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

Wojciech Stempniak

December 5th, 2024

The study on the UD corpora

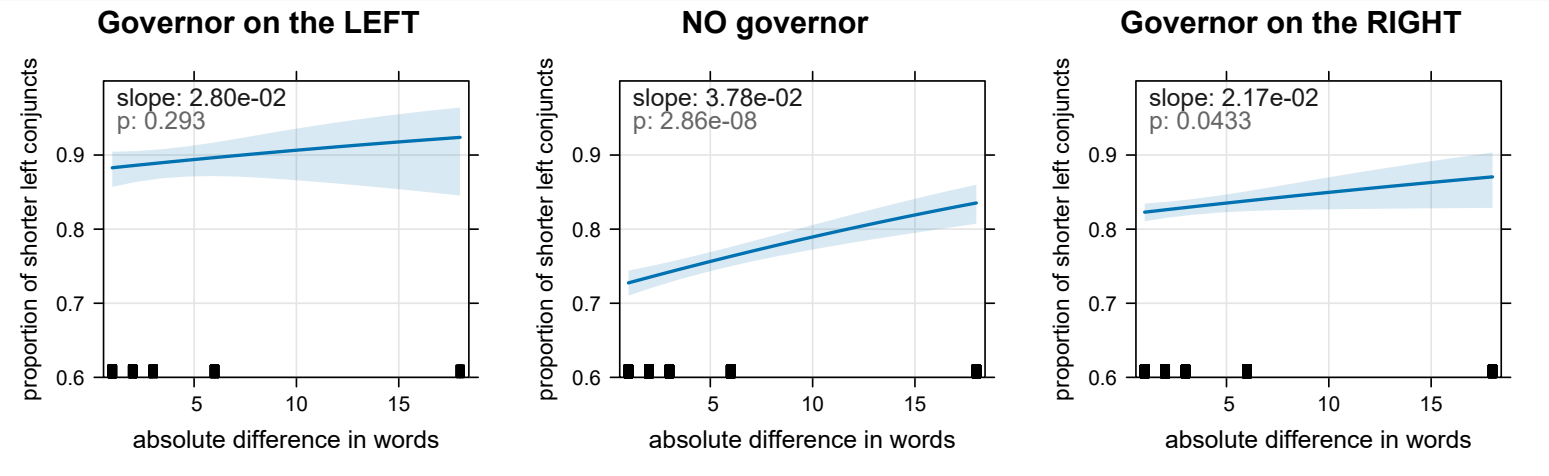
	tokens	coords	L/–	–/R	L/R	R
it	864K	25,426	$+^{***}$	+	$+^{*}$	$+^{**}$
la	983K	39,510	$+^{*}$	$+^{***}$	$+^{***}$	–
pt	1,361K	29,255	$+^{***}$	+	$+^{**}$	$+^{**}$
ro	938K	37,247	+	$+^{***}$	$+^{***}$	+
es	1,002K	28,666	$+^{***}$	+	$+^{*}$	$+^{**}$
en	718K	21,013	–	$+^{**}$	$+^{**}$	–
is	1,183K	43,852	$+^{***}$	–	$+^{*}$	$+^{*}$
cs	2,249K	90,566	$-^{***}$	$+^{***}$	$+^{***}$	$-^{*}$
pl	497K	16,684	–	$+^{*}$	+	+
ru	1,896K	61,004	+	$+^{***}$	$+^{***}$	–

(Przepiórkowski et al. 2024)

The study on the UD corpora – head-final languages

- Korean

- Turkish



The study on the UD corpora – head-final languages

	tokens	coords	L/–	–/R	L/R	R
ko	447K	21.5K	+	+	+	+**
tr	730K	19.6K	–	+	+	+*

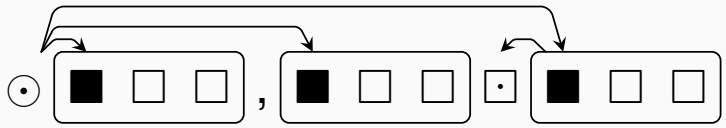
Question: How do we interpret them?

Intuition: Coordinations in head final languages “should” have “opposite” tendencies to those in head-initial languages.

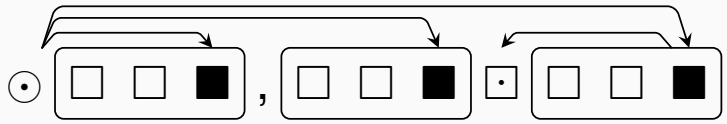
Coordination in head-final languages

Assumption 1. The head has more dependents on its left side than on its right side.

Head-initial languages:



Head-final languages:



Coordination in head-final languages

Assumption 1 was checked using the average relative head position defined by the formula

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

where:

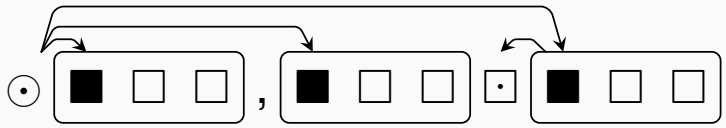
- P – the relative position of the head;
- H – the absolute position of the head within the conjunct;
- N – the conjunct length.

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

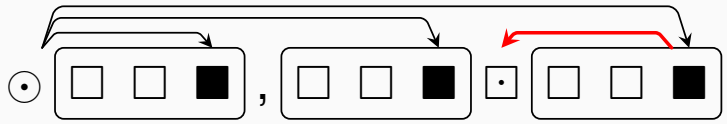
Coordination in head-final languages

Assumption 1. The conjunct head has more dependents on its left side than on its right side.

Head-initial languages:

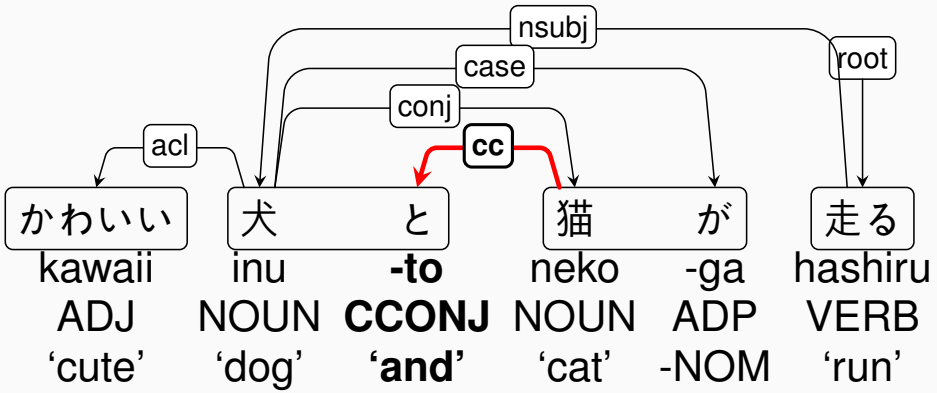


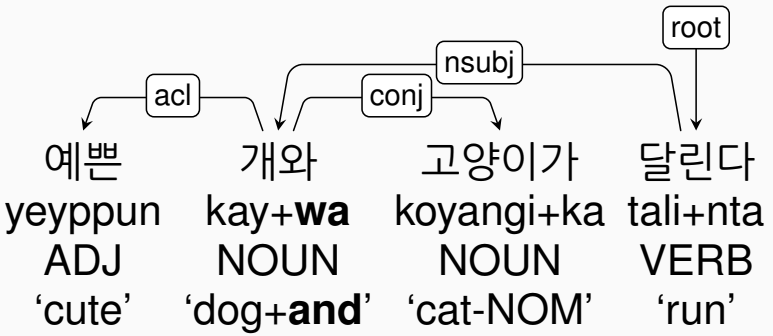
Head-final languages:



Coordination in head-final languages

Problem: Conjunctions in head-final languages do not work like that: (Kanayama et al. 2018)

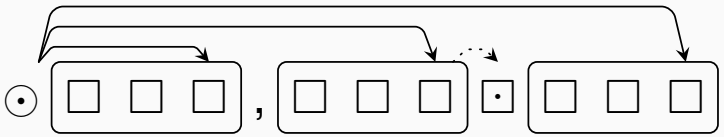




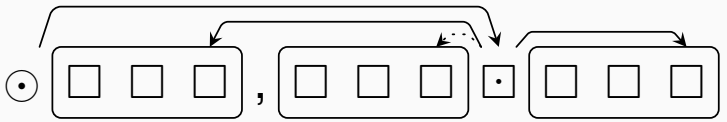
Coordination in head-final languages

Assumption 2. The conjunction is a dependent of the *closest* token head.

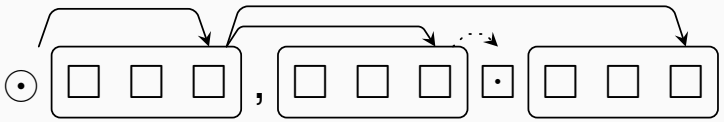
London (Multi-headed):



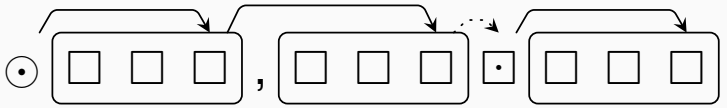
Prague (Conjunction-headed):



Stanford (Bouquet):



Moscow (Chain):

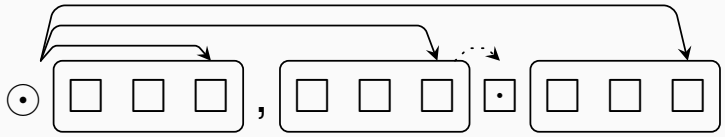


(Popel et al. 2013, Przepiórkowski and Woźniak 2023)

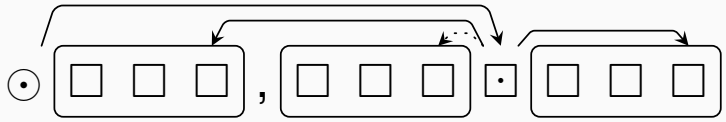
Coordination in head-final languages

Assumption 3. We must consider different approaches.

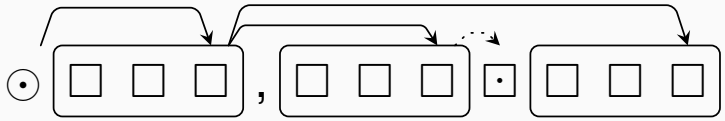
London (Multi-headed):



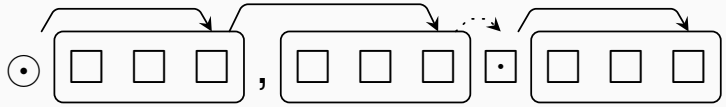
Prague (Conjunction-headed):



Stanford (Bouquet):



Moscow (Chain):

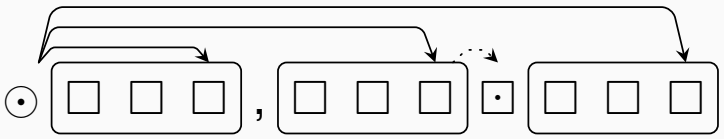


(Popel et al. 2013, Przepiórkowski and Woźniak 2023)

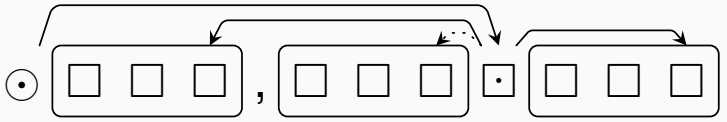
Coordination in head-final languages

Assumption 3. We must consider different approaches.

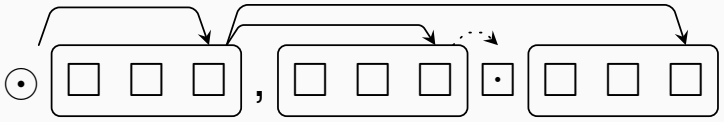
London (Multi-headed):



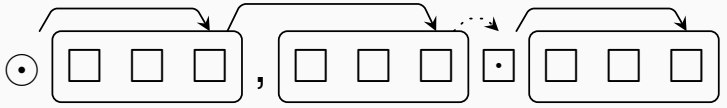
Prague (Conjunction-headed):



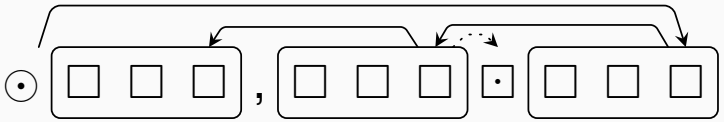
Stanford (Bouquet):



Moscow (Chain):



Inverted Moscow (Right-headed):

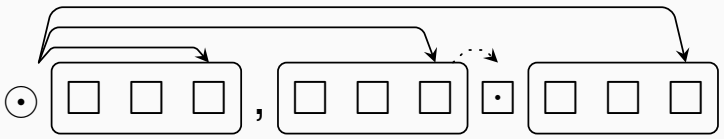


(Choi and Palmer, 2011)

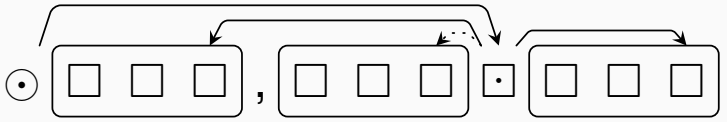
Coordination in head-final languages

The analysis covers only binary coordinations.

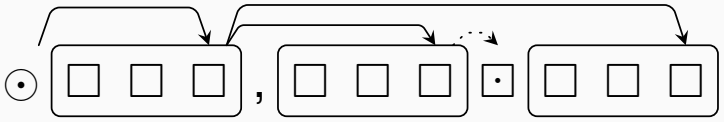
London (Multi-headed):



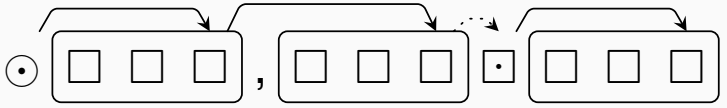
Prague (Conjunction-headed):



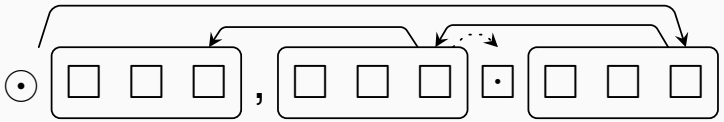
Stanford (Bouquet):



Moscow (Chain):



Inverted Moscow (Right-headed):

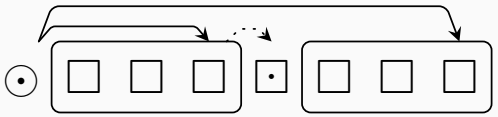


(Choi and Palmer, 2011)

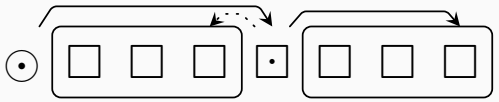
Coordination in head-final languages

The analysis covers only binary coordinations.

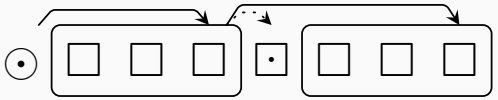
London (Multi-headed):



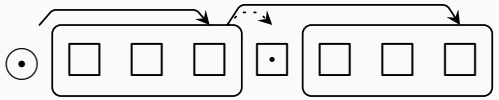
Prague (Conjunction-headed):



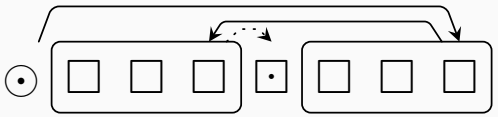
Stanford (Bouquet):



Moscow (Chain):



Inverted Moscow (Right-headed):

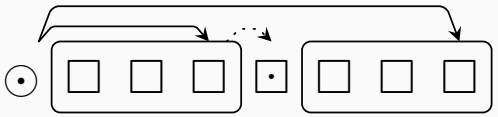


(Choi and Palmer, 2011)

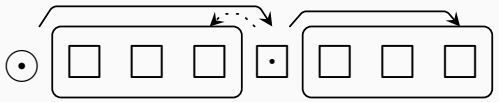
Coordination in head-final languages

The analysis covers only binary coordinations.

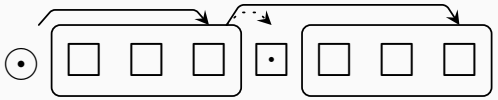
London (Multi-headed):



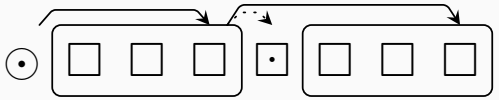
Prague (Conjunction-headed):



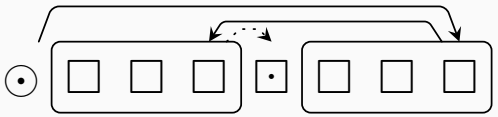
Stanford (Bouquet):



Moscow (Chain):



Inverted Moscow (Right-headed):

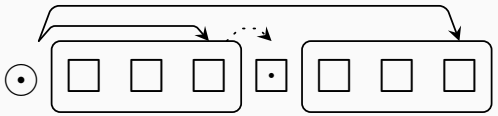


(Choi and Palmer, 2011)

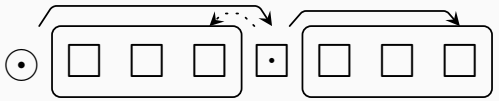
Coordination in head-final languages

Considered approaches:

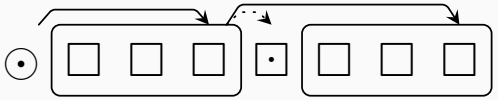
London (Multi-headed):



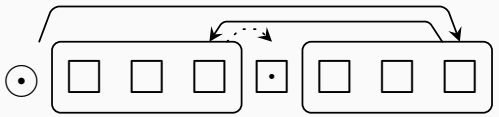
Prague (Conjunction-headed):



Stanford (Bouquet):



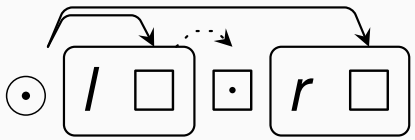
Inverted Moscow (Right-headed):



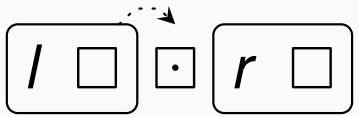
Computing the tendencies

For a given approach (here – London), we consider three types of coordinations:

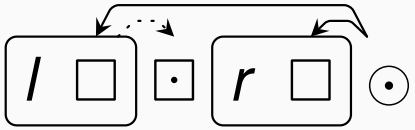
with governor on the left:



with no governor:



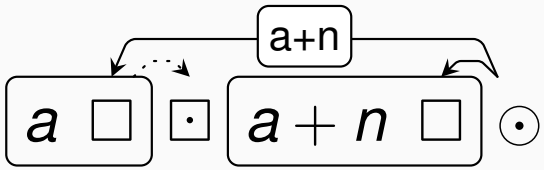
with governor on the right:



Computing the tendencies

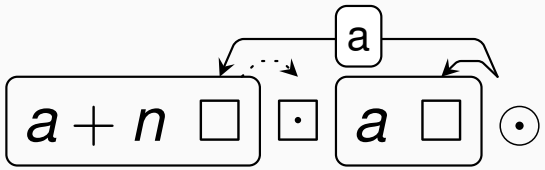
For every type of coordination, we compare the sum of total *relevant* dependencies:

shorter conjunct on the left



$S = a + n$

shorter conjunct on the right

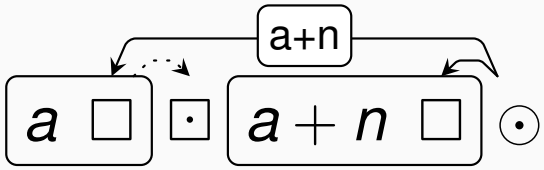


$S = a$

$n \geq 0$ is the absolute difference of the conjuncts length.

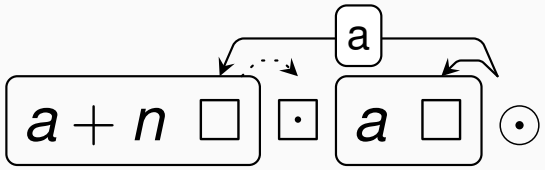
Computing the tendencies

shorter conjunct on the left



$S_R = a + n$

shorter conjunct on the right



$S'_R = a$

For a coordination with the governor position $* \in \{L, -, R\}$, the change in the tendency to put shorter conjunct at the beginning of coordination with the growth of the absolute difference of the conjuncts length $p_*(n)$ is predicted by the sign of the function $e_* = S'_* - S_*$.

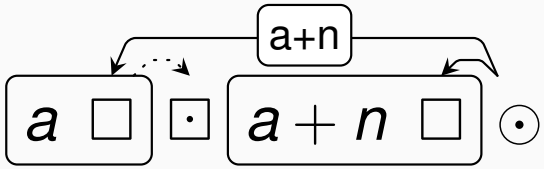
E.g. the London approach predicts that the $p_R(n)$ tendency is decreasing:

$$e_R(n) = S'_R - S_R = a - (a + n) = -n < 0$$

Computing the $p_*(n)$ tendencies

Let $\overline{l_*}$ ($\overline{r_*}$) be the number of relations going *over* the body of the left (right) conjunct.

shorter conjunct on the left



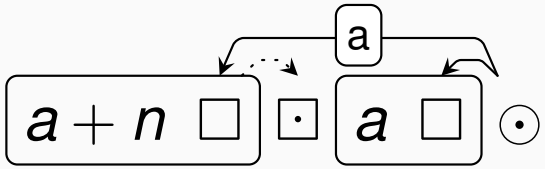
$$S_R = a + n$$

$$S_R = 0a + 1(a + n)$$

$$S_R = \overline{l_R}a + \overline{r_R}(a + n)$$

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

shorter conjunct on the right



$$S'_R = a$$

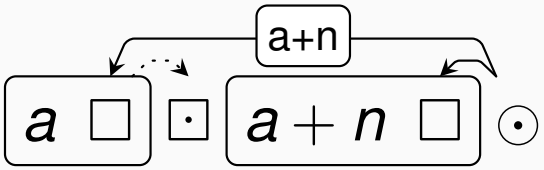
$$S'_R = 1a + 0(a + n)$$

$$S'_R = \overline{l_R}(a + n) + \overline{r_R}a$$

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

Computing the p_* tendencies

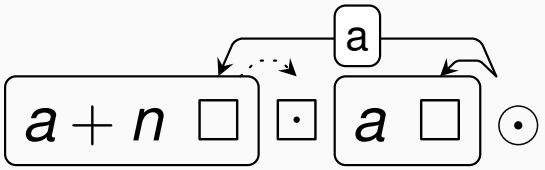
shorter conjunct on the left



$$S_R = a + n$$

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

shorter conjunct on the right



$$S'_R = a$$

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

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

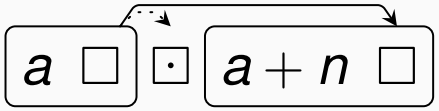
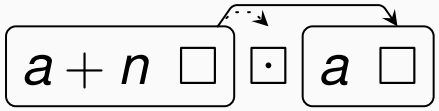
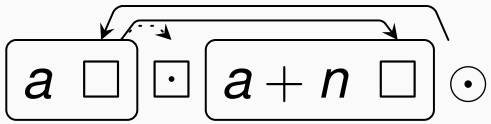
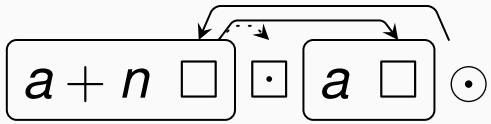
$$e_*(n) = \overline{l_*}(a + n) + \overline{r_*}a - (\overline{l_*}a + \overline{r_*}(a + n))$$

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

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

Since $n \geq 0$, the prediction of the model of the direction of $p_*(n)$ slope is equal to sign of $\overline{l_*} - \overline{r_*}$

Computing the contrasts between $p_*(n)$ tendencies

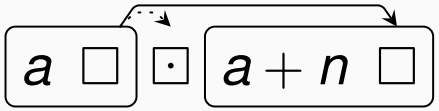
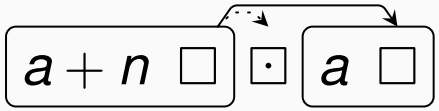
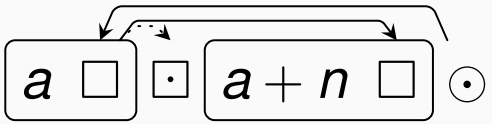
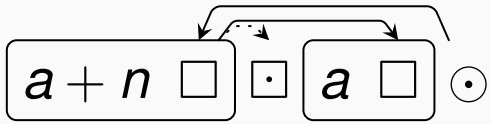
governor	shorter conjunct on the left	shorter conjunct on the right
absent		
right		

$$e_{-}(n) = a - (a + n) = -n$$

$$e_{R}(n) = 2a - 2(a + n) = -2n$$

The slope of $p_R(n)$ is expected to be *more* negative than the slope of $p_{-}(n)$ (assuming DLM's monotonicity).

Computing the contrasts between p_* tendencies

governor	shorter conjunct on the left	shorter conjunct on the right
absent		
right		

Generalize this using function:

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

$$e_{*/\dagger}(n) = (\overline{l_*} - \overline{r_*} - (\overline{l_\dagger} - \overline{r_\dagger})) \cdot n \text{ (for } *, \dagger \in \{L, -, R\} \text{ and } n > 0)$$

Predictions of the approaches

- Head initial-languages

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

- Head final-languages

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

The study on the UD corpora – head-final languages

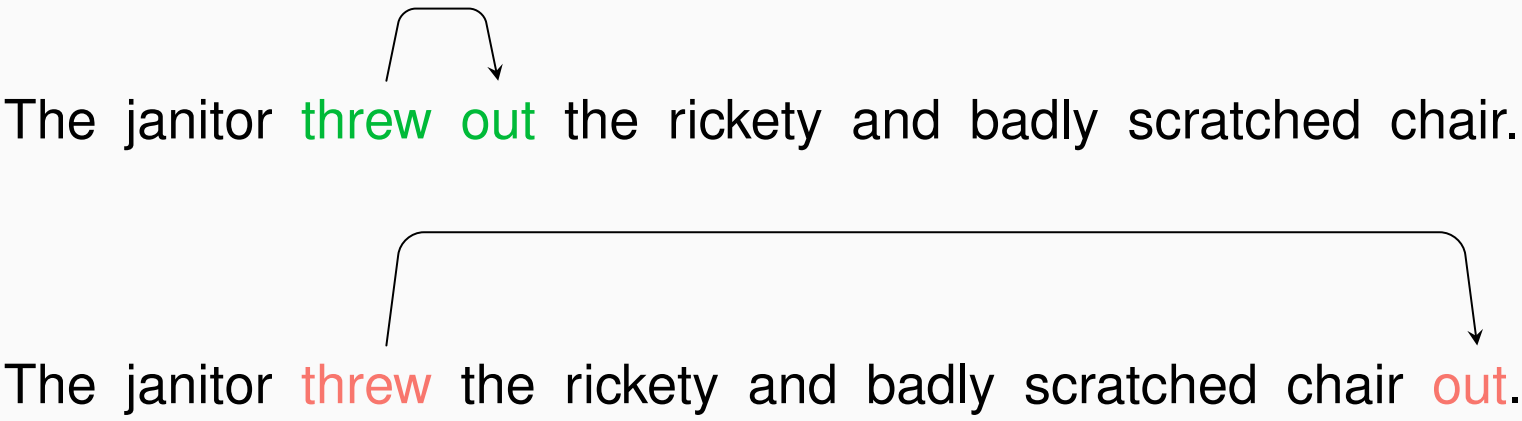
	tokens	coords	L/–	–/R	L/R	R
ko	447K	21.5K	+	+	+	+**
tr	730K	19.6K	–	+	+	+*

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

Dependency Length Minimization

We prefer sentences with shorter dependencies, because they are easier to process.

- The at-use DLM (Hunter and Prideaux, 1983):



- The at-grammar DLM (Hawkins, 1994):
  * Did $_{S}$[that John failed his exam] surprise Mary?

    Did $_{NP}$[that fact] surprise Mary?

Predictions of the approaches

- Head initial-languages

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

- Head final-languages

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

Summary and Limitations

- The experimental part of the study remains negative
- The algorithm used for the extraction is inaccurate and based on heuristics
  - Only 58% of all Turkish coordinations were extracted correctly.
  - Recurring issue: two unrelated simple sentences are treated as one coordination without a governor
- Problems with UD corpora (Kanayama et al. 2018)
  - No coordinations in Japanese corpora made analysis of tis language impossible

Bibliography

Choi, J. D. and Palmer, M. (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.

Hawkins, J. A. (1994). *A Performance Theory of Order and Constituency*. Cambridge University Press, Cambridge.

Hunter, P. J. and Prideaux, G. D. (1983). Empirical constraints on the verbparticle construction in English. Journal of the Atlantic Provinces Linguistic Association.

Kanayama, H., Han, N.-R., Asahara, M., Hwang, J. D., Miyao, Y., Choi, J. D, and Matsumoto, Y. (2018). Coordinate structures in Universal Dependencies for head-final languages. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 75–84.

Popel, M., Mareček, D., Štěpánek, J., Zeman, D., and Žabokrtskỳ, Z. (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.

Przepiórkowski, A. and Woźniak, M. (2023). Conjunct lengths in English, Dependency Length Minimization, and dependency structure of coordination. In Rogers, A., Boyd-Graber, J., and Okazaki, N., editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15494–15512, Toronto, Canada. Association for Computational Linguistics.

Przepiórkowski, A., Borysiak, M., Okrasinski, A., Pobożniak, B., Stempniak, W., Tomaszek, K., and Głowacki, A. (2024) 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.