cost model. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2009): Volume 1, Long Papers*, pages 688–698, Valencia, Spain.

Richard Futrell, Roger P. Levy, and Edward Gibson. 2020. Dependency locality as an explanatory principle for word order. *Language*, 96(2):371–412.

Jan Hajič, Jarmila Panevová, Eva Hajičová, Petr Sgall, Petr Pajas, Jan Štěpánek, Jiří Havelka, Marie Mikulová, Zdeněk Žabokrtský, Magda Ševčíková Razímová, and Zdeňka Urešová. 2006. Prague Dependency Treebank 2.0 (PDT 2.0).

John A. Hawkins. 1994. *A Performance Theory of Order and Constituency*. Cambridge University Press, Cambridge.

Richard Hudson. 1984. *Word Grammar*. Blackwell, Oxford.

John P. A. Ioannidis. 2005. Why most published research findings are false. *PLoS Medicine*, 2(8):0696–0701.

Arne Lohmann. 2014. *English Coordinate Constructions: A Processing Perspective on Constituent Order*. Cambridge University Press, London.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19:313–330.

Alan Boag Munn. 1993. *Topics in the Syntax and Semantics of Coordinate Structures*. Ph.D. dissertation, University of Maryland.

Ad Neeleman, Joy Philip, Misako Tanaka, and Hans van de Koot. 2023. Subordination and binary branching. *Syntax*, 26(1):41–84.

Minh Van Nguyen, Viet Lai, Amir Pouran Ben Veyseh, and Thien Huu Nguyen. 2021. Trankit: A light-weight transformer-based toolkit for multilingual natural language processing. In *Proceedings of the 16th Conference of the European*

Chapter of the Association for Computational Linguistics: System Demonstrations*.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal Dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation, LREC 2016*, pages 1659–1666, Portorož, Slovenia. European Language Resources Association (ELRA).

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(17).

Adam Przepiórkowski and Agnieszka Patejuk. 2018. Arguments and adjuncts in Universal Dependencies. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018)*, pages 3837–3852, Santa Fe, NM. (Best position paper at COLING 2018).

Adam Przepiórkowski and Agnieszka Patejuk. 2019. Nested coordination in Universal Dependencies. In *Proceedings of the Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019)*, pages 58–69. Association for Computational Linguistics.

Adam Przepiórkowski and Michał Woźniak. 2023. Conjunct lengths in English, Dependency Length Minimization, and dependency structure of coordination. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15494–15512, Toronto, Canada. Association for Computational Linguistics.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A Python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108.

Lukas Sönning and Valentin Werner. 2021. The replication crisis, scientific revolutions, and linguistics. *Linguistics*, 59(5):1179–1206.

Daniel Zeman, Joakim Nivre, Mitchell Abrams, Elia Ackermann, Noëmi Aepli, Hamid Aghaei, Željko Agić, Amir Ahmadi, Lars Ahrenberg, Chika Kennedy Ajede, Gabrielė Aleksandravičiūtė, Ika Alfina, Avner Algom, Erik Andersen, Lene Antonsen, Katya Aplonova, Angelina Aquino, Carolina Aragon, Glyd Aranes, Maria Jesus Aranzabe, Bilge Nas Arıcan, Þórunn Arnardóttir, Gashaw Arutie, Jessica Naraiswari Arwidarasti, Masayuki Asahara, Deniz Baran Aslan, Cengiz Asmazoğlu, Luma Ateyah, Furkan Atmaca, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Keerthana Balasubramani, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Starkaður Barkarson, Rodolfo Basile, Victoria Basmov, Colin Batchelor, John Bauer, Seyyit Talha Bedir, Kepa Bengoetxea, Yifat Ben Moshe, Gözde Berk, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Agnė Bielinskienė, Kristín Bjarnadóttir, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Anouck Braggaar, Kristina Brokaitė, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Lauren Cassidy, Tatiana Cavalcanti, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Neslihan Cesur, Savas Cetin, Özlem Çetinoğlu, Fabricio Chalub, Shweta Chauhan, Ethan Chi, Taishi Chika, Yongseok Cho, Jinho Choi, Jayeol Chun, Juyeon Chung, Alessandra T. Cignarella, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Daniela Corbetta, Marine Courtin, Mihaela Cristescu, Philemon Daniel, Elizabeth Davidson, Mathieu Dehouck, Martina de Laurentiis, Marie-Catherine de Marneffe, Valeria de Paiva, Mehmet Oguz Derin, Elvis de Souza, Arantza Diaz de Ilarraza, Carly Dickerson, Arawinda Dinakaramani, Elisa Di Nuovo, Bamba Dione, Peter Dirix, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Hanne Eckhoff, Sandra Eiche, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Olga Erina, Tomaž Erjavec, Aline Etienne, Wograine Evelyn, Sidney Facundes, Richárd Farkas, Federica Favero, Jannatul Ferdaousi, Marília Fernanda, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Kazunori Fujita, Katarína Gajdošová, Daniel Galbraith, Federica Gamba, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Fabrício Ferraz Gerardi, Kim Gerdes, Filip Ginter, Gustavo Godoy, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Bernadeta Griciūtė, Matias Grioni, Loïc Grobol, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Tunga Güngör, Nizar Habash, Hinrik Hafsteinsson, Jan Hajič, Jan Hajič jr., Mika Hämäläinen, Linh Hà Mỹ, Na-Rae Han, Muhammad Yudistira

Hanifmuti, Takahiro Harada, Sam Hardwick, Kim Harris, Dag Haug, Johannes Heinecke, Oliver Hellwig, Felix Hennig, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Jena Hwang, Takumi Ikeda, Anton Karl Ingason, Radu Ion, Elena Irimia, Ọlájídé Ishola, Kaoru Ito, Siratun Jannat, Tomáš Jelínek, Apoorva Jha, Anders Johannsen, Hildur Jónsdóttir, Fredrik Jørgensen, Markus Juutinen, Sarveswaran K, Hüner Kaşıkara, Andre Kaasen, Nadezhda Kabaeva, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Neslihan Kara, Ritván Karahóǧa, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Elena Klementieva, Elena Klyachko, Arne Köhn, Abdullatif Köksal, Kamil Kopacewicz, Timo Korkiakangas, Mehmet Köse, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Parameswari Krishnamurthy, Sandra Kübler, Oğuzhan Kuyrukçu, Aslı Kuzgun, Sookyoung Kwak, Veronika Laippala, Lucia Lam, Lorenzo Lambertino, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phuong Lê Hồng, Alessandro Lenci, Saran Lertpradit, Herman Leung, Maria Levina, Cheuk Ying Li, Josie Li, Keying Li, Yuan Li, KyungTae Lim, Bruna Lima Padovani, Krister Lindén, Nikola Ljubešić, Olga Loginova, Stefano Lusito, Andry Luthfi, Mikko Luukko, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Menel Mahamdi, Jean Maillard, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Büşra Marşan, Cătălina Mărănduc, David Mareček, Katrin Marheinecke, Stella Markantonatou, Héctor Martínez Alonso, Lorena Martín Rodríguez, André Martins, Jan Mašek, Hiroshi Matsuda, Yuji Matsumoto, Alessandro Mazzei, Ryan McDonald, Sarah McGuinness, Gustavo Mendonça, Tatiana Merzhevich, Niko Miekka, Karina Mischenkova, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Maria Mitrofan, Yusuke Miyao, AmirHossein Mojiri Foroushani, Judit Molnár, Amirsaeid Moloodi, Simonetta Montemagni, Amir More, Laura Moreno Romero, Giovanni Moretti, Keiko Sophie Mori, Shinsuke Mori, Tomohiko Morioka, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Robert Munro, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Mariam Nakhlé, Juan Ignacio Navarro Horñiacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Manuela Nevaci, Luong Nguyễn Thị, Huyền Nguyễn Thị Minh, Yoshihiro Nikaido, Vitaly Nikolaev, Rattima Nitisaroj, Alireza Nourian, Hanna Nurmi, Stina Ojala, Atul Kr. Ojha, Adédayọ Olúòkun, Mai Omura, Emeka Onwuegbuzia, Noam Ordan, Petya Osenova, Robert Östling, Lilja Øvrelid, Şaziye Betül Özateş, Merve Özçelik, Arzucan Özgür, Balkız Öztürk Başaran, Teresa Paccosi, Alessio Palmero Aprosio, Hyunji Hayley Park, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Giulia Pedonese, Angelika Peljak-Łapińska, Siyao Peng, Cenel-Augusto Perez, Natalia Perkova, Guy Perrier, Slav Petrov, Daria Petrova, Andrea Peverelli, Jason Phelan, Jussi Piitulainen, Tommi A Pirinen, Emily Pitler, Barbara Plank, Thierry Poibeau, Larisa Ponomareva, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyysalo, Peng Qi, Andriela Rääbis, Alexandre Rademaker, Mizanur Rahoman, Taraka Rama, Loganathan Ramasamy, Carlos Ramisch, Fam Rashel, Mohammad Sadegh Rasooli, Vinit Ravishankar, Livy Real, Petru Rebeja, Siva Reddy, Mathilde Regnault, Georg Rehm, Ivan Riabov, Michael Rießler, Erika Rimkutė, Larissa Rinaldi, Laura Rituma, Putri Rizqiyah, Luisa Rocha, Eiríkur Rögnvaldsson, Mykhailo Romanenko, Rudolf Rosa, Valentin Roșca, Davide Rovati, Ben Rozonoyer, Olga Rudina, Jack Rueter, Kristján Rúnarsson, Shoval Sadde, Pegah Safari, Benoît Sagot, Aleksi Sahala, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Ezgi Sanıyar, Dage Särg, Baiba Saulīte, Yanin Sawanakunanon, Shefali Saxena, Kevin Scannell, Salvatore Scarlata, Nathan Schneider, Sebastian Schuster, Lane Schwartz, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Syeda Shahzadi, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Yana Shishkina, Muh Shohibussirri, Dmitry Sichinava, Janine Siewert, Einar Freyr Sigurðsson, Aline Silveira, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Maria Skachedubova, Aaron Smith, Isabela Soares-Bastos, Shafi Sourov, Carolyn Spadine, Rachele Sprugnoli, Vivian Stamou, Steinþór Steingrímsson, Antonio Stella, Milan Straka, Emmett Strickland, Jana Strnadová, Alane Suhr, Yogi Lesmana Sulestio, Umut Sulubacak, Shingo Suzuki, Daniel Swanson, Zsolt Szántó, Chihiro Taguchi, Dima Taji, Yuta Takahashi, Fabio Tamburini, Mary Ann C. Tan, Takaaki Tanaka, Dipta Tanaya, Mirko Tavoni, Samson Tella, Isabelle Tellier, Marinella Testori, Guillaume Thomas, Sara Tonelli, Liisi Torga, Marsida Toska, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Utku Türk, Francis Tyers, Sumire Uematsu, Roman Untilov, Zdeňka Urešová, Larraitz Uria, Hans Uszkoreit, Andrius Utka, Elena Vagnoni, Sowmya Vajjala, Rob van der Goot, Martine Vanhove, Daniel van Niekerk, Gertjan

van Noord, Viktor Varga, Uliana Vedenina, Eric Villemonte de la Clergerie, Veronika Vincze, Natalia Vlasova, Aya Wakasa, Joel C. Wallenberg, Lars Wallin, Abigail Walsh, Jing Xian Wang, Jonathan North Washington, Maximilan Wendt, Paul Widmer, Shira Wigderson, Sri Hartati Wijono, Seyi Williams, Mats Wirén, Christian Wittern, Tsegay Woldemariam, Tak-sum Wong, Alina Wróblewska, Mary Yako, Kayo Yamashita, Naoki Yamazaki, Chunxiao Yan, Koichi Yasuoka, Marat M. Yavrumyan, Arife Betül Yenice, Olcay Taner Yıldız, Zhuoran Yu, Arlisa Yuliawati, Zdeněk Žabokrtský, Shorouq Zahra, Amir Zeldes, He Zhou, Hanzhi Zhu, Anna Zhuravleva, and Rayan Ziane. 2022. Universal Dependencies 2.11. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Niina Ning Zhang. 2009. *Coordination in Syntax*. Cambridge University Press, Cambridge.

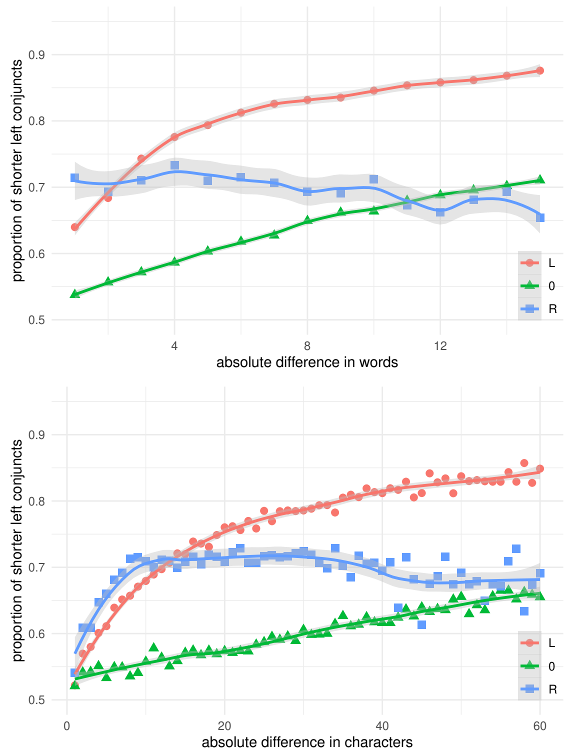


Figure 5: Observed and loess-smoothed proportions of coordinations in COCA **news** with left conjuncts shorter, depending on the absolute difference of conjunct lengths.

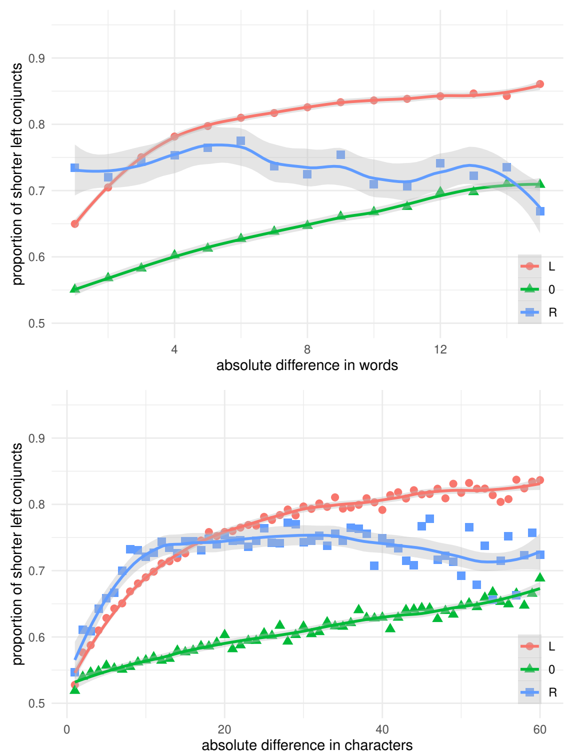


Figure 6: Observed and loess-smoothed proportions of coordinations in COCA **mag** with left conjuncts shorter, depending on the absolute difference of conjunct lengths.

A. Appendix: Genres

Figures 5–10 present the dynamics of proportions of shorter left conjuncts separately for each of the 6 genres included in the current study: newspapers, magazines, academic writings, fiction, blogs, and other web pages. All of them confirm that, when the governor is on the left or absent, these proportions grow monotonically with the length difference between conjuncts. However, not all of them show clearly that – after an initial increase – these proportions decrease when the governor is on the right: this effect is clearly visible in newspapers and magazines (Figures 5–6), less clearly in academic writings (present when length is measured in words, but less clear when it is measured in characters), blogs, and other web pages (Figures 7–9), and it is almost absent in fiction, where the curves for the two positions of the governor, left and right, are very similar (Figure 10). We leave investigation of these differences for future work.

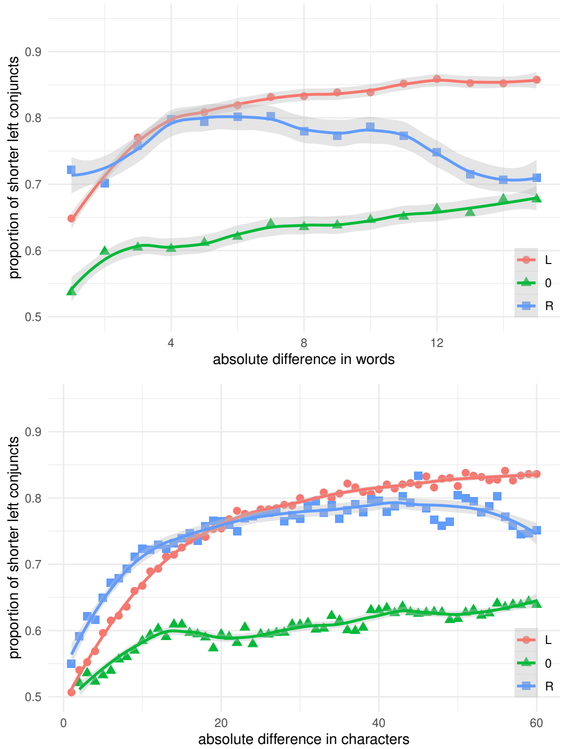


Figure 7: Observed and loess-smoothed proportions of coordinations in COCA **acad** with left conjuncts shorter, depending on the absolute difference of conjunct lengths.

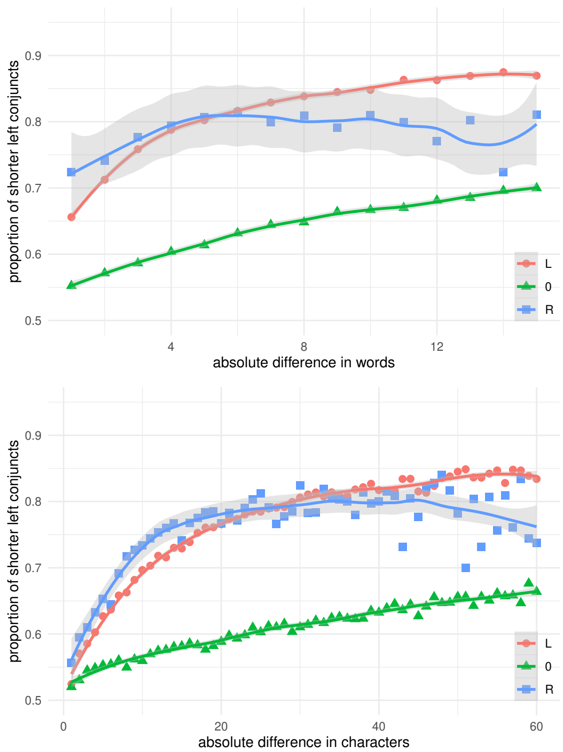


Figure 8: Observed and loess-smoothed proportions of coordinations in COCA **blog** with left conjuncts shorter, depending on the absolute difference of conjunct lengths.

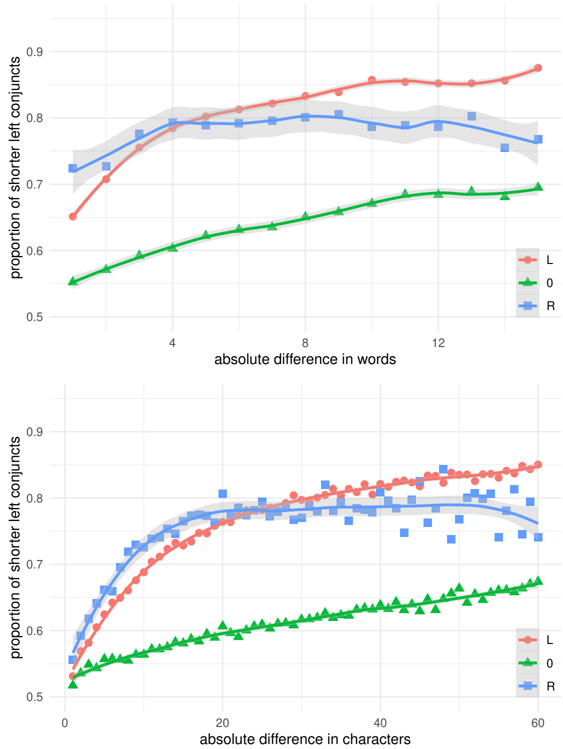


Figure 9: Observed and loess-smoothed proportions of coordinations in COCA **web** with left conjuncts shorter, depending on the absolute difference of conjunct lengths.

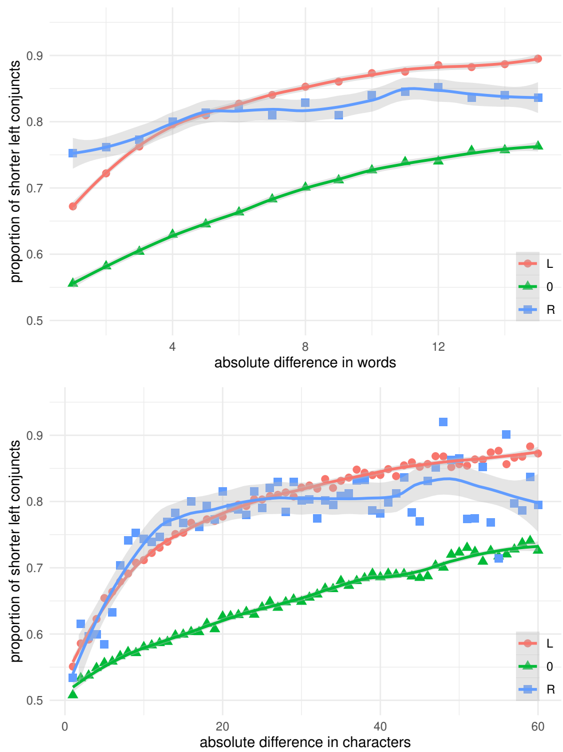


Figure 10: Observed and loess-smoothed proportions of coordinations in COCA **fic** with left conjuncts shorter, depending on the absolute difference of conjunct lengths.