Conjunct Lengths in English, Dependency Length Minimization, and Dependency Structure of Coordination

Adam Przepiórkowski
ICS Polish Academy of Sciences
and University of Warsaw
adamp@ipipan.waw.pl

Michał Woźniak
University of Warsaw
m.wozniak60@student.uw.edu.pl

Abstract

This paper confirms that, in English binary coordinations, left conjuncts tend to be shorter than right conjuncts, regardless of the position of the governor of the coordination. We demonstrate that this tendency becomes stronger when length differences are greater, but only when the governor is on the left or absent, not when it is on the right. We explain this effect via Dependency Length Minimization and we show that this explanation provides support for symmetrical dependency structures of coordination, as opposed to structures headed by the first conjunct.

1 Introduction

It has been observed for various particular types of coordination in English that left-most conjuncts tend to be shorter than right-most conjuncts (e.g., Gibson et al. 1996, Temperley 2005, Lohmann 2014). This is illustrated in (1) from the Penn Treebank (PTB; Marcus et al. 1993), where the left conjunct, *ship*, is shorter than the right conjunct, *hope I get paid*, in terms of the number of words (1 vs. 4), the number of syllables (1 vs. 4), and the number of characters (4 vs. 15, including spaces).

(1) I'm going *to* [[ship] and [hope I get paid]].

However, to the best of our knowledge, there are no demonstrations of this effect that would take all kinds of coordinations into account and that would use various length metrics. Filling this gap is the first contribution of this paper.

There is even less work that asks whether it is really the left-to-right order of conjuncts that matters here, or whether it is perhaps the closeness to the external head – henceforth, *governor*. (1), where the governor *to* is on the left, lends support to both – "leftness" and "closeness" – hypotheses. But the two hypotheses make different predictions when the governor is on the right, as in (2)–(3).

(2) [[Walter Sisulu] and [the African National Congress]] *came* home yesterday.

(3) [[My younger daughter] and [I]] *are* fine.

(2), where the left conjunct is shorter (2 vs. 4 words, 5 vs. 9 syllables, 13 vs. 29 characters), only supports the "leftness" hypothesis, while (3), where the right conjunct is shorter (1 vs. 3 words, 1 vs. 5 syllables, 1 vs. 19 characters), only supports the "closeness" hypothesis.

The second contribution of this paper is to establish two facts regarding the influence of the governor on conjunct lengths in English. The first fact is that left conjuncts tend to be shorter even when the governor is on the right, which immediately invalidates the "closeness" hypothesis. The second observation is more interesting and has important consequences for dependency theories of coordination: the position of the governor is crucial in how this tendency for left conjuncts to be shorter changes with differences in conjunct lengths. When the governor is on the left, this tendency becomes stronger with increasing length differences between conjuncts, but when the governor is on the right, this effect disappears.

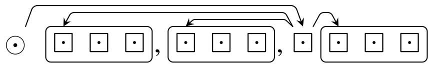
The third contribution of this paper is to provide an explanation of this effect in terms of Dependency Length Minimization (DLM) – the robustly demonstrated tendency for natural languages to strive for maximally local dependencies. Our explanation is compatible with the view that such distance minimization pressure is at work both at the level of use and at the level of grammar (Hawkins 1994, Futrell et al. 2020).

The final novel contribution is to demonstrate that this explanation is possible only on two of the four main linguistic approaches to coordination, namely, on the approaches schematically represented in (4)–(5), but not on the other two approaches schematized in (6)–(7).[1]

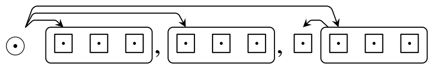
[1] In these diagrams, the governor is marked as ⊙, tokens within the coordination as ⊡, and tokens within each conjunct are grouped. Names of the approaches in (4) and (6)–(7) are based on those in Popel et al. 2013 and they

Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics
Volume 1: Long Papers, pages 15494–15512
July 9-14, 2023 ©2023 Association for Computational Linguistics

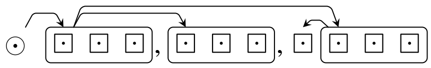
(4) **Conjunction-headed/Prague**:



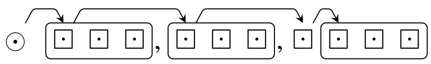
(5) **Multi-headed/London**:



(6) **Bouquet/Stanford**:



(7) **Chain/Moscow**:



We also show that DLM considerations favour the representation of coordination in the enhanced version of the current treebank annotation standard, Universal Dependencies (UD; `https://universaldependencies.org/`; Nivre et al. 2016, de Marneffe et al. 2021, Zeman et al. 2022), over the basic UD representation of coordination. Hence, the empirical results of this paper lend support to the issue of the appropriate dependency structure of coordination both in theoretical linguistics and in corpora.

2 Data

Work reported here is based on one of standard syntactic resources, the Penn Treebank. More precisely, we utilized the version of PTB, which we call $\text{PTB}_{\&}$, made available by Ficler and Goldberg (2016).[2] It incorporates earlier corrections of the internal structure of nominal phrases (Vadas and Curran 2007) and – importantly – it improves on PTB by offering explicit and relatively consistent information about coordinations (see Ficler and Goldberg 2016 for details). Unlike treebanks within Universal Dependencies, $\text{PTB}_{\&}$ makes extents of conjuncts unambiguous and, in particular, it explicitly marks shared dependents of conjuncts, but it does not explicitly mark heads of constructions, so also information about governors of coordinate structures is not explicitly available. However, it is relatively easy to construct rules finding governors. A simplified example of such a rule is: "if the mother of coordination c is a PP, try to locate a sister of c of category IN or TO (cf. (8)), or – failing that – of category VBG (cf. (9))".[3]

(8) Flesh goes to total alert *for* [[flight] or [fight]].

(9) The visitors then listed technologies up for sale, *including* [[launch services] and [propulsion hardware]].

The final evaluation of the program implementing these rules, performed on previously unseen random 100 coordinations, gave the 92% accuracy in locating a specific governor and, crucially, 97% in deciding whether the governor is absent, on the left, or on the right of the coordinate structure.[4]

Out of about 49,200 sentences (1.25M tokens) in $\text{PTB}_{\&}$, 19,095 contain at least one coordination, with the total of 24,446 coordinations. All coordinate structures (i.e., *–CCP nodes) with exactly two conjuncts – 21,825 altogether – where automatically extracted from $\text{PTB}_{\&}$, together with information about the location of the governor, if any.[5] There were 17,825 coordinations with a governor on the left (13,106, i.e., 73.5%) or on the right (4,719, i.e., 26.5%).

The length of each conjunct was measured as in §1: in characters (textual length), in syllables (as in, e.g., Benor and Levy 2006 and Lohmann 2014; an approximation of spoken length), and in words (as in, e.g., Gibson et al. 1996 and Temperley 2005; common in discussions of DLM). Implementing character and word metrics was straightforward. Syllable counting was done with the help of two Python libraries: Inflect, for converting numbers written with digits to words, and CMUdict, for

reflect where a given approach is conspicuously assumed: (4) – in the Prague Dependency Treebank (`https://ufal.mff.cuni.cz/prague-dependency-treebank`; Hajič et al. 2006), (6) – in the Stanford dependency parser (`https://nlp.stanford.edu/software/lex-parser.shtml`; de Marneffe et al. 2006), (7) – in the Meaning–Text Theory originally developed in Moscow (Mel'čuk 1974, 1988, 2009). In the same spirit, we call the approach in (5) "London", as it is often associated with Word Grammar (Hudson 2010, 1990, 1984: 225), developed at University College London.

[2] `https://github.com/Jess1ca/CoordinationExtPTB`

[3] PP stands for *prepositional phrase*, IN is the tag used in PTB for prepositions, TO – for various uses of the word *to*, and VBG – for gerunds.

[4] The gold standard for this evaluation was created by two annotators. Their initial Inter-Rater Agreement was $\kappa = 0.71$ for the task of locating a specific governor (or deciding that there is none) and $\kappa = 0.74$ for deciding on the presence and position of the governor with respect to the coordination. In order to create the gold standard, all differences in annotation where then discussed and resolved. Against this gold standard, values of precision, recall, and F_1 of the algorithm for the task of deciding that the governor is on the left were: $P = 0.98$, $R = 0.97$, $F_1 = 0.98$, on the right: $P = 0.94$, $R = 0.94$, $F_1 = 0.94$, and absent: $P = 0.95$, $R = 1.00$, $F_1 = 0.97$.

[5] For example, a matrix coordination of sentences has no governor, as in (i).

(i) [[It scared brokers], but [most survived].]

	median left	median right	mean left	mean right	V	p
All coordinations (N = 21,825)						
characters	15	20	26.68	32.34	7.4e+07	3.7e-262
syllables	5	7	8.35	9.85	6.5e+07	5.5e-171
words	3	3	4.42	5.36	3.2e+07	6.6e-234
No governor (N = 4,000)						
characters	49	56	54.78	64.66	3e+06	1.9e-33
syllables	15	16	16.20	18.88	2.8e+06	1.8e-28
words	8	10	9.23	11.00	2.4e+06	2.3e-40
Governor on the left (N = 13,106)						
characters	14	18	22.40	28.01	2.5e+07	8.2e-214
syllables	5	6	7.29	8.75	2.2e+07	6.5e-127
words	3	3	3.70	4.63	1.1e+07	1.9e-207
Governor on the right (N = 4,719)						
characters	9	10	14.76	16.94	3.3e+06	3.2e-49
syllables	3	4	4.64	5.24	2.6e+06	3.5e-35
words	1	1	2.35	2.60	5.6e+05	4.4e-23

Table 1: Medians and means of lengths of left and right conjuncts in binary coordinations in $PTB_\&$

looking up the number of syllables for particular words. Additional heuristics were implemented for tokens unknown to CMUdict (abbreviations, special symbols including $, etc.; see the Appendix).

3 Basic Statistics

Table 1 shows that, in binary coordinations in $PTB_\&$, left conjuncts tend to be shorter than right conjuncts. This is true about the whole population of binary coordinations, as well as about each of its three subpopulations: with no governor, with the governor on the left, and – crucially – with the governor on the right. As noted above, this last result immediately invalidates the "closeness" hypothesis. In each population and for any length unit, the median of left conjunct lengths is smaller than or equal to the median of right conjunct lengths, and in each case the mean of left conjunct lengths is smaller than the mean of right conjunct lengths. All 12 differences between means in Table 1 are highly significant ($p \ll 0.001$), as established by the one-sided Wilcoxon test (with the values of V statistics and p reported in the table).[6] This confirms and extends partial results of previous works, which focused on particular constructions and used particular length metrics: there is a general tendency in English for left conjuncts to be shorter.

[6]The non-parametric Wilcoxon test was used here and elsewhere, because the relevant distributions are not normal, as ascertained by the Anderson-Darling test for normality (with $p \ll 0.001$). (The Anderson-Darling test is reported in Razali and Yap 2011 and Yap and Sim 2011 to be of comparable performance to the more frequently used Shapiro-Wilk test, which however could not be used here, as some sample sizes are greater than 5,000.) The Wilcoxon test was applied here in one-sided mode, with the alternative hypothesis being that the length of the left conjunct is shorter than that of the right conjunct. All statistics reported in this paper were calculated using R (R Core Team 2022).

	governor on the left prop	governor on the left N	governor on the right prop	governor on the right N	$\chi^2(1)$	p
characters	0.632	12140	0.603	4236	11.5	0.0007
syllables	0.599	11027	0.600	3671	0.0	0.87
words	0.674	8377	0.625	1754	15.1	0.0001

Table 2: Proportions of shorter conjuncts occurring on the left (vs. right) depending on the position of the governor, in coordinations with conjuncts of different lengths

4 Dependence on Governor Position

The previous section showed that left conjuncts tend to be shorter, even when the governor is on the right. However, this section will demonstrate that the position of the governor matters and that the governor does attract shorter conjuncts to some extent. We first report on an unsuccessful attempt to make this demonstration.

If the governor attracts shorter conjuncts, then we might expect more left conjuncts to be shorter when the governor is on the left than when the governor is on the right. Table 2 shows that this expectation is *partially met*: when length is measured in characters or words, the proportion of shorter left conjuncts is indeed greater when the governor is on the left; these two differences are highly significant ($p < 0.001$), as ascertained by the two-sided proportions test.[7] However, when length is measured in syllables, the proportion of shorter left conjuncts is slightly higher in the opposite scenario, i.e., when the governor is on the right, but this difference is not statistically significant.

Moreover, by the same reasoning, when there is no governor, we might expect relevant proportions to be somewhere between those in Table 2. As shown in Table 3, this expectation is *not met*: when there is no governor, the proportions of shorter left conjuncts are smaller than when there is a governor on the left or on the right. The reason for this is that coordinations without a governor are very

[7]N is the number of coordinations with a given position of the governor in which conjuncts have different lengths according to a given measure.

	no governor		vs. on the left		vs. on the right	
	prop	N	$\chi^2(1)$	p	$\chi^2(1)$	p
characters	0.575	3932	41.5	1.2e-10	6.7	0.0098
syllables	0.569	3764	10.4	0.0012	7.6	0.0057
words	0.587	3574	82.4	1.1e-19	7.2	0.0072

Table 3: Proportions of shorter conjuncts occurring on the left (vs. right) in the absence of governor, in coordinations with conjuncts of different lengths

specific: 97% of them are coordinations of matrix Ss (sentences; 57%) or VPs (verbal phrases; 40%), and many of such coordinations are constituted by a main sentence or VP, with another – shorter – added as a comment to the first one (see (10)) or heavily elided (see (11)).[8]

(10) [[Mr. Straszheim expects he will take some heat], and [he's right]].

(11) The bank [[employs 8,000 people in Spain] and [2,000 abroad]].

In summary, comparing total proportions of shorter left conjuncts does not provide us with an argument for the influence of the governor.

However, such an influence is clear when we investigate how these proportions change with absolute length differences. Figure 1 contains the results of fitting monofactorial binary logistic regression models to the three subpopulations differing in the presence and position of the governor (see the three rows in this figure); this is done for all three length metrics (see the three columns).[9] The figure shows that when there is no governor (the first row of plots) and when it is on the left (the second row), the proportions of shorter left conjuncts grow steadily; in all six cases the probability p that this positive tendency is accidental is well below 0.001. Interestingly, the situation changes drastically when we consider coordinations with a governor on the right (the third row). Here, the correlation is not significantly positive, and in the case of words it is even (insignificantly) negative.

Additional multifactorial binary logistic regression analysis confirmed the very significant interaction of governor position and absolute length difference ($p < 0.001$ for characters and syllables, $p < 0.01$ for words). Moreover, the analysis of slope contrasts (performed with R's `emmeans::emtrends`) shows that the slope is statistically significantly flatter when the governor is on the right than when it is on the left or missing. Finally, in the case of lengths measured in characters and syllables, the slope is significantly steeper when the governor is on the left than when there is no governor.[10]

In summary, the tendency for shorter conjuncts to occur on the left grows with absolute length difference between conjuncts when there is no governor and – even more so – when it is on the left, but not when it is on the right.

5 Dependency Length Minimization

Dependency Length Minimization (DLM) is the tendency for natural languages to prefer structures with shorter dependencies to those with longer dependencies (for overviews see, e.g., Liu et al. 2017, Temperley and Gildea 2018). This tendency has long been noted in linguistics (Behaghel 1909, 1932: 4), has been confirmed by numerous corpus studies (some of the earliest being Hawkins 1994 and Ferrer-i-Cancho 2004) combined with computer simulations (starting with Gildea and Temperley 2007, 2010 and Liu 2008), and has received various psycholinguistic and statistical explanations (e.g., Hawkins 1994, Gibson 1998, Futrell and Levy 2017, Futrell 2019). As argued in Hawkins 1994 and Futrell et al. 2020, this tendency operates both at the level of grammar and at the level of use.

At the level of use, when both orders of two dependents are grammatical, the shorter dependent tends to occur closer to the governor – both in head-initial languages, where the short–long tendency is observed (e.g., Bever 1970, Hawkins 1994, Arnold et al. 2000), and in head-final languages, where the long–short tendency is seen (e.g., Hawkins 1994, Yamashita and Chang 2001, Yamashita 2002). For example, when two PP dependents of a verb are of similar lengths, both orders are perfectly fine (e.g., *sing [in the club] [for an hour]* and *sing [for an hour] [in the club]*), but as length differences between the two PPs increase, so does the tendency for the shorter to occur next to the verb (e.g., *sing*

[8] By contrast, such S or VP coordinations only constitute 24% of coordinate structures with the governor on the left, and 10% with the governor on the right.

[9] Due to the low number of coordinations with large length differences when the governor is on the right, observations were collected into five buckets defined by the vector $\vec{\delta} = \langle 0, 1, 2, 3, 6, 25 \rangle$ for words, where bucket i contains coordinations with absolute length differences within the interval $(\vec{\delta}_i, \vec{\delta}_{i+1}]$; for syllables and characters the relevant vectors are $2 * \vec{\delta}$ and $6 * \vec{\delta}$, as an average English word consists of approximately 2 syllables and 6 characters.

[10] In the case of words, the slight opposite tendency is observed, but it is not (even marginally) statistically significant.

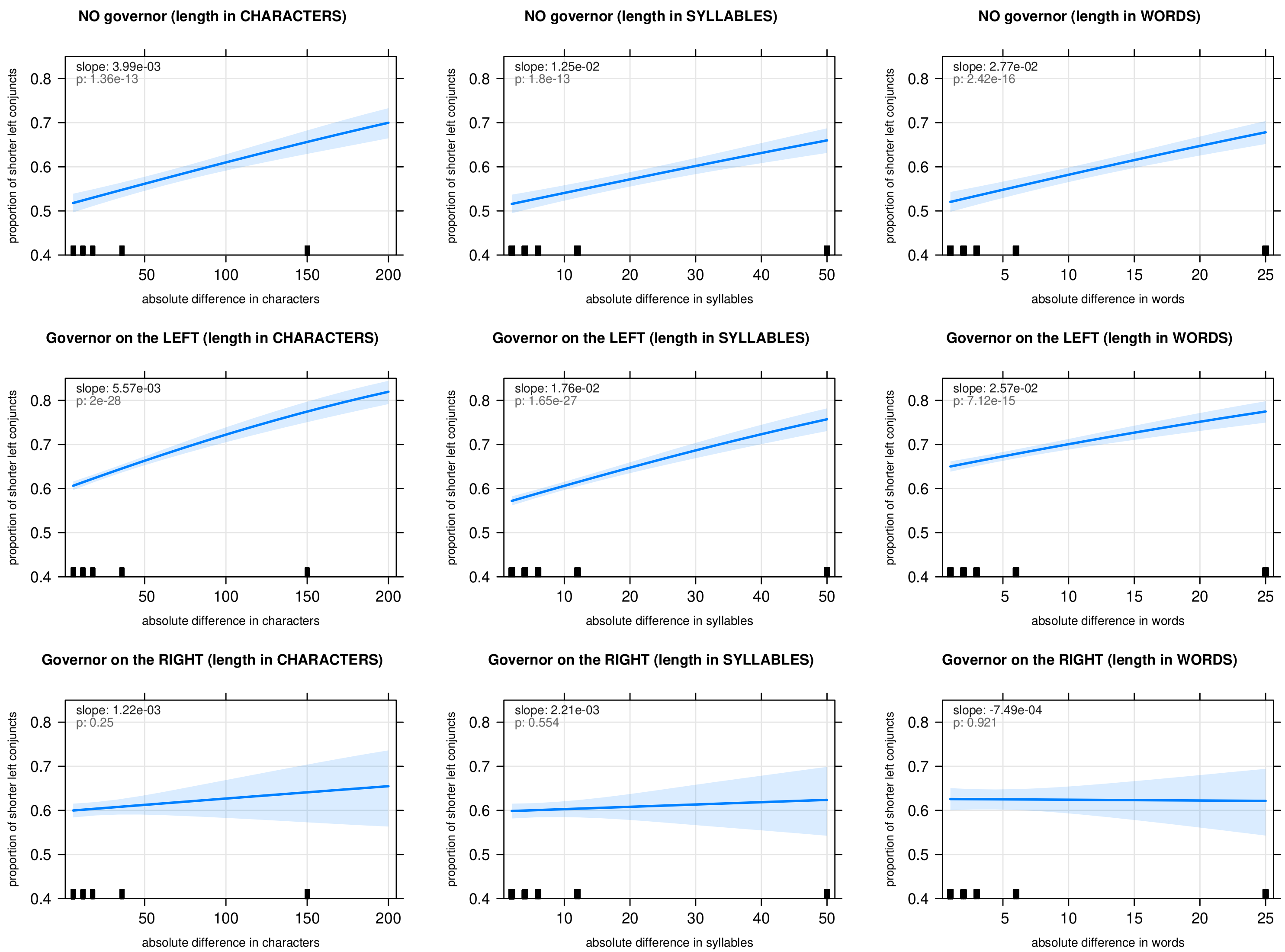


Figure 1: Proportions of shorter left conjuncts depending on the absolute difference of conjunct lengths (with confidence bands)

[for an hour] [in the most famous jazz club in the whole of USA] is more likely to occur than *sing [in the most famous jazz club in the whole of USA] [for an hour]*).

At the level of grammar, certain conventionalized word orders turn out to minimize dependency lengths on average. For example, when an NP (nominal phrase) 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.

We hypothesize that the same processes are at play in coordination. That is, the dependency structure of coordination must be such that shorter left conjuncts minimize dependency lengths – the more so, the bigger the length difference – when the governor is on the left (see the middle row of plots in Figure 1) or absent (see the top row), but not when it is on the right (see the bottom row). In the next section we will investigate which of the dependency approaches to coordination are compatible with such a DLM-based explanation of the effects illustrated in Figure 1.

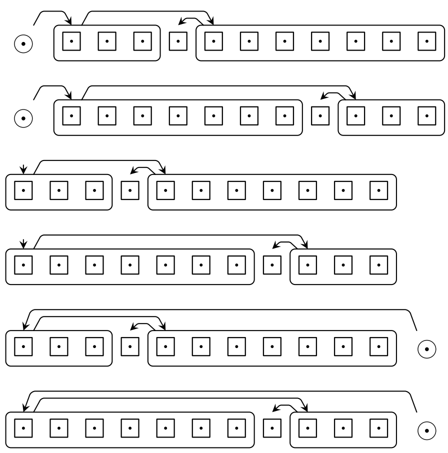
6 Dependency Structure of Coordination

The following reasoning is based on the observation that, in English, heads of both conjuncts are on average situated the same – usually short – distance from the left periphery. In the case of PPs, VPs, CPs (complementizer phrases, e.g., *that he came*; marked as SBAR in PTB) and NPs on their analysis as determiner phrases (Abney 1987, Hudson 1990), this will usually be the left-most word. In the case of NPs analysed as headed by the noun, this will be the second word on average (first, in the case of determinerless plurals and mass terms, third when both a determiner and an adjective is present, etc.), in the case of typical sentences the head will be offset by the subject, etc.

Let us first consider the Bouquet/Stanford approach. As can be seen in (12a–b), which illustrates coordination with the governor on the left, the total dependency length is smaller when the shorter conjunct occurs on the left (as in (12a)) than when it occurs on the right (as in (12b)). The same holds for coordinations with no governor, in which case the head of the first conjunct is the root; see (12c–d). This agrees with the tendencies illustrated in Figure 1. However, also when the governor is on the right, as in (12e–f), shorter left conjuncts minimize dependency length. Moreover, the gain is the same as in the other two situations: it is the length difference of the conjuncts. So, if the Stanford approach accurately reflected dependencies in coordination, we would expect the third row in Figure 1 to look the same as the first two rows, contrary to facts.

(12) **Bouquet/Stanford**:

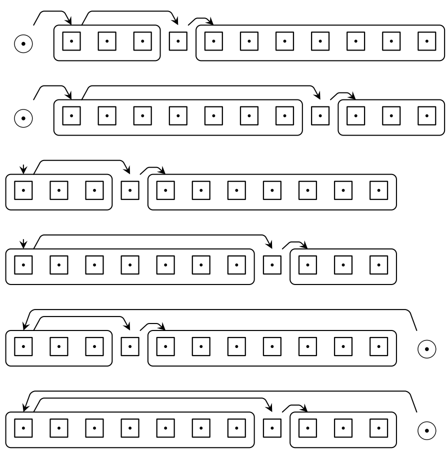
a.
b.
c.
d.
e.
f.



Exactly the same reasoning applies to the Chain/Moscow approach, illustrated in (13).[11]

(13) **Chain/Moscow**:

a.
b.
c.
d.
e.
f.

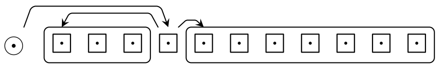


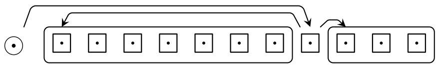
[11]As well as to its variant in Surface Syntactic Universal Dependencies (SUD; Gerdes et al. 2018, 2021; https://surfacesyntacticud.github.io/guidelines/u/particular_phenomena/coord/).

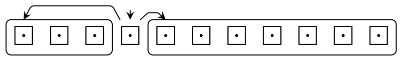
Hence, the Moscow approach also does not provide a good linguistic model of our empirical findings.

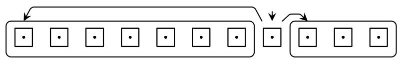
On the other hand, the Conjunction-headed/Prague approach is compatible with our corpus observations. In this case, when the governor is on the left (see (14a–b)), the dependency minimization gain is twice greater than when there is no governor (see (14c–d)): it is twice the length difference between conjuncts. This larger minimization gain may explain the above observation that the slopes in Figure 1 tend to be steeper when there is a governor on the left than when there is no governor.

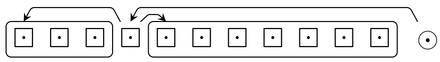
(14) **Conjunction-headed/Prague**:

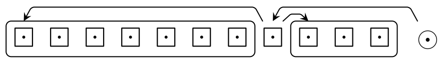
a. 

b. 

c. 

d. 

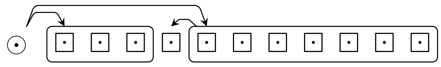
e. 

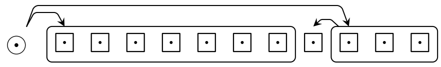
f. 

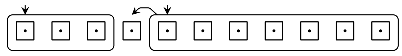
Interestingly, unlike in the case of the previous two approaches to coordination, relative conjunct lengths do not matter for dependency length minimization when the governor is on the right (see (14e–f)). Hence, the Prague approach makes it possible to explain the effects observed in Figure 1 and, moreover, the explanation is purely at the level of use: it does not invoke any conventionalized effects of DLM.

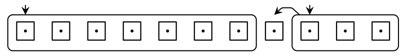
Let us finally consider the Multi-headed/London approach. At first, it seems incompatible with our empirical findings: when the governor is on the left (see (15a–b)), shorter left conjuncts minimize dependency length, as confirmed by empirical observations, but when there is no governor (see (15c–d)), lengths of conjuncts do not matter for dependency minimization, and the significantly positive slopes in plots in the top row of Figure 1 remain unexplained. Moreover, when the governor is on the right (see (15e–f)), shorter *right* conjuncts seem to minimize dependencies, so the lines in the third row of Figure 1 are expected to have significantly negative slopes, symmetrical to the positive slopes of lines in the middle row, contrary to facts.

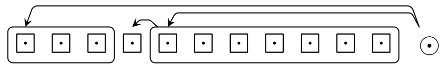
(15) **Multi-headed/London**:

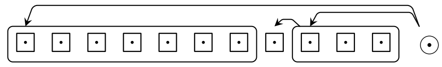
a. 

b. 

c. 

d. 

e. 

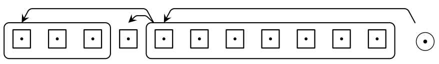
f. 

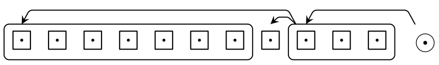
However, the London approach to coordination is compatible with Figure 1 on the assumption that the at-use pressure for shorter left conjuncts in the majority of governed coordinate structures is conventionalized as an at-grammar preference for shorter left conjuncts. That is, just as in the case of NP and PP dependents of verbs, DLM may be assumed to work in coordination both at the level of use and at the level of grammar. In the case of coordination with no governor, where there are no at-use preferences for relative conjunct lengths (see (15c–d)), this at-grammar preference still has the effect that shorter conjuncts are preferred on the left, as seen in the top row of Figure 1. Moreover, this at-grammar preference makes it possible to explain the steeper slopes when the governor is on the left (the middle row), namely, as an additive effect of the at-use pressure (see (15a–b)) and the at-grammar convention. Finally, in the case of governor on the right (see (15e–f)), the at-use pressure for shorter right conjuncts is at conflict with the at-grammar preference for shorter left conjuncts, and they seem to largely cancel each other out, as observed in the third row of Figure 1.

In partial summary, out of the four standard dependency approaches to coordination, only two make it possible to explain the curious tendencies in Figure 1 as DLM effects, with DLM operational either only at use (the Prague approach) or both at use and at grammar (the London approach). Paradoxically, while these tendencies are not symmetrical – shorter left conjuncts are preferred when the governor is on the left (or absent), but there is no clear preference either way when it is on the right – the two approaches that are compatible with it are symmetrical in the sense that they treat both conjuncts on par. On the Prague approach, coordination is headed by the conjunction and each conjunct is its dependent, while on the London approach coordination is multi-headed – equally by the head of each conjunct. This should be contrasted with Stanford and Moscow approaches, in which coordination is – asymmetrically – headed by the first conjunct.

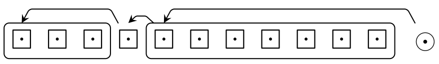
Interestingly, more symmetrical variants of these two asymmetric approaches to coordination are compatible with the empirical observations in Figure 1. To see that, consider the simple variants of the "first conjunct is the head" approaches in which coordination is still headed by the first conjunct when there is no governor, but otherwise it is headed by the conjunct closer to the governor. Such variants may be justified by the occasional observations that, in various languages, the governor seems to have a special relation to the closest conjunct, e.g., in the phenomenon of closest conjunct agreement (see Nevins and Weisser 2019 for an overview). These variants differ from the original approaches only when the governor is on the right, i.e., only in configurations in which the original approaches were incompatible with corpus findings:

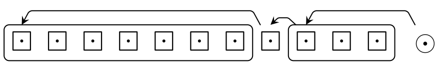
(16) **Bouquet/Stanford (closest)**:

e. 

f. 

(17) **Chain/Moscow (closest)**:

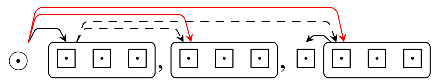
e. 

f. 

So-modified Stanford and Moscow approaches behave like the Prague approach: shorter left conjuncts still minimize dependency lengths when the governor is on the left or absent, but which conjunct is shorter does not matter when the governor is on the right. Hence, on the assumption that there is no conventionalized preference for shorter left conjuncts, such approaches are compatible with the lack of clear tendency in the third row of Figure 1.

Let us finally consider the enhanced version of the current treebank annotation standard, Universal Dependencies (UD; `https://universaldependencies.org/`; Nivre et al. 2016, de Marneffe et al. 2021, Zeman et al. 2022), appropriately called Enhanced Universal Dependencies (EUD; Schuster and Manning 2016). It combines Stanford and London approaches to coordination. The basic EUD structure of coordination is schematically represented in (18):

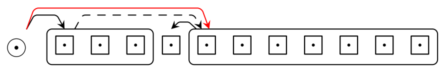
(18) **Enhanced UD**:

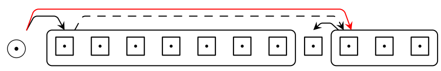


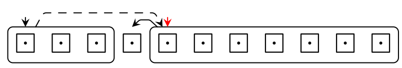
The vanilla UD standard follows the Stanford approach to coordination. Enhanced UD retains all dependencies of the vanilla UD and adds dependencies from the governor to all non-initial conjuncts.[12] In (18), there are two such EUD-specific dependencies from the governor to non-initial conjuncts (drawn in red above all other dependencies). Note that removing some of the basic UD dependencies, namely, those which are shown in (18) as dashed arcs, would make the structure purely multi-headed. But, as it stands, the full EUD structure contains all dependencies of both approaches: Stanford and London. It turns out that, because of that, such structures are compatible with the empirical findings of this paper.

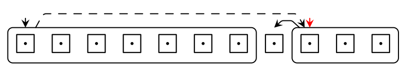
As can be seen in (19), shorter left conjuncts again minimize dependency lengths when the governor is on the left (see (19a–b)) or absent (see (19c–d)). However, when the governor is on the right, the total dependency length does not depend on the order of conjuncts (see (19e–f)). This is compatible with the empirical findings of this paper in a way fully analogous to how the Prague approach fits these empirical findings.

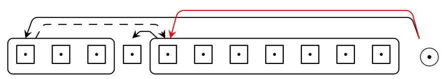
(19) **Enhanced UD**:

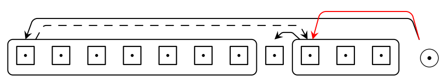
a. 

b. 

c. 

d. 

e. 

f. 

We conclude that, out of the theoretical linguistic approaches to the dependency structure of English coordinations, both Conjunction-headed/Prague

[12] EUD also adds dependencies from conjuncts to their shared dependents.

and Multi-headed/London approaches may explain the patterns observed in $PTB_{\&}$. In particular, the results of this paper provide an argument against the theoretical linguistic validity of the basic version of the Universal Dependencies standard, commonly used in NLP applications, given that basic UD implements the Bouquet/Stanford approach. However, they also provide support for the enhanced version of this standard, which – on top of dependencies present in basic UD – adds dependencies from the governor to each conjunct, thus making the structure effectively multi-headed.

7 Previous Work

This work carries out the research program briefly suggested in Temperley and Gildea 2018: 78: "[G]iven the strong evidence for DLM, a finding that one syntactic analysis of (say) coordinate phrases resulted in shorter dependencies than another analysis [. . .] could be regarded as a strong point in its favor." The only earlier attempt to do that that we are aware of is Temperley 2005. It identifies 6 constructions with the governor on the left and 3 with the governor on the right and attempts to extract them from the original PTB, reporting that "in many cases [relevant dependencies] can be inferred quite reliably from the constituent structures". Temperley shows that – in these constructions – left conjuncts tend to be shorter in terms of words, even when the governor is on the right. Temperley considers two dependency approaches to coordination, Multi-headed/London and Chain/Moscow, and argues that both are compatible with this tendency: the Moscow approach at the level of use, and the London approach at the level of grammar. By contrast, we show that DLM must be at work in English coordination at both levels in order to be compatible with the Multi-headed/London approach and, crucially, that the observed facts cannot be explained via DLM by asymmetrical approaches such as Chain/Moscow.[13]

While Temperley (2005) does not consider the dependence of the proportion of shorter left conjuncts on the length difference between conjuncts, this effect is demonstrated by Gibson et al. (1996: 88–90). They show that, in NP coordinations in PTB and in the Brown corpus (Kučera and Francis 1967), proportions of shorter left conjuncts grow proportionally to length differences between conjuncts. While they do not consider the dependence of this effect on the presence and position of the governor, so they do not notice that this tendency disappears when the governor is on the right, their results are broadly compatible with ours.[14]

Apart from these two highly relevant previous publications, there is little corpus work on the order of conjuncts in coordination, the most important being Lohmann 2014. While it is limited to English nominal coordinations, it confirms the short–long tendency (measured in syllables there) observed elsewhere, but also shows that semantic factors have a stronger effect (where applicable), which explains why – even when differences in length are very large – the proportion of shorter left conjuncts is still much lower than 1.

8 Conclusion

This paper makes the following empirical contributions: 1) it demonstrates – more robustly than previous work – the general short–long tendency in binary coordinations in English, and 2) that this tendency also holds for coordinations followed by their governor. The novel observation is 3) how this effect depends on differences in conjunct lengths and on the position of the governor: the strong statistically highly significant positive correlation between length differences and proportions of shorter left conjuncts disappears when the governor is on the right. On the theoretical side, 4) we argued that this effect is explained by DLM, possibly operating at both levels: use and grammar, and 5) that this explanation is only possible when the symmetrical dependency structures of coordination are assumed. To the extent to which no other explanations of the effects observed in Figure 1 are forthcoming, this provides an argument for the linguistic validity of symmetrical approaches to coordination such as Conjunction-headed/Prague and Multi-headed/London and against those approaches which assume that coordinations are headed by the first conjunct, such as Bouquet/Stanford and Chain/Moscow approaches.

[13] Temperley (2005) also argues for the superiority of the multi-headed approach over the chain approach, but his crucial argument for that relies on certain – not universally shared – assumptions about the structure of nominal phrases and about non-projectivity.

[14] For lack of space, we do not provide plots for the whole population of coordinations in $PTB_{\&}$, but such plots would be similar to the first two rows in Figure 1, as coordinations with the governor on the right ($N = 4{,}719$), which display a weakish and statistically insignificant correlation between proportions and length differences, would be dominated by the other coordinations ($N = 17{,}106$), where the correlation is much stronger and statistically significant.

Acknowledgements

This paper benefited from comments by Joakim Nivre, Agnieszka Patejuk, David Temperley, the three anonymous ACL 2023 reviewers, as well as members of the Linguistic Engineering Group at ICS Polish Academy of Sciences (especially, Małgorzata Marciniak) and students of the Cognitive Science Program at the University of Warsaw (especially, Grzegorz Kasperski and Wojciech Stempniak).

9 Limitations

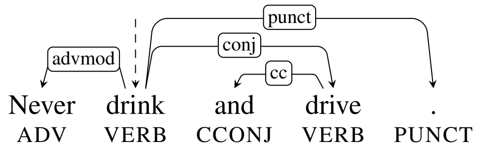
9.1 English and PTB

The main limitation of the research reported here is that it is based not only on just a single language, English, but on just a single corpus containing single-genre texts (from the Wall Street Journal), $\mathrm{PTB}_{\&}$. It might seem that this kind of research should instead be performed on Universal Dependencies (UD; `https://universaldependencies.org/`; Nivre et al. 2016, de Marneffe et al. 2021, Zeman et al. 2022) – a collection of over 200 dependency treebanks for over 100 languages, currently the favourite resource for investigating DLM (see, e.g., Futrell et al. 2020).

Unfortunately, UD is ill-suited for the task at hand. In order to investigate conjunct lengths in relation to the position of the governor, a resource is needed which, for each coordination, makes it possible to locate its governor, if any, and to identify the exact extent of each conjunct (in order to measure its length). UD excels on the first requirement but fails on the second: it is not clear whether dependents on the left of the head of the left conjunct are a part of this conjunct or whether they modify the whole coordinate structure (or are shared by all conjuncts). This problem is illustrated with the following example from the UD English GUM corpus (Zeldes 2017), where it is not clear whether *never* is part of the left conjunct only (in which case the left conjunct is longer than the right conjunct) or whether it is shared by the two conjuncts (in which case they are of equal length).[15]

[15]Note that, technically, UD follows the asymmetrical Bouquet/Stanford approach to coordination, even though, conceptually, coordination is assumed to be symmetric: "UD in principle assumes a symmetric relation between conjuncts, which have equal status as syntactic heads of the coordinate structure. However, because the dependency tree format does not allow this analysis to be encoded directly, the first conjunct in the linear order is by convention always treated as the parent of all other conjuncts" (de Marneffe et al. 2021: 276).

(20)



Our initial experiments suggest that it is much easier to find governors of coordinations in $\mathrm{PTB}_{\&}$, than to decide which dependents should be shared by conjuncts in UD treebanks;[16] hence the use of $\mathrm{PTB}_{\&}$ rather than UD in the current paper.

The Enhanced UD (EUD) format makes it possible to explicitly encode which dependents are shared, but currently almost all EUD treebanks seem to be the result of error-prone automatic conversion from basic UD treebanks based on simple heuristics (Schuster and Manning 2016), so such information is currently not fully reliable. For example, the EUD version of GUM (in the current 2.11 release of UD) does not contain information that *never* in the above example is shared between conjuncts (the intended interpretation of this sentence), although it does contain information about, say, shared subjects. However, given the recent steps towards the creation of EUD treebanks with more reliable coordination information (Grünewald et al. 2021), future versions of EUD treebanks might become better-suited for the task at hand.

Moreover, some EUD treebanks – for some languages other than English – are the result of direct translation from formats that do distinguish between modifiers of left conjuncts and modifiers of whole coordinations (or all conjuncts), so it is possible that they preserve these distinctions. Probably the largest such treebank in the current release 2.11 of UD is UD_Czech-PDT (87,913 sentences; Hajič et al. 2006), other such treebanks including: UD_Czech-CAC (24,709; Hladká et al. 2008), UD_Polish-PDB (22,152; Wróblewska 2018), UD_Polish-LFG (17,246; Przepiórkowski and Patejuk 2020), UD_Dutch-Alpino (13,603; Bouma and van Noord 2017), UD_Czech-FicTree (12,760; Jelínek 2017), UD_Slovak-SNK (10,604; Zeman 2017), UD_Arabic-PADT (7,664; Hajič et al. 2009), UD_Dutch-LassySmall (7,341; Bouma and van Noord 2017), and UD_Lithuanian-ALKSNIS (3,642; Bielinskienė et al. 2016). Furthermore, there are some treebanks – including reasonable-sized French treebanks – natively en-

[16]See also Findlay and Haug 2021: §3 on the relatively low performance of heuristics aiming at deciding whether a given dependent is private to one conjunct or shared by all conjuncts in UD treebanks.

coded in the surface-syntactic alternative to UD, Surface Syntactic Universal Dependencies (SUD; Gerdes et al. 2018, 2021), which explicitly represents information about shared dependents. Moreover, there are some constituency treebanks (thus unambiguously representing extents of conjuncts), which also contain explicit information about heads of constructions (which makes identifying governors easy), that could be used for the purposes of the current research; these include the Polish Składnica treebank (Woliński et al. 2011, Woliński and Hajnicz 2021) and the Swedish Eukalyptus treebank (Adesam et al. 2015). Finally, all the relevant information is represented in Head-driven Phrase Structure Grammar (HPSG; Pollard and Sag 1994, Müller et al. 2021) and Lexical Functional Grammar (LFG; Kaplan and Bresnan 1982, Dalrymple et al. 2019) parsebanks – e.g., the English Redwood treebank (Oepen et al. 2004), the Bulgarian BulTreeBank (Simov et al. 2002), or the Polish LFG Treebank (Patejuk and Przepiórkowski 2015) – but the internal format of structures in such treebanks is usually much more complex than in the case of PTB or (E)UD, and less well-documented, which makes it more difficult to extract relevant information automatically. Nevertheless, future work should investigate to what extent these treebanks contain reliable and unambiguous information about coordinate structures and their governors and, if they do, it should attempt to replicate the results reported here on the basis of those treebanks.

9.2 Confounding Factors

Another limitation of the research reported above is that it does not consider possible confounding factors. There are at least two such factors that should be carefully investigated, although in both cases we give reasons below why we believe these factors should not dramatically influence our conclusion regarding the ability of symmetrical dependency approaches to coordination to explain the patterns observed in Figure 1.

The first possible confounding factor is the "old/new" discourse status of conjuncts. It is well known that phrases expressing discourse-old ("given") information tend to be shorter than discourse-new phrases and they tend to occur earlier in the sentence (see, e.g., Arnold et al. 2000). Applied to conjuncts, this might explain both the fact that left conjuncts tend to be shorter regardless of the presence and position of the governor (see Table 1 in the main text) and that the proportion of shorter left conjuncts is greater than 0.5 regardless of the presence and position of the governor (see Tables 2–3). The explanation would be simple: when the two conjuncts have different discourse status, the statistically shorter discourse-old conjunct tends to occur earlier in the sentence, i.e., as the left conjunct, which influences the hypothetically otherwise balanced distribution of conjunct lengths in coordinations. However, Temperley 2005: 587–588 shows that even when the discourse status of nominal conjuncts is controlled for (e.g., by considering only indefinite NP conjuncts, which tend to be discourse-new), left conjuncts tend to be shorter, and the differences between means of length ratios are statistically significant.

More importantly, it is not clear how this discourse factor could alone explain the tendencies seen in Figure 1. In order to explain the first two rows, it would have to be assumed that the larger the length difference between conjuncts, the greater the probability that the shorter conjunct is discourse-old and the longer is discourse-new and, hence, the larger the proportion of shorter left conjuncts. We are not aware of a claim to this effect being made in the literature. But even if it were true, this hypothesis alone would not suffice, as it is directly contradicted by the third row, where proportions of shorter left conjuncts do not grow with their length differences in a statistically significant way, and even seem to shrink in the case of length differences measured in words.

On the other hand, it is possible to combine this discourse-based hypothesis with the at-use DLM considerations above, i.e., to adopt this hypothesis as an alternative to the grammar-level convention "prefer shorter left conjuncts proportionally to length difference". (Recall that this convention had to be assumed in order to make the empirical results of this paper compatible with the Multi-headed/London approach to coordination.) Then the at-use operation of DLM and this general discourse preference for shorter left conjuncts would conspire in the explanation based on the London approach, but only on the assumption – crucial for that explanation – that the strength of the discourse preference grows with the absolute difference between lengths of conjunct. The opposite assumption, namely, that the strength of this discourse preference does not depend on such length differences, is needed in the case of the Conjunction-

	governor					
	none		on the left		on the right	
	#	%	#	%	#	%
ADJP	7	0.18	355	2.71	714	15.13
ADVP	4	0.10	123	0.94	38	0.81
NP	79	1.98	7886	60.17	2966	62.85
NX	3	0.07	326	2.49	61	1.29
PP	3	0.07	446	3.40	24	0.51
QP	5	0.12	91	0.69	59	1.25
S	2267	56.67	504	3.85	337	7.14
SBAR	3	0.07	373	2.85	20	0.42
UCP	4	0.10	278	2.12	340	7.20
VP	1609	40.23	2693	20.55	153	3.24

Table 4: Numbers (#) and percentages (%) of coordinations of different categories depending on the presence and position of the governor

headed/Prague approach and the three variants of the "coordination headed by the first conjunct" approaches considered at the end of §6. This is because, on these approaches, at-use DLM does not say anything about the relative lengths of conjuncts when the governor is on the right, so – if the discourse preference for shorter left conjuncts is really at work here – its effect must be relatively constant, as witnessed in the third row of plots in Figure 1: the slopes there are not significantly positive.

In summary, the potential confounding factor of discourse-newness may influence which of the symmetrical approaches to coordination explains the data better: London (if the strength of discourse effects depends on conjunct length differences) or Prague and variants (if it does not). On the other hand, this potential confounding factor should not influence the general conclusion that symmetrical approaches have more explanatory power than asymmetrical approaches such as Bouquet/Stanford or Chain/Moscow.

The second possible confounding factor is the type of coordination: whether it is a coordination of NPs, VPs, sentences, etc. Table 4 gives numbers and percentages of different kinds of coordination, depending on the position of the governor, and Figure 2 presents a mosaic plot visualizing this data.[17] It shows that the distributions of categories of coordination vary considerably depending on the presence and position of the governor. For example, as mentioned in §4, 97% of all coordinations without a governor are coordinations of sentences (57%)

[17]All categories occurring at least 20 times with the governor on the right and at least 20 times with the governor on the left where taken into account. All other categories occur less than ten times in *each* of the *three* data sets.

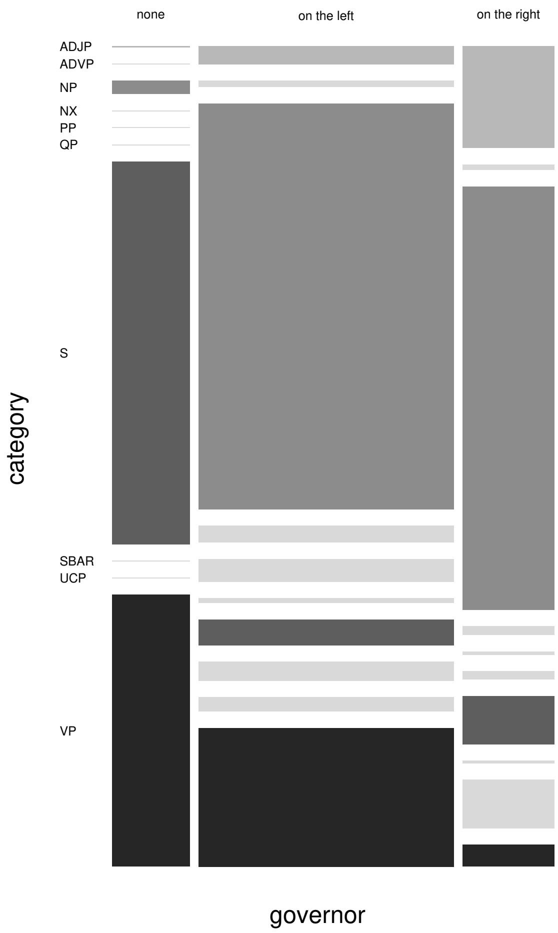


Figure 2: Proportions of coordinations of different categories depending on the presence and position of the governor; ADJP, NP, S, and VP coordinations are shown in (increasingly) darker shades of grey

and VPs (40%), while this is true of only 24% of coordinations with the governor on the left, and only 10% – on the right. As sentences and VPs are typically much longer than phrases bearing other categories, also conjuncts in coordinations of these types are much longer, and their mean length differences are also larger (see Table 1 in the main text). Moreover, as also observed in the main text, many of such coordinations have shorter right conjunct (see examples (10)–(11)), which results in such coordinations having a lower proportion of shorter left conjuncts than coordinations with a governor (contrast Tables 2 vs. 3). So it is clear that the different distributions of categories of coordination, which depend on the presence and position of the governor, have an impact on the statistics reported in Tables 1–3 and, thus, should be carefully investigated.

It is less clear how this potentially confounding factor could influence the main effect observed in Figure 1, i.e., that there is a statistically significant positive correlation between proportions of shorter left conjuncts and absolute length differences when there is no governor or when it is on the left, but no such correlation when the governor is on the right. One way to verify that there is no such influence would be to examine plots analogous to those in Figure 1, but separately for each category of coordination. Unfortunately, there is no category that would be sufficiently well represented in all three populations; for example, while there are as many as 7,886 NP coordinations with the governor on the left and 2,966 when it is on the right, there are only 79 such coordinations without a governor in $\text{PTB}_{\&}$. Not only is this last number much too low for a statistically valid investigation, but also the 2,966 coordinations with the governor on the right do not contain enough data (enough in the sense explained in Harrell 2015: 72–73; cf. Gries 2021: 71) for the case where the right conjunct is shorter: there are only 24 such coordinations with length difference (in words) 3, only 14 when the difference is 4, only 7 when it is 5, etc.

Nevertheless, once larger corpora with good quality annotation of coordinations become available, the effect of coordination category should be investigated in more detail.

References

Steven Abney. 1987. *The English Noun Phrase in its Sentential Aspect*. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Yvonne Adesam, Gerlof Bouma, and Richard Johansson. 2015. Defining the Eukalyptus forest — the Koala treebank of Swedish. In *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015)*, pages 1–9, Vilnius, Lithuania. Linköping University Electronic Press, Sweden.

Jennifer E. Arnold, Thomas Wasow, Anthony Losongco, and Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76(1):28–55.

Otto Behaghel. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen*, 25:110–142.

Otto Behaghel. 1932. *Deutsche Syntax: eine geschichtliche Darstellung. Band IV: Wortstellung. Periodenbau*. Carl Winters Universitätsbuchhandlung, Heidelberg.

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

Thomas G. Bever. 1970. The cognitive basis for linguistic structures. In John R. Hayes, editor, *Cognition and the Development of Language*, pages 279–362. Wiley, New York.

Agnė Bielinskienė, Loïc Boizou, Jolanta Kovalevskaitė, and Erika Rimkutė. 2016. Lithuanian dependency treebank ALKSNIS. In Inguna Skadiņa and Roberts Rozis, editors, *Human Language Technologies – The Baltic Perspective*, pages 107–114. IOS Press.

Gosse Bouma and Gertjan van Noord. 2017. Increasing return on annotation investment: The automatic construction of a Universal Dependency treebank for Dutch. In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*, pages 19–26, Gothenburg, Sweden. Association for Computational Linguistics.

Mary Dalrymple, John J. Lowe, and Louise Mycock. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford University Press, Oxford.

Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation, LREC 2006*, pages 449–454, Genoa. ELRA.

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

Ramon Ferrer-i-Cancho. 2004. Euclidean distance between syntactically linked words. *Physical Review E*, 70:056135.

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.

Jamie Y. Findlay and Dag T. T. Haug. 2021. How useful are Enhanced Universal Dependencies for semantic interpretation? In *Proceedings of the Sixth International Conference on Dependency Linguistics (DepLing, Syntax Fest 2021)*, pages 22–34, Sofia, Bulgaria.

Richard Futrell. 2019. Information-theoretic locality properties of natural language. In *Proceedings of the First Workshop on Quantitative Syntax (Quasy, SyntaxFest 2019)*, pages 2–15, Paris, France. Association for Computational Linguistics.

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.

Kim Gerdes, Bruno Guillaume, Sylvain Kahane, and Guy Perrier. 2018. SUD or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 66–74. Association for Computational Linguistics.

Kim Gerdes, Bruno Guillaume, Sylvain Kahane, and Guy Perrier. 2021. Starting a new treebank? Go SUD! In *Proceedings of the Sixth International Conference on Dependency Linguistics (DepLing, Syntax Fest 2021)*, pages 35–46, Sofia, Bulgaria.

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

Edward Gibson, Carson T. Schütze, and Ariel Salomon. 1996. The relationship between the frequency and the processing complexity of linguistic structure. *Journal of Psycholinguistic Research*, 25(1):59–92.

Daniel Gildea and David Temperley. 2007. Optimizing grammars for minimum dependency length. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pages 184–191, Prague.

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

Stefan Th. Gries. 2021. *Statistics for Linguistics with R*, 3rd edition. De Gruyter Mouton, Berlin/Boston.

Stefan Grünewald, Prisca Piccirilli, and Annemarie Friedrich. 2021. Coordinate constructions in English enhanced Universal Dependencies: Analysis and computational modeling. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 795–809. Association for Computational Linguistics.

Jan Hajič, Otakar Smrž, Petr Zemánek, Petr Pajas, Jan Šnaidauf, Emanuel Beška, Jakub Krácmar, and Kamila Hassanová. 2009. Prague Arabic dependency treebank 1.0.

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).

Frank E. Harrell, Jr. 2015. *Regression Modeling Strategies*, 2nd edition. Springer.

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

Barbora Hladká, Jan Hajič, Jirka Hana, Jaroslava Hlaváčová, Jiří Mírovskỳ, and Jan Raab. 2008. The Czech academic corpus 2.0 guide. *The Prague Bulletin of Mathematical Linguistics*, 89(1):41–96.

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

Richard Hudson. 1990. *English Word Grammar*. Blackwell, Oxford.

Richard Hudson. 2010. *An Introduction to Word Grammar*. Cambridge University Press, Cambridge.

Tomás Jelínek. 2017. FicTree: A manually annotated treebank of Czech fiction. In *ITAT*, pages 181–185.

Ronald M. Kaplan and Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*, pages 173–281. MIT Press, Cambridge, MA.

Henry Kučera and W. Nelson Francis. 1967. *Computational Analysis of Present-day American English*. Brown University Press, Providence, RI.

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

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

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.

Igor Mel'čuk. 1974. *Opyt teorii lingvističeskix modelej «Smysl ⇔ Tekst»*. Nauka, Moscow.

Igor Mel'čuk. 1988. *Dependency Syntax: Theory and Practice*. The SUNY Press, Albany, NY.

Igor Mel'čuk. 2009. Dependency in natural language. In Alain Polguère and Igor Mel'čuk, editors, *Dependency in Linguistic Description*, pages 1–110. John Benjamins, Amsterdam.

Stefan Müller, Anne Abeillé, Robert D. Borsley, and Jean-Pierre Koenig, editors. 2021. *Head-Driven Phrase Structure Grammar: The Handbook*. Language Science Press, Berlin.

Andrew Nevins and Philipp Weisser. 2019. Closest conjunct agreement. *Annual Review of Linguistics*, 5:219–241.

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).

Stephan Oepen, Dan Flickinger, Kristina Toutanova, and Christoper D. Manning. 2004. LinGO Redwoods: A rich and dynamic treebank for HPSG. *Research on Language and Computation*, 4(2):575–596.

Agnieszka Patejuk and Adam Przepiórkowski. 2015. Parallel development of linguistic resources: Towards a structure bank of Polish. *Prace Filologiczne*, LXV:255–270.

Carl Pollard and Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. Chicago University Press / CSLI Publications, Chicago, IL.

Martin Popel, David Mareček, Jan Štěpánek, Daniel Zeman, and Zděněk Žabokrtský. 2013. Coordination structures in dependency treebanks. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 517–527, Sofia, Bulgaria.

Adam Przepiórkowski and Agnieszka Patejuk. 2020. From Lexical Functional Grammar to enhanced Universal Dependencies: The UD-LFG treebank of Polish. *Language Resources and Evaluation*, 54:185–221.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Nornadiah Mohd Razali and Bee Wah Yap. 2011. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1):21–33.

Sebastian Schuster and Christopher D. Manning. 2016. Enhanced English Universal Dependencies: An improved representation for natural language understanding tasks. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation, LREC 2016*, pages 2371–2378, Portorož, Slovenia. European Language Resources Association (ELRA).

Kiril Simov, Gergana Popova, and Petya Osenova. 2002. HPSG-based syntactic treebank of Bulgarian (BulTreeBank). In Andrew Wilson, Paul Rayson, and Tony McEnery, editors, *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*, pages 135–142. Lincom-Europa, Munich.

David Temperley. 2005. The dependency structure of coordinate phrases: A corpus approach. *Journal of Psycholinguistic Research*, 34(6):577–601.

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

David Vadas and James Curran. 2007. Adding noun phrase structure to the Penn Treebank. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pages 240–247, Prague.

Marcin Woliński, Katarzyna Głowińska, and Marek Świdziński. 2011. A preliminary version of Składnica—a treebank of Polish. In *Proceedings of the 5th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, pages 299–303, Poznań, Poland.

Marcin Woliński and Elżbieta Hajnicz. 2021. Składnica: a constituency treebank of Polish harmonised with the Walenty valency dictionary. *Language Resources and Evaluation*, 55:209–239.

Alina Wróblewska. 2018. Extended and enhanced Polish dependency bank in Universal Dependencies format. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 173–182. Association for Computational Linguistics.

Hiroko Yamashita. 2002. Scrambled sentences in Japanese: Linguistic properties and motivations for production. *Text*, 22:597–634.

Hiroko Yamashita and Franklin Chang. 2001. "Long before short" preference in the production of a head-final language. *Cognition*, 81:B45–B55.

Bee Wah Yap and Chiaw Hock Sim. 2011. Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12):2141–2155.

Amir Zeldes. 2017. The GUM corpus: Creating multilayer resources in the classroom. *Language Resources and Evaluation*, 51(3):581–612.

Daniel Zeman. 2017. Slovak dependency treebank in Universal Dependencies. *Journal of Linguistics/Jazykovedný casopis*, 68(2):385–395.

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, Hórunn 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 Sil-

veira, 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, Steinh͡ór 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.

Appendix: Counting Syllables

The syllabic length of a constituent was measured by adding syllabic lengths of all leaves of this constituent, whose labels do not start and end with a hyphen. Such technical labels include -LRB- and -RRB- (left and right round brackets.)

Syllables of each string were counted using the CMUdict Python package,[18] which relies on data from the Carnegie Mellon Pronouncing Dictionary.[19] This dictionary contains phonetic information; in particular, stressed phonemes, which can be treated as syllabic, are marked with numbers (e.g., *Peter* is represented as `P-IY1-T-ER0`). Therefore, using CMUdict, the number of syllables can be determined by counting the phonemes that end with a number.

Some PTB tokens were absent in CMUdict, in which case the following heuristics were applied.

In the case of tokens containing non-alphanumeric characters, these characters were first removed (apart from commas and dots in numbers), resulting in multiple tokens (see the example at the end of this appendix).

Tokens consisting of up to four upper-case English letters were treated as abbreviations, so their syllabic length was calculated as the sum of syllables in each letter, according to CMUdict. For example, the length of *WWW* is six syllables. In other cases of tokens consisting of letters, a simple heuristic was applied to try to estimate the number of syllables: each substring of vowels was counted as one syllable. If a word ended with *-e* (other than *-le*), then this ending was not counted as a syllable. According to this heuristic, the syllabic length of *beautiful* is 3, of *tackle* – 2, and of *fare* – 1.

Numbers were recognized as 1) strings of digits with optional commas and at most one dot after all commas, e.g., *100,000.99*, 2) possibly ending in *-st*, *-nd*, *-rd*, *-th* (ordinal numbers), 3) as well as strings of digits ending with *-s* and optionally starting with an apostrophe (e.g., *'80s*; such affixes are removed, as not affecting the number of syllables).[20] Every such number was converted to words using Python's Inflect package (e.g., *100,000.99* would be converted to *one hundred thousand point nine nine*).[21] As an exception, numbers in the range 1960–1999 were treated as years (e.g., 1984 as *nineteen eighty-four* rather than *one thousand nine hundred eighty-four*).

After the conversion, syllabic length was determined as in the case of other words, i.e., using CMUdict or the above heuristics.

Finally, a small dictionary with syllabic lengths was created for 18 strings: *$*, *%*, *&*, *½*, *¼*, *¾*, *'s* and 11 abbreviations of month names (*sans May*).

For example, consider the hypothetical token *O.K.-177,000\KTF+NATO-iron*. It would first be split on non-alphanumeric characters into the following tokens: *O*, *K*, *177,000*, *KTF*, *NATO*, and *iron*. *177,000* is recognized as a number and converted by Inflect to *one hundred and seventy seven thousand*, which corresponds to 11 syllables. *O*, *K*, *NATO*, and *iron* are all in CMUdict, with – respectively – 1, 1, 2, and 2 syllables, i.e., 6 in total. *KTF* is not in CMUdict, so it is split into letters, whose syllabic lengths are 1 according to CMUdict, so 3 in total. Hence, altogether, the whole initial token is estimated to consist of 20 syllables.

[18] `https://pypi.org/project/cmudict/`

[19] `http://www.speech.cs.cmu.edu/cgi-bin/cmudict`

[20] An attempt to recognize unknown tokens as numbers in 2) (e.g., *80th*) and 3) (e.g., *'80s*) was made *before* the initial stage of splitting tokens on non-alphanumeric characters.

[21] `https://pypi.org/project/inflect/`

ACL 2023 Responsible NLP Checklist

A For every submission:

☑ A1. Did you describe the limitations of your work?
9

☐ A2. Did you discuss any potential risks of your work?
Not applicable. We cannot think of any potential risks.

☑ A3. Do the abstract and introduction summarize the paper's main claims?
1

☒ A4. Have you used AI writing assistants when working on this paper?
Left blank.

B ☑ Did you use or create scientific artifacts?

We used – not created – scientific artefacts, as described in Section 2.

☑ B1. Did you cite the creators of artifacts you used?
2

☒ B2. Did you discuss the license or terms for use and / or distribution of any artifacts?
We used the extension of Penn Treebank described in Ficler and Goldberg 2016 (cited in the paper). Unfortunately, the license status of this extension is not clear. The webpage of this extension (https://github.com/Jess1ca/CoordinationExtPTB) requires one to first obtain another PTB extension from a URL that does not seem to be functional (http://schwa.org/projects/resources/wiki/NounPhrases). In the end we obtained the extension described in Ficler and Goldberg 2016 directly from the authors and we made sure that our institution obtained the original PTB from LDC, so that no rights were violated. We felt there was no need (and no space) to include the above description of the unclear license in the paper, but we will be willing to do so if reviewers make this request.

☑ B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
2 and 9.1

☒ B4. Did you discuss the steps taken to check whether the data that was collected / used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect / anonymize it?
We used a version of Penn Treebank, which 1) only consists of newspaper texts, 2) is widely available and has been used in countless tasks before, and – for these reasons – does not require any anonymization.

☐ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
Not applicable. Not applicable, because we do not make any new artifacts available.

☑ B6. Did you report relevant statistics like the number of examples, details of train / test / dev splits, etc. for the data that you used / created? Even for commonly-used benchmark datasets, include the number of examples in train / validation / test splits, as these provide necessary context for a reader to understand experimental results. For example, small differences in accuracy on large test sets may be significant, while on small test sets they may not be.
See Tables 1–3, fn.8, and Figure 1.

The Responsible NLP Checklist used at ACL 2023 is adopted from NAACL 2022, with the addition of a question on AI writing assistance.

C ☒ Did you run computational experiments?

Left blank.

☐ C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
No response.

☐ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
No response.

☐ C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?
No response.

☐ C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation), did you report the implementation, model, and parameter settings used (e.g., NLTK, Spacy, ROUGE, etc.)?
No response.

D ☒ Did you use human annotators (e.g., crowdworkers) or research with human participants?

Left blank.

☐ D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?
No response.

☐ D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?
No response.

☐ D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating? For example, if you collected data via crowdsourcing, did your instructions to crowdworkers explain how the data would be used?
No response.

☐ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
No response.

☐ D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
No response.