Dependency Structure of Coordination in Head-final Languages: a Dependency-Length-Minimization-Based Study

Wojciech Stempniak
University of Warsaw
and Saarland University
w.stempniak@student.uw.edu.pl

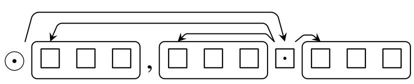
Abstract

There is no single accepted model of the dependency structure of coordination. Universal Dependencies (UD, De Marneffe et al. 2021) enforces in its corpora an asymmetrical model privileging the coordination's first conjunct as a standard. Kanayama et al. (2018) criticize that approach stating that this model is incompatible with the grammatical structure of head-final languages. Recent research (Przepiórkowski and Woźniak 2023, Przepiórkowski et al. 2024a) provides a DLM-based argument for the symmetrical models of the dependency structure of English coordination. This paper shows the result of the analysis of coordinations found in UD corpora of two head-final languages, namely Korean and Turkish. Based on the analysis of coordinations and theoretical arguments, an alternative approach to the dependency structure of coordination in head-final languages is suggested.

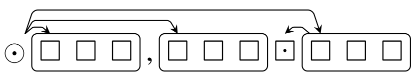
1 Introduction

There is no single universally accepted approach to the dependency structure of coordination. Przepiórkowski and Woźniak (2023) (henceforth PW23) enumerate four main models[1]:

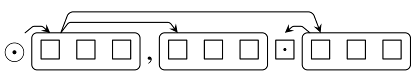
(1) a. **Conjunction-headed/Prague**



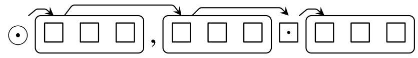
b. **Multi-headed/London**



c. **Bouquet/Stanford**



d. **Chain/Moscow**

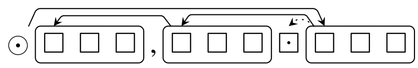


PW23 show that the asymmetrical approaches (1c)–(1d) cannot describe the English coordination structure correctly. Their argument is based on Dependency Length Minimization (DLM) – an universal and well-documented tendency to order words in sentences in a way so that long dependencies are avoided. (Temperley 2007, Futrell et al. 2015).

PW23's findings are replicated by subsequential studies including Przepiórkowski et al. (2024a) (from now on PBG24). The latter indicate that the London approach is probably the best description of the English coordination structure.

However, these conclusions cannot be extended for head-final languages such as Korean or Turkish. Kanayama et al. (2018) suggest that the coordination in head-final languages (HFL) may be asymmetrical. They propose a different approach taken from the work of Choi and Palmer (2011)[2]:

(2) **Right-headed/Inverted Moscow**



This paper aims to show that the approach shown in (2) might be the only one describing dependency coordination structure in HFL correctly. Using the methodology of PW23 and assuming the Dependency Length Minimization, it is demonstrated that using this approach the change in the tendency to put shorter conjunct at the beginning of coordination can be predicted more accurately than when using the other approaches.

[1]The following diagrams are based on those in the work of PW23. The governor is marked by ⊙, the conjunction by ⊡, and other tokens by □. Tokens belonging to the same conjunct are grouped. The names of the approaches in (1a)–(1d) are based on those in PW23. Apart from the approach shown in (1b), they were originally named by Popel et al. (2013).

[2]This approach assumes the head of the right conjunct to be the technical head of coordination and that each token is a dependent of the subsequent conjunct head. Those assumptions are inverted with respect to the Moscow approach. Note that Choi and Palmer (2011) do not specify which token is the governor of the conjunction. For reasons explained in §6.1 is assumed that in this approach the conjunction is the dependent of the head of its closest conjunct.

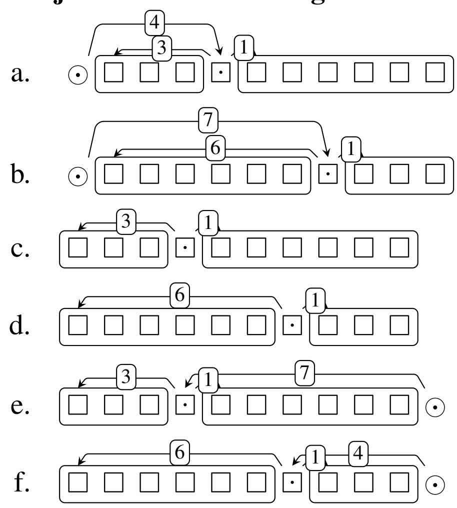
Proceedings of the 22nd Workshop on Treebanks and Linguistic Theories (TLT 2024), pages 65 - 75
December 5-6, 2024 ©2024 Association for Computational Linguistics

2 Previous Work

PW23 examine the tendency to put the shorter conjunct of the coordination as the first. They show that, assuming DLM, each approach to the dependency structure of coordination can predict the change in this tendency as the absolute difference of the conjunct length grows. Their study takes into account only binary coordinations.

To summarize their method, let me present the predictions of one dependency structure of coordination model, namely the Prague approach. They compare the total dependency length in six cases:[3]

(3) **Conjunction-headed/Prague**



PW23 compare the total dependency length in cases with the same governor position, i.e. (3a) vs (3b), (3c) vs (3d) and (3e) vs (3f). E.g. the total length of dependencies in (3a) is $4 + 3 + 1 = 8$ tokens, and in (3b) it is $7+6+1 = 14$ tokens, so the absolute length difference is equal to $|8 - 14| = 6$ tokens. In scheme (3) the difference of the conjunct length is $|3 - 6| = 3$ tokens. This means that the Prague approach assumes that when the governor is on the left, the total dependency length is smaller when the shorter conjunct is on the left (3a) than when it is on the right (3b). With the growth of the conjunct length difference, the total dependency length difference also grows.

PW23 point out that out of each pair, the arrangement with the smallest total dependency length is the more probable the greater the difference between the conjuncts' length is[4]. Therefore, because of DLM, the greater the difference is, the more coordinations are expected to have the shorter conjunct on the left. This can be demonstrated as a change in a function $p_*(n)$, where $n > 0$ is the absolute difference between the conjunct lengths and $* \in \{L, -, R\}$ is the governor position.[5] The function value is the proportion of the coordinations with the shorter conjunct on the left to all coordinations with a given governor position.

PW23 show that each approach can predict the direction of $p_*(n)$ function slope by comparing the total dependency length in pairs. Moreover, they determine the true values of the proportions function by analyzing 21.8K English coordinations in PennTree Bank. Table 1 summaries the predictions of the direction of the $p_*(n)$ tendencies in English. The predictions are compared with the actual tendencies found by PW23 and PBG24.

	L	$-$	R
Prague	+	+	0
London	+	0	−
Stanford	+	+	+
Moscow	+	+	+
PW23	+	+	0
PBG24	+	0	−

Table 1: Predictions of the change of the $p_*(n)$ tendencies in English and the tendencies observed in previous works.

PW23 argue that only the symmetric approaches (namely the Prague and London models) predict the changes in the proportions correctly. While the predictions of the Prague approach match the observed tendencies, there is a difference between the actual changes in the proportions and the predictions of the London approach. They state that this difference can be explained by the DLM effect at grammar. PW23 point out that coordinations with the left governor are most frequent in English and the $p_L(n)$ is positive. Therefore, a tendency to put the shorter conjunct as the first has became a gen-

[3]The governor can be in one of the three positions (left, absent, right) and the shorter conjunct can be either the first (left) or the last (right). Technically, coordinations with the governor in the middle (between conjuncts) are possible but they are too uncommon to be analyzed. Edge labels show the length of the dependency measured in tokens.

[4]Note that the DLM is not the only factor taken onto account while ordering the conjuncts. There is a general tendency to put the shorter conjunct as the first one, which has multiple explanations (Lohmann, 2014). However, PW23 explain that the DLM is the only factor that depends on the governor position an the conjunct length at the same time. They assume that in the analysis of thousands of coordinations the influence of the remaining factors even out.

[5]The possible governor positions shall be understood as follow: L means the governor is on the left, R denotes a governor on the right and $-$ stands for a coordination with no governor. The $p_*(n)$ function is not defined explicitly in PW23 and it is taken from the work of Przepiórkowski et al. (2024b).

eral, grammatical rule in English. With the growth of the difference between the conjuncts length, the tendency is stronger. Hence, the observed $p_{-}(n)$ and $p_R(n)$ tendencies are distorted. This means that the important thing to compare is not the actual and predicted $p_*(n)$ tendencies, but rather the actual and predicted differences between various $p_*(n)$ tendencies.

PBG24 replicate PW23's study, analyzing the larger (11.5M coordinations) Corpus of Contemporary American English (COCA). Tendencies observed by them match the predictions of the London model without the need to refer to DLM-at-grammar. Because of that, they narrow down possible models to the London approach.

PW23 research only covers the matter of the structure of coordination in English, which is a head-initial language. Kanayama et al. (2018) claim that the dependency structure of coordination in head-final languages can be different. They point out that the development of the models in (1) was based on the research, arguments and intuitions regarding only head-final languages. They especially criticize the asymmetrical Stanford approach used in Universal Dependencies as incompatible with the head-final languages' conditions.

Kanayama et al. (2018) claim that forcing Japanese and Korean UD annotators to use the Stanford approach resulted in lowering the quality of their corpora. They show linguistic and empirical arguments towards an alternative approach proposed by Choi and Palmer (2011) and urge UD to allow using that model in HFL corpora.

This paper argues that allowing the Right-headed approach to the dependency structure of coordination in UD corpora of head-final languages would be beneficial. The claim is based on the results of the analysis of Turkish and Korean UD corpora using PW23 and PBG24's methodology and theoretical arguments taken from the work of Kanayama et al. (2018).

3 Data

Three Korean (Kaist, GSD and PUD) and nine Turkish (Kenet, Penn, Tourism, Atis, GB, FrameNet, BOUN, IMST and PUD) corpora have been analyzed. All corpora have been annotated in Universal Dependencies v. 2.13 and downloaded from UD's website (`https://universaldependencies.org/` in December 2023). The data has been annotated manually. Four Turkish corpora (Atis, GB, BOUN and PUD) have been annotated natively in UD style, others have been automatically converted from different style. In total 21.5K Korean and 19.6K Turkish coordinations have been analyzed.

Table 2 shows the number of coordinations with a specific position of the governor and the shorter conjunct.[6]

shorter conjunct	governor position left	absent	right
Korean			
left	294	4054	3999
right	89	1093	964
Turkish			
left	894	3263	4257
right	111	1052	880

Table 2: The number of coordinations with a specific position of the governor and the shorter conjunct in the HFL corpora.

4 Methods

In order to replicate the methodology of PW23, only binary coordinations should be taken into account. However, ignoring every coordination with more than two conjuncts could severely impact the result of the study. On the other hand, to analyze the impact of the DLM effect on the length of every conjunct of the coordination a new methodology would be needed. Therefore, in this study every coordination is treated as binary, i.e. no matter how many conjuncts are in it, the first conjunct is considered the left one and the last conjunct is considered to be the right one. If a coordination have more than two conjuncts, the middle ones are ignored.

To determine the slope of all $p_*(n)$ functions as well as the differences between them the coordinations are extracted and the lengths of their conjuncts are measured.

The process of delimiting the conjunct basing on dependency trees is non-trivial and cannot be automated with high accuracy[7] (Patejuk and

[6]Since the DLM effect can only be noticed when there is a difference between the length of the conjuncts, the coordinations in which both conjuncts have the same length are not taken into account in this analysis. Also, because of the small number of the coordinations with the governor in the middle, those are also ignored.

[7]E.g. in a phrase such as *long days and nights* it is not syntactically determined if the word *long* describes only the word *days*, or both *days* and *nights*. The conjuncts can be delimited as either *[[long days] and [nights]]* or *[long [[days]*

Przepiórkowski 2018, Przepiórkowski and Woźniak 2023). Therefore a heuristic-based algorithm has been used to extract coordinations. It is an HLF-adjusted version of the algorithm used in PBG24's analysis. It is depicted in the Appendix A.

Since the automated process is not fully reliable, it has been evaluated by a native speaker of the Turkish language. 60 coordinations have been sampled randomly and evaluated using two criteria: 1. the governor position has been determined correctly and 2. both conjuncts have been delimited exactly as they should be (putting aside the punctuation, as it does not affect the word count). 35 of Turkish coordinations have been extracted correctly, which resulted in overall 58% algorithm accuracy. The algorithm has not been evaluated in the Korean language analysis.

The dependency length can be measured in various ways (e.g. characters, syllables, words). The DLM effect is mostly connected with the number of new objects in the discourse. Since objects correspond to words, for the DLM analysis dependency length is measured in words understood as non-punctuation tokens (Futrell et al., 2020).

Once the conjunct lengths and governor positions are determined, the monofactorial logistical regression is performed[8] to calculate the slope of each of the $p_*(n)$ functions.

The code used for extracting coordinations and statistical analysis is publicly available in the repository at `https://github.com/wjstempniak/Dependency-Structure-of-Coordination`.

5 Results

Figure 1 depicts changes in the tendency to put the shorter conjunct at the beginning of the coordination with the growth of the absolute difference between the conjuncts length.

Table 3 shows the differences between the slopes of $p_L(n)$, $p_-(n)$ and $p_R(n)$ the tendencies computed using R's `emmeans::emtrends` function (Lenth, 2024). The $L/-$, $-/R$ and L/R columns show the difference of the steepness of two respective $p_*(n)$ tendencies, e.g. if $L/-$ is positive, the $p_L(n)$ tendency is more increasing than the $p_*(n)$ tendency. The difference of the steepness of the $p_*(n)$ and $p_\dagger(n)$ tendencies is henceforth referred as the $*/\dagger$ *contrast*.

Although it may seem that for both languages almost all contrasts are positive,[9] it is crucial to notice that the differences between the tendencies are highly insignificant. For all three contrasts, both in Korean and Turkish, p was greater than 0.2. In Turkish, for the L/R contrast (which had been expected to be the greatest and therefore most significant) p was equal to 0.83.

The insignificance of differences can either mean that there is no difference or may indicate the lack of sufficient quality and quantity of the data. How-

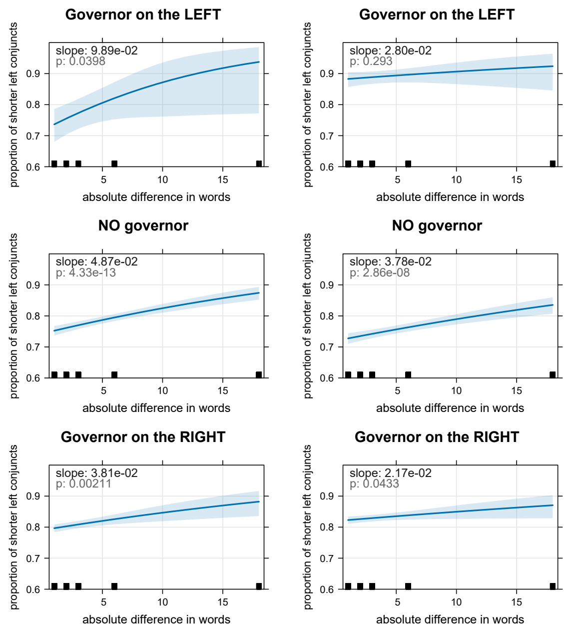


Figure 1: Observed $p_*(n)$ in Korean (left) and Turkish (right).

	$L/-$		$-/R$		L/R	
	diff	p	diff	p	diff	p
ko	0.05	0.30	0.01	0.45	0.06	0.22
tr	-0.01	0.72	0.02	0.21	0.01	0.83

Table 3: Differences between the steepness of observed $p_*(n)$ tendencies in HFL.

and [nights]]].

[8]In PW23, "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 $\delta = \langle 0, 1, 2, 3, 6, 25\rangle$", where "bucket i contains coordinations with absolute length differences within the interval $(i, i+1]$". In the HFL analysis, due to a low number of coordinations with the governor on the left, similar method was applied. To fit the data, the values of the δ vector have been adjusted to $\langle 0, 1, 2, 3, 6, 18\rangle$. The computations were performed using the R's `glm` function (R Core Team, 2023).

[9]Note that the only negative tendencies are the L and $L/-$ tendencies in Turkish. However, those are the tendencies concerning Turkish coordinations with the governor on the left. Because Turkish is a HFL, this type of coordination structure is rare in its corpora – in this study, there are only 1.8k Turkish coordinations with the governor on the left (opposed to 12k coordinations with the governor on the right).

ever, the total number of coordinations analyzed in Korean and Turkish was similar to the number of English coordinations analyzed by PW23 (Korean: 21.5K, Turkish: 19.6K vs. PW23: 21.8K). This suggest that if such differences exist, examining around 20k coordinations should be enough to find them. The fact that in this analysis no significant differences were found does not *prove* that there are no differences, but certainly indicates that this is probable. Given that, since most of the (insignificant) observed tendencies are positive, it is more probable that those differences are positive or neutral than negative.

The next section explains how these results can be interpreted in the context of the approaches to the dependency structure of coordination shown it (1) and (2).

6 Discussion

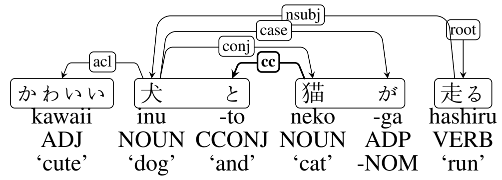
6.1 Dependency structure of coordination in head-final languages

For the analysis of binary coordinations in HFL several assumptions have to be made.

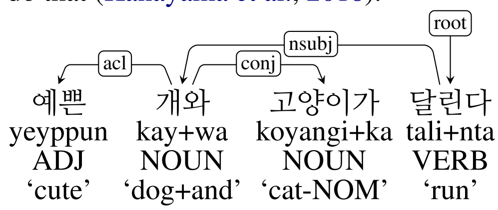
Firstly, in case of binary coordinations the Stanford and Moscow approaches are essentially the same.[10] The predictions of these approaches are the same regardless of the position of the governor, so for the analysis' sake one of them can be omitted.

Additionally, it is known that in HFL heads tend to be at the end of phrases. For the sake of the analysis, the head is assumed to be the final token of a coordination conjunct.[11]

Finally, it is safe to assume the conjunction is always dependent to the conjunct head next to it. This is a conclusion from the fact that in HFL the conjunction is often an agglutinate, suffix, or part of the word or phrase unit[12] that is the head of the conjunct. This is visible in Japanese and Korean examples below.[13]

(4)


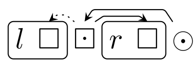
In (4) the conjunction token 'と' ('and') is a part of *bunsetsu* '犬と' ('dog and'). According to the Stanford approach 'と'('and') should be treated as a dependent of '猫' ('cat'). Assuming common sense and basic semantic intuition, there is no reason to do that (Kanayama et al., 2018).

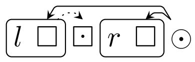
(5)


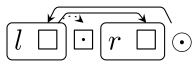
In (5), the conjunctive particle '와' is a suffix in *eojeol* '개와' ('dog and'), therefore it does not constitute an individual token and is not a dependent to any conjunct head.

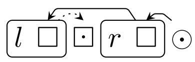
For these reasons assume that in HFL if the conjunction is a separate token, it is connected to the right head. Thus, the dependency length between the conjunction and the head is constant and its influence on the total dependency length is negligible. Therefore, that dependency is ignored.

Taking into account the assumptions stated above, the following approaches are analyzed:[14]

(6) a. **Prague** 

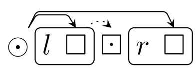
b. **London** 

c. **Stanford** 

d. **Inverted** 

6.2 Predictions of different approaches

In order to determine the predictions, the pairs of cases with the three different governor positions are compared (now using the London approach as the example):

(7) a. 

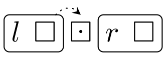
[10]The only difference between these models is the two relations between the head of the right conjunct and the conjunction. The sum of these relations is the same in all six cases so the influence of this relations on the total dependency length is none.

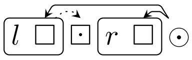
[11]It is a simplification, because the head is not always the final token. However, the relevant factor is that there are more potential dependents on the left than on the right side of the head. Appendix B shows the information about the relative position of the head within the conjuncts in the used data.

[12]Such as Japanese *bunsetsu* or Korean *eojeol*. See Kanayama et al. (2018) for details and examples.

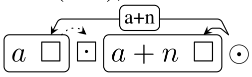
[13]The example sentences are adopted from the work of Kanayama et al. (2018) and are annotated according to the Stanford approach.

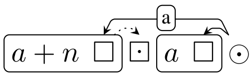
[14]The non-head tokens which are parts of the conjuncts (henceforth called the conjunct *body*) are replaced with l or r symbol for clarity. In these schemata, the governor is placed on the right side of the coordination to reflect the fact in HFL it strongly tends to be in that position. The analysis still covers coordinations with all three governor positions, so that does not interfere with the models' predictions.

b. 

c. 

For each possible governor position, the difference of the total dependency length is compared between the case where the first conjunct is shorter and the case where it is longer. Let a and $a+n$ be the length of the conjuncts, where $n > 0$. For the case shown in (14a), those cases are:

(8) a. 

b. 

In (8a) the total dependency length is equal to $a{+}n$ and in (8b) it is equal to a. Therefore, the absolute difference is equal to n. The total dependency length in (8a) is greater than the total dependency length in (8b). Thus the prediction of the London approach for HFL is that when the governor is on the right, the proportion of coordinations with the shorter conjunct on the left is decreasing with the growth of the absolute difference of the conjunct length. In other words, the London approach predicts that $p_R(n)$ is decreasing.

To shorten the calculations let me formalize them. Let S_* be the sum of the *relevant*[15] dependencies' length in the case where the first conjunct is shorter, and S'_* be the sum of the *relevant* dependencies' length in the other case. The model predicts that the function $p_*(n)$ is increasing if and only if S'_* is greater than S_*. Finally, let $e_*(n)$ be a function such that

(9) $e_*(n) = S'_* - S_*$.

The function $e_*(n)$ estimates the direction of the slope of $p_*(n)$ in the way that the sign of $e_*(n)$ is equal to the direction of the slope of $p_*(n)$ for $* \in \{L, -, R\}$.

Let $\overline{l_*}$ ($\overline{r_*}$) be the number of dependencies that go *over* the left (right) conjunct's body. In (8a) there are 0 dependencies going over the left conjunct body and there is 1 dependency going over the right conjunct's body. Since the conjunction is ignored, to compute the total dependency length the product of the number of dependencies and their length, which is equal to the conjunct length, is simply added. The total dependency length in (8a) is $0a + 1(a+n) = a+n$.

This can be generalized as

(10) a. $S_* = \overline{l_*}a + \overline{r_*}(a+n)$ and

b. $S'_* = \overline{l_*}(a+n) + \overline{r_*}a$.

Recall from (9) that $e_*(n) = S'_* - S_*$. Thus

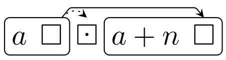
(11) a. $e_*(n) = \overline{l_*}(a+n) + \overline{r_*}a - (\overline{l_*}a + \overline{r_*}(a+n))$, which can be simplified to

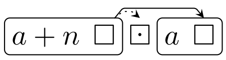
b. $e_*(n) = (\overline{l_*} - \overline{r_*})((a+n) - a)$, so

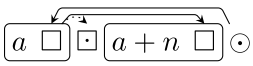
c. $e_*(n) = (\overline{l_*} - \overline{r_*}) \cdot n$

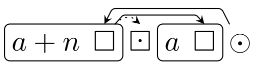
Because $n > 0$, the estimating function $e_*(n)$ is increasing if and only if $\overline{l_*} - \overline{r_*} > 0$. That means the prediction of the model of the direction of $p_*(n)$ slope is equal to sign of $\overline{l_*} - \overline{r_*}$.

The DLM can be understood as a probability function from total dependency length in a given case to a probability that this case occurs in natural language. Assuming that the function is monotonous, the differences between the slopes for each governor position pair can be predicted. Let me show how to do that using the example of the $-/R$ contrast in the Stanford approach for HFL:

(12) a. 

b. 

c. 

d. 

In (12), the $e_-(n) = a - (a+n) = -n$, so the slope of $p_-(n)$ is expected to be negative. However, the $e_R(n) = 2a - 2(a+n) = 2n$, so the slope of $p_R(n)$ is expected to be negative and smaller than the slope of $e_-(n)$.

This observation can be generalized using the $e_{*/\dagger}(n)$ function such as

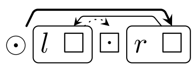
(13) a. $e_{*/\dagger}(n) = e_*(n) - e_\dagger(n)$ or

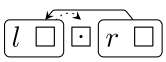
b. $e_{*/\dagger}(n) = (\overline{l_*} - \overline{r_*} - (\overline{l_\dagger} - \overline{r_\dagger})) \cdot n$

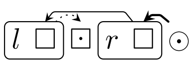
for $*, \dagger \in \{L, -, R\}$ and $n > 0$.

Hence, the sign of the contrast between $p_*(n)$ and $p_\dagger(n)$ functions is equal the sign of $e_{*/\dagger}(n)$.

Consider again the three cases described in (7).[16]

(14) a. 

b. 

c. 

[15]The dependencies within the conjunct are ignored as they are constant and independent from changes in the conjunct order and the governor position.

[16]In the following example, the coordinations are annotated according to the Inverted approach. However, the implications below are true for every model of the dependency structure of coordination.

All dependencies that are present in (14b) are also present in (14a) and (14c). Moreover, the dependencies going *over* the left conjunct body can be divided into two groups: 1. those present when there is no governor (in (14b) there are $\overline{l_-}$ of them) and 2. those connecting the governor with its dependents (in (14a) and (14c), they are thickened). Let $\overline{L_*}$ (and $\overline{R_*}$) be the number of relations connecting the governor with its dependents going *over* the left (right) body conjunct. From the observation above it is visible that

(15) a. $\overline{l_L} = \overline{l_-} + \overline{L_L}$, or $\overline{L_L} = \overline{l_L} - \overline{l_-}$
and, similarly
b. $\overline{r_L} = \overline{r_-} + \overline{R_L}$, or $\overline{R_L} = \overline{r_L} - \overline{r_-}$.

Since the dependencies present in (14b) are also present in (14a) and (14c), when comparing the difference between different slopes they can be omitted. In other words, while computing the contrast between $p_*(n)$ functions the only *relevant* dependencies are those connecting the governor with its dependents.

Recall from (11c) and (13) that:

(16) $e_{*/\dagger}(n) = e_*(n) - e_\dagger(n)$

(17) $e_*(n) = (\overline{l_*} - \overline{r_*}) \cdot n$

From (11c):

(18) a. $e_L(n) = (\overline{l_L} - \overline{r_L}) \cdot n$,
b. $e_-(n) = (\overline{l_-} - \overline{r_-}) \cdot n$,
c. $e_R(n) = (\overline{l_R} - \overline{r_R}) \cdot n$.

From (13):

(19) a. $e_{L/-}(n) = e_L(n) - e_-(n)$,
b. $e_{L/-}(n) = (\overline{l_L} - \overline{r_L}) \cdot n - (\overline{l_-} - \overline{r_-}) \cdot n$,
c. $e_{L/-}(n) = (\overline{l_L} - \overline{l_-} - (\overline{r_L} - \overline{r_-})) \cdot n$,
d. $e_{L/-}(n) = (\overline{L_L} - \overline{R_L}) \cdot n$.

Similarly, it is provable that

(20) a. $e_{-/R}(n) = e_-(n) - e_R(n)$,
b. $e_{-/R}(n) = (\overline{l_-} - \overline{r_-}) \cdot n - (\overline{l_R} - \overline{r_R}) \cdot n$,
c. $e_{-/R}(n) = (\overline{l_-} - \overline{l_R} - (\overline{r_-} - \overline{r_R})) \cdot n$,
d. $e_{-/R}(n) = (\overline{L_R} - \overline{R_R}) \cdot n$.

To sum up, the predictions of a given model are signs of functions shown in (18), (19d) and (20d) for $n > 0$. Using these formulae, predictions for HFL can be computed for every model (see Table 4).

Model	$e_L(n)$	$e_-(n)$	$e_R(n)$	$e_{-/R}(n)$	$e_{L/-}(n)$
Prague	0	$-n$	$-2n$	$-n$	$-n$
London	n	0	$-n$	$-n$	$-n$
Stanford	0	$-n$	$-2n$	$-n$	$-n$
Inverted	$-n$	$-n$	$-n$	0	0

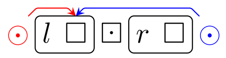
Table 4: Values of estimating function for HFL.

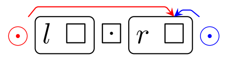
Recall from Table 1 that all three observed $p_*(n)$ tendencies are positive. This means that none of the considered approaches predict the slope direction itself correctly. This may be due to a strong, universal tendency to put the shorter conjunct at the beginning of the coordination which has a different cause.[17] For this reason it is important to compare the predicted and observed differences between the tendencies steepness (i.e. the contrasts) rather than the predicted and observed $p_*(n)$ tendencies.

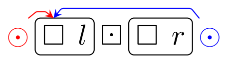
As it is visible in Table 3, every approach predicts that the $L/-$ and $-/R$ contrasts are either negative or none. However, the results of the study suggest that the contrast is is more likely to be positive. One might say that there can be other approaches to the dependency structure of coordination in HFL which predict the contrasts to be positive. The following section proves that such an approach is impossible.

6.3 All possible approaches

To cover all approaches, let me analyze possible dependents of the governor of the coordination. The governor's dependent can be either to the left of the left conjunct's body (as in 21c), to the right of the right conjunct body (as in 21b), or between the conjuncts' bodies (as in 21a).

(21) a. 

b. 

c. 

From (21a)–(21c) it is visible that irrespective of the assumed approach for every dependency connecting the governor on the left going *over* the left conjunct body either this dependency is also *over* the right conjunct body (21b) or there is another dependency *over* the right conjunct body when the governor is on the right (21a).

The same can be said about the dependency connecting the governor on the right going *over* the right conjunct body and the dependency goes *over* the left conjunct body when the governor is on the left. This can be written as

(22) a. $\overline{L_L} + \overline{R_L} = \overline{L_R} + \overline{R_R}$, or
b. $\overline{L_L} - \overline{L_R} = \overline{R_R} - \overline{R_L}$.

[17]This tendency is observed in multiple previous works and explained in a numerous ways, including arguments based on pragmatics (Lohmann, 2014), psycholinguistics (McDonald et al., 1993), stress patterns (Wright et al., 2005) and DLM-at-grammar (PW23).

Given that:

(23) a. $e_{L/-}(n) = (\overline{L_L} - \overline{L_R}) \cdot n$
b. $e_{L/-}(n) = (\overline{R_R} - \overline{R_L}) \cdot n$
c. $e_{L/-}(n) = e_{-/R}(n)$

Furthermore, from (21b) it is visible that if there is a dependency connecting the left governor going *over* the right conjunct body, this dependency goes also *over* the left conjunct body. Thus

(24) a. $\overline{L_L} \geq \overline{L_R}$ or $\overline{L_L} - \overline{L_R} \leq 0$, and since
b. $e_{L/-}(n) = \overline{L_L} - \overline{L_R}$ and
c. $e_{L/-}(n) = e_{-/R}(n)$, therefore
d. $e_{L/-}(n)$ and $e_{-/R}(n)$ are non-increasing functions.

Therefore, irrespective of the assumed approach to the dependency structure of coordination, a model can either predict that the slope of $p_L(n)$ can be either the same or more decreasing than the slope of $p_-(n)$. The same can be said about the $p_-(n)$ and $p_R(n)$ slopes. This is true for all possible models consistent with the assumptions made in this paper. This leads to the conclusion that a model predicting positive contrasts between the slopes is impossible.

It is well known that the DLM affect word order both in individual sentences (Futrell et al., 2015) and at the grammatical level (PW23). However, it is possible that the DLM influences the shape of the grammatical structure itself as well. [18]

In case of head-initial languages, a symmetric coordination allows to use DLM efficiently, as putting the short conjunct first in coordinations with the governor on the left indeed shortens the total length of the dependencies (because the $L/-$ and $L/-$ contrasts are negative). However, in case of HFL, there is no possible approach that would allow shortening the total length of the dependencies by putting the short conjunct last in coordinations with the right governor (because the $-/R$ and L/R contrasts are also negative or none). Using any of the four main approaches shown in (1) in HFL would cause almost every coordination to have excessively long dependencies.

This may explain why head-final languages may have formed an inverted-approach dependency structure of coordination, opposed to head-initial languages which evolved a symmetric one.

7 Limitations

The main drawback of the presented research is the quality and quantity of used data. As stated in the work of Kanayama et al. (2018), UD-imposed bounds "tied the hands" of HFL corpora annotators and forced them to work out compromises, which reduced the corpora quality. This include "dropping the conjunction category entirely in the case of Japanese" (Kanayama et al., 2018, p. 82), which made an analysis of coordination in Japanese UD corpora impossible. Given these restrictions, the relatively small[19] Korean and Turkish corpora were analyzed. Once more corpora of the head-final languages with coordinations marked consistently will be created, revisiting this study with more and better-quality data will become possible.

Apart from that, the algorithm used for extracting coordinations was highly imperfect. As stated before, determining the exact conjunct length based solely on dependency trees is a well-known common issue and cannot be solved automatically (Kanayama et al. 2018, Patejuk and Przepiórkowski 2018). The evaluation process showed that 58% of all Turkish coordinations were extracted correctly.[20] However, the algorithm was not evaluated for the Korean language analysis.

Lastly, the author does not know Turkish nor

[18]It is intuitive that a grammatical structure for a simple clause should have a simple dependency structure with as short dependencies as possible. For that reason the Stanford approach declares the head of the left conjunct as the "technical head of the coordination"(De Marneffe et al., 2021) – because in the head-initial languages the governor of coordination tends to be at the left, and the head of the left conjunct tends to be at the beginning of this conjunct, i.e. next to the governor. Therefore, the head of the left conjunct is a dependent to the governor. In case of HFL, it is exactly opposite – the governor tends to be at the end of a coordination, which is next to the head of the right conjunct. Because of that, the Stanford and Inverted approaches can be intuitive for respectively head-initial and head-final languages users and not intuitive for the other language group users. The need to minimize the length of dependencies is also the reason why Choi and Palmer (2011) decided to invert the Moscow approach, and not the Stanford model – because a hypothetical Inverted Stanford model would have to long dependencies between conjunct heads in case of long coordinations.

[19]There are 446K tokens in Korean and 735K in Turkish UD corpora opposed to 2.6M in Japanese UD corpora.

[20]The recurrent issue with Turkish corpora has been that two unrelated simple sentences have been treated as one coordination without a governor. In fact, 20% of coordinations in the sample have been a part of a run-on sentence that had been incorrectly marked as a coordinations. However, this is an issue with corpora, not with the algorithm. However, this problem does not necessarily affect the results of the study. There is no pattern in the incorrectly extracted coordinations, so in a long run all influence from the invalid data points should even out. Moreover, this effect seems to apply only to the coordinations with no governor. If the issue affected strongly the $p_-(n)$ slope, the $L/-$ and $-/R$ contrasts would be significantly different. This, however, did not happened. Overall, while there is no reason to state that this might affect the study result, that cannot be ruled out.

Korean and therefore does not have head-final intuitions internalized, leaving space for an error arising from head-initial-based unconscious assumptions. A native speaker of Turkish has been consulted for the development of heuristics and has performed the evaluation of the algorithm. However, no Korean native speaker was involved in the analysis.

8 Conclusions

The main subject of the study is the analysis of the contrast between the tendencies to put the shorter coordination conjunct at the beginning as the absolute difference between the conjunct length grows with a given governor position. For head-initial languages there are approaches that predicts that the $L/-$ and $-/R$ contrasts are negative, so it may be intuitive to say that for the head-final languages there are approaches that predict that the contrast are positive. The main novel contributions of this paper is the proof that no model of the dependency structure of coordination can predict that. In other words, this paper proves that irrespective of an assumed approach either all tendencies are predicted to be either the same or more descending in the order: left-absent-right governor. This is true for both head-initial and head-final languages.

Additionally, the paper explains why the arguments for a symmetric approach to the dependency structure of coordination provided by PW23 and PBG24 cannot be extend to head-final languages. Moreover, the negative replication of aforementioned works in addition to experiments provided by Kanayama et al. (2018) and their theoretical arguments leads to a conclusion that Choi and Palmer's (2011) Inverted Stanford/Moscow approach probably describes best the dependency structure of coordination in head-final languages. However, the lack of significant differences in slopes does not *prove* that there is no difference between them, so the result of the experimental part of this research remains negative. A further analysis with more and better-quality data is needed to strengthen these claims.

As stated in Kanayama et al. (2018), better data could be obtained if Universal Dependencies allowed the HFL corpora annotators to annotate coordinations according to approaches that are more intuitive for HFL users. That would increase the number of correctly annotated coordinations in Korean and Japanese at the cost of universality. However, since it is possible for head-initial and head-final languages to have a different coordination structure, the dependency structure of coordination may not be universal across languages. Therefore, there is a need for a possibility to annotate coordinations differently in the Universal Dependency standard.

Acknowledgements

The analysis described in this paper was part of the author's bachelor's thesis at the Cognitive Science programme at the University of Warsaw (Stempniak, 2024). The idea to conduct this research was devised by the supervisor of the thesis, Professor Adam Przepiórkowski. The paper benefited from comments by him, the three anonymous TLT 2024 reviewers, and Magdalena Borysiak. I am grateful to Berke Şenşekerci for his help with the development of the heuristics for the head-final languages analysis as well as for evaluating the algorithm for the analysis of Turkish coordinations.

References

Jinho D Choi and Martha Palmer. 2011. Statistical dependency parsing in Korean: From corpus generation to automatic parsing. In *Proceedings of the Second Workshop on Statistical Parsing of Morphologically Rich Languages*, pages 1–11.

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

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

Richard Futrell, Kyle Mahowald, and Edward Gibson. 2015. Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112(33):10336–10341.

Hiroshi Kanayama, Na-Rae Han, Masayuki Asahara, Jena D Hwang, Yusuke Miyao, Jinho D Choi, and Yuji Matsumoto. 2018. Coordinate structures in Universal Dependencies for head-final languages. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 75–84.

Russell V. Lenth. 2024. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.10.0.

Arne Lohmann. 2014. *English coordinate constructions*. Cambridge University Press.

Janet L McDonald, Kathryn Bock, and Michael H Kelly. 1993. Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, 25(2):188–230.

Agnieszka Patejuk and Adam Przepiórkowski. 2018. *From Lexical Functional Grammar to Enhanced Universal Dependencies: Linguistically informed treebanks of Polish*. Instytut Podstaw Informatyki Polskiej Akademii Nauk, Warszawa.

Martin Popel, David Mareček, Jan Štěpánek, Daniel Zeman, and Zdeněk Žabokrtskỳ. 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.

Adam Przepiórkowski, Magdalena Borysiak, and Adam Głowacki. 2024a. An argument for symmetric coordination from dependency length minimization: A replication study. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 1021–1033.

Adam Przepiórkowski, Magdalena Borysiak, Adam Okrasiński, Bartosz Pobożniak, Wojciech Stempniak, Kamil Tomaszek, and Adam Głowacki. 2024b. Symmetric dependency structure of coordination: Crosslinguistic arguments from dependency length minimization. In *Proceedings of the 22nd Workshop on Treebanks and Linguistic Theories (TLT 2024)*, Hamburg, Germany.

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.

R Core Team. 2023. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Wiedeń, Austria.

Wojciech Stempniak. 2024. Struktura zależnościowa koordynacji – analiza korpusów universal dependencies. Bachelor's Thesis, University of Warsaw.

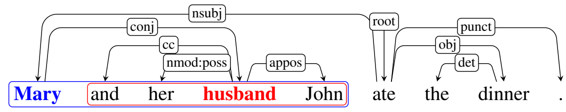
David Temperley. 2007. Minimization of dependency length in written English. *Cognition*, 105(2):300–333.

Saundra K Wright, Jennifer Hay, and Tessa Bent. 2005. Ladies first? Phonology, frequency, and the naming conspiracy. *Linguistics*, 43(3):531–561.

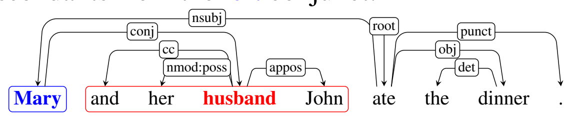
A Algorithm for determining the conjunct contents

This algorithm assumes that the coordination structure is annotated according to the current UD guidelines.

Consider all descendants of the conjuncts' **heads**:

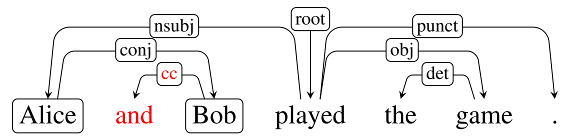


Exclude the head of the right conjunct and its descendants from the left conjunct:



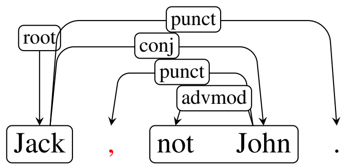
Then apply the following heuristics:

(H1) A conjunct cannot begin with a conjunction (a word that is connected to the head of the conjunct by a cc relationship).



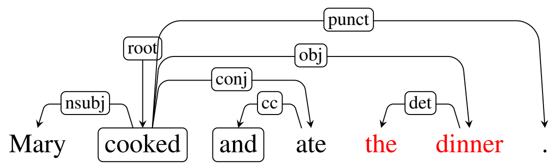
In the example above, *and* is not considered to be a part of the right conjunct because of (H1).

(H2) A conjunct cannot begin with a punctuation mark (specifically, a comma, semicolon, colon, or dash).



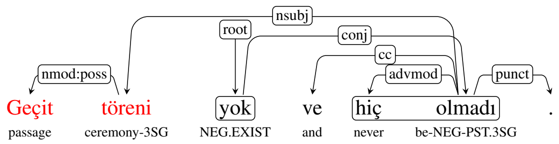
Though the comma is a descendant of the right conjunct head *John*, it is not a part of right conjunct because of (H2).

(H3) Left head descendants on the right side of the right conjunct are not a part of the left conjunct.



The goal of this heuristic is to exclude the tokens describing the both conjuncts that are dependents of the head of the left conjunct. The intuition supporting it it that the "private" dependents of the head of the left conjunct are almost always near this head and almost never at the right side of the right conjunct.

(H4) Right head descendants on the left side of the left conjunct are not a part of the right conjunct.

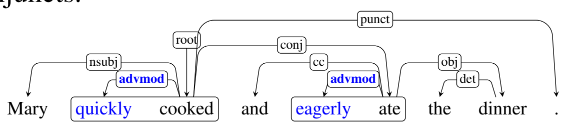


'There is no parade and there never was.'

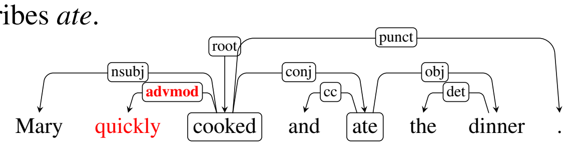
This heuristic has been developed specifically for head-final languages. It is an inverted version of (H4).

(H5) The child of the left conjunct head on the left side of the left head is not a part of left conjunct, if its relation with left head is *unique*, i.e. there is no relation between any other head and its child identical to it.

This is by far the most unreliable heuristic. Its goal is to tell apart the dependents of the head of the left conjunct describing the left conjunct exclusively from those describing all coordination's conjuncts.

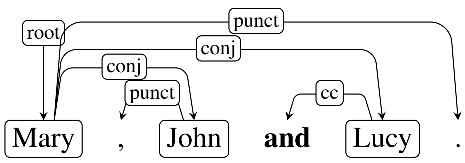


In the example above, the heads of both conjuncts have a dependent with an `advmod` relation. This means that this relation is not *unique*. Therefore, *quickly* describes *cooked*, and *eagerly* describes *ate*.

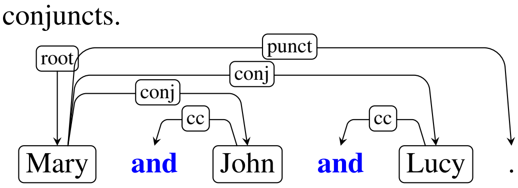


In this example, only the left conjunct head has a dependent with the `advmod` relation. Because of that, this relation is considered to be *unique*. According to (H5), *quickly* describes both *cooked* and *ate*.

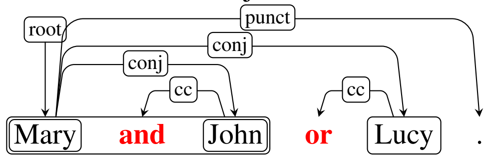
(H6) If there are multiple different **conjunctions** in a coordination, there is an extra coordination nested in it.



There is only one conjunction in this coordination – *and*. Therefore, this is one coordination with 3 conjuncts.



Here, there are two instances of the same conjunction – *and*. Because of this, this is also one coordination with 3 conjuncts.



In the example above, there are two distinct conjunctions – *and* and *or*. Therefore, in this sentence there are two coordinations with 2 conjuncts each, one nested in another.

B The relative head position within the conjunct

To confirm the assumption that most of the descendants of the conjunct heads are at the left side of the head, the relative position of the head within the conjunct is computed. The relative head position is defined by the formula

$$P = \frac{H-1}{N-1} \text{ for } N \geq 2,$$

where P denotes a relative position of the head; H is equal the absolute position of the head within the conjunct (i.e. the ordinal number of the token that is the head); and N is the conjunct length measured in tokens.

Table 5 shows the mean relative position of the heads of the right and left conjuncts of the coordinations found in the Korean and Turkish corpora.

	left conjunct		right conjunct	
	N	mean	N	mean
ko	6801	**0.78**	12951	**0.65**
tr	7994	**0.64**	12763	**0.69**

Table 5: The relative position of the head within the conjunct in HFL.

In all cases, the heads tend to be in the right half of the conjunct. This is confirmed using the Student's t-test testing the difference from 0.5 (i.e. the middle of the conjunct). All differences are highly significant ($p < 0.001$). Computations were performed using the R's `t.test` function (R Core Team, 2023).