

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

Adam Przepiórkowski,
Magdalena Borysiak, Adam Głowacki



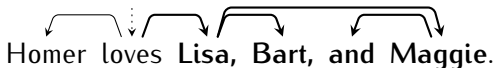
INSTITUTE OF COMPUTER SCIENCE
POLISH ACADEMY OF SCIENCES
ul. Jana Kazimierza 5, 01-248 Warsaw



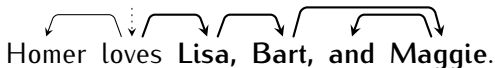
UNIVERSITY
OF WARSAW

LREC-COLING 2024

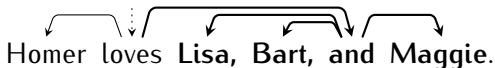
- Bouquet/Stanford (Universal Dependencies):



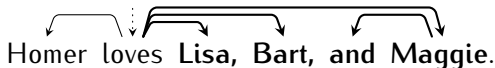
- Chain/Moscow:



- Conjunction-headed/Prague:



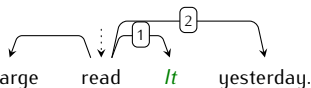
- Multi-headed/London:



Word order tends to minimize dependency lengths:



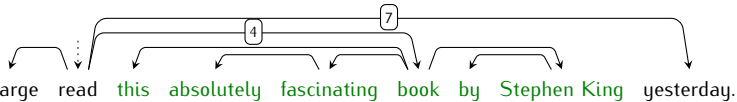
Word order tends to minimize dependency lengths:

- Marge read *It* yesterday. good


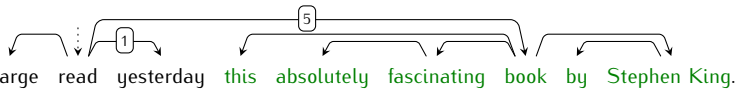
The diagram shows a dependency parse tree for the sentence "Marge read It yesterday.". The root node is "read". It has three children: "Marge", "It", and "yesterday.". The dependency between "read" and "It" is labeled with a boxed "1". The dependency between "read" and "yesterday." is labeled with a boxed "2".

- Marge read yesterday *It*. bad


The diagram shows a dependency parse tree for the sentence "Marge read yesterday It.". The root node is "read". It has three children: "Marge", "yesterday", and "It". The dependency between "read" and "yesterday" is labeled with a boxed "1". The dependency between "read" and "It" is labeled with a boxed "2".

- Marge read *this absolutely fascinating book by Stephen King* yesterday. good



The diagram shows a dependency parse tree for the sentence "Marge read this absolutely fascinating book by Stephen King yesterday.". The root node is "read". It has three children: "Marge", "this absolutely fascinating book by Stephen King", and "yesterday.". The dependency between "read" and "this" is labeled with a boxed "4". The dependency between "read" and "yesterday." is labeled with a boxed "7".

- Marge read yesterday *this absolutely fascinating book by Stephen King*. good


The diagram shows a dependency parse tree for the sentence "Marge read yesterday this absolutely fascinating book by Stephen King.". The root node is "read". It has three children: "Marge", "yesterday", and "this absolutely fascinating book by Stephen King". The dependency between "read" and "yesterday" is labeled with a boxed "1". The dependency between "read" and "this" is labeled with a boxed "5".

Bouquet/Stanford (Universal Dependencies):


- 
(governor on the left)

- 
 - When the difference of lengths increases,
 - the pressure for the shorter conjunct to be on the left grows
 - so, in a corpus, the proportions of coordinations with shorter left conjunct should **grow**.

- 
(governor absent)

- 
 - As above.

- 
(governor on the right)

- 
 - As above.

Bouquet/Stanford (Universal Dependencies):

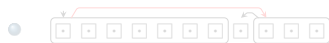
- 
(governor on the left)



- When the difference of lengths increases,
 - the pressure for the shorter conjunct to be on the left grows
 - so, in a corpus, the proportions of coordinations with shorter left conjunct should grow.



(governor absent)



- As above.



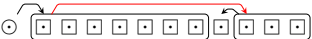
(governor on the right)



- As above.

Bouquet/Stanford (Universal Dependencies):


- 
(governor on the left)

- 
 - When the difference of lengths increases,
 - the pressure for the shorter conjunct to be on the left **grows**
 - so, in a corpus, the proportions of coordinations with shorter left conjunct should **grow**.

- 
(governor absent)

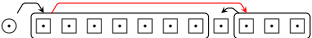
- 
 - As above.

- 
(governor on the right)

- 
 - As above.

Bouquet/Stanford (Universal Dependencies):


- 
(governor on the left)

- 
 - When the difference of lengths increases,
 - the pressure for the shorter conjunct to be on the left **grows**
 - so, in a corpus, the proportions of coordinations with shorter left conjunct should **grow**.

- 
(governor absent)


- 
 - As above.

- 
(governor on the right)

- 
 - As above.

Bouquet/Stanford (Universal Dependencies):


- 
(governor on the left)

- 
 - When the difference of lengths increases,
 - the pressure for the shorter conjunct to be on the left **grows**
 - so, in a corpus, the proportions of coordinations with shorter left conjunct should **grow**.

- 
(governor absent)

- 
 - As above.

- 
(governor on the right)

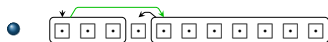
- 
 - As above.

Bouquet/Stanford (Universal Dependencies):

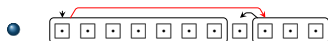
-  (governor on the left)



- When the difference of lengths increases,
- the pressure for the shorter conjunct to be on the left **grows**
- so, in a corpus, the proportions of coordinations with shorter left conjunct should **grow**.



(governor absent)



- As above.



(governor on the right)



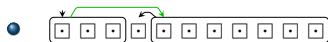
- As above.

Bouquet/Stanford (Universal Dependencies):

- 
(governor on the left)



- When the difference of lengths increases,
- the pressure for the shorter conjunct to be on the left **grows**
- so, in a corpus, the proportions of coordinations with shorter left conjunct should **grow**.



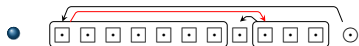
(governor absent)



- As above.




(governor on the right)



- As above.

Bouquet/Stanford (Universal Dependencies):


- 
(governor on the left)

- 
 - When the difference of lengths increases,
 - the pressure for the shorter conjunct to be on the left **grows**
 - so, in a corpus, the proportions of coordinations with shorter left conjunct should **grow**.

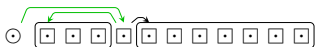





- 
(governor absent)

- 
 - As above.

- 
(governor on the right)

- 
 - As above.

Conjunction-headed/Prague:

- 
(governor on the left)
- 
 - As above.
- 
(governor absent)
- 
 - As above.
- 
(governor on the right)
- 
 - When the difference of lengths increases,
 - the pressure for the shorter conjunct to be on the left does not change
 - so, in a corpus, the proportions of coordinations with shorter left conjunct should be constant.

Conjunction-headed/Prague:



- As above.



(governor absent)



- As above.



(governor on the right)



- When the difference of lengths increases,
- the pressure for the shorter conjunct to be on the left does not change
- so, in a corpus, the proportions of coordinations with shorter left conjunct should be constant.

Conjunction-headed/Prague:



- As above.



(governor absent)



- As above.



(governor on the right)



- When the difference of lengths increases,
- the pressure for the shorter conjunct to be on the left does not change
- so, in a corpus, the proportions of coordinations with shorter left conjunct should be constant.

Conjunction-headed/Prague:



- As above.



- As above.



- When the difference of lengths increases,
- the pressure for the shorter conjunct to be on the left does not change
- so, in a corpus, the proportions of coordinations with shorter left conjunct should be constant.

Conjunction-headed/Prague:



- As above.

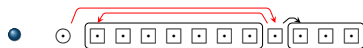


- As above.



- When the difference of lengths increases,
- the pressure for the shorter conjunct to be on the left does not change
- so, in a corpus, the proportions of coordinations with shorter left conjunct should be constant.

Conjunction-headed/Prague:



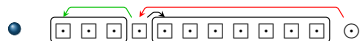
- As above.



(governor absent)

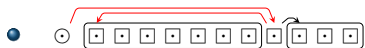


- As above.



- When the difference of lengths increases,
 - the pressure for the shorter conjunct to be on the left does not change
 - so, in a corpus, the proportions of coordinations with shorter left conjunct should be constant.

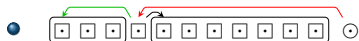
Conjunction-headed/Prague:



- As above.



- As above.



- When the difference of lengths increases,
- the pressure for the shorter conjunct to be on the left **does not change**
- so, in a corpus, the proportions of coordinations with shorter left conjunct should be constant.

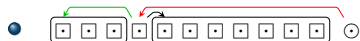
Conjunction-headed/Prague:



- As above.



- As above.



- When the difference of lengths increases,
- the pressure for the shorter conjunct to be on the left **does not change**
- so, in a corpus, the proportions of coordinations with shorter left conjunct should **be constant**.

Przepiórkowski and Woźniak 2023 (ACL 2023):

Statistics about coordination extracted from an enhanced version of the Penn Treebank (PTB; Marcus *et al.* 1993, Fidler and Goldberg 2016):

- left conjuncts tend to be shorter (observed before),
- this tendency grows with length difference (briefly noticed in Gibson *et al.* 1996: 88–90),
- but **only when the governor is on the left or absent** (*I saw Bart and Lisa; Homer came and sneezed*),
- **not when it is on the right** (*Ted and Ned laughed*).

Przepiórkowski and Woźniak 2023 (ACL 2023):

Statistics about coordination extracted from an enhanced version of the Penn Treebank (PTB; Marcus *et al.* 1993, Fidler and Goldberg 2016):

- **left conjuncts tend to be shorter** (observed before)
- this tendency grows with length difference (briefly noticed in Gibson *et al.* 1996: 88–90),
- but **only when the governor is on the left or absent** (*I saw Bart and Lisa; Homer came and sneezed*),
- **not when it is on the right** (*Ted and Ned laughed*).

Przepiórkowski and Woźniak 2023 (ACL 2023):

Statistics about coordination extracted from an enhanced version of the Penn Treebank (PTB; Marcus *et al.* 1993, Fidler and Goldberg 2016):

- **left conjuncts tend to be shorter** (observed before),
- **this tendency grows with length difference** (briefly noticed in Gibson *et al.* 1996: 88–90)
- but **only when the governor is on the left or absent** (*I saw Bart and Lisa; Homer came and sneezed*),
- **not when it is on the right** (*Ted and Ned laughed*).

Przepiórkowski and Woźniak 2023 (ACL 2023):

Statistics about coordination extracted from an enhanced version of the Penn Treebank (PTB; Marcus *et al.* 1993, Fidler and Goldberg 2016):

- **left conjuncts tend to be shorter** (observed before),
- this **tendency grows with length difference** (briefly noticed in Gibson *et al.* 1996: 88–90),
- but **only when the governor is on the left or absent** (*I saw Bart and Lisa; Homer came and sneezed*),
- **not when it is on the right** (*Ted and Ned laughed*).

Przepiórkowski and Woźniak 2023 (ACL 2023):

Statistics about coordination extracted from an enhanced version of the Penn Treebank (PTB; Marcus *et al.* 1993, Fidler and Goldberg 2016):

- **left conjuncts tend to be shorter** (observed before),
- this **tendency grows with length difference** (briefly noticed in Gibson *et al.* 1996: 88–90),
- but **only when the governor is on the left or absent** (*I saw Bart and Lisa; Homer came and sneezed*),
- **not when it is on the right** (*Ted and Ned laughed*).

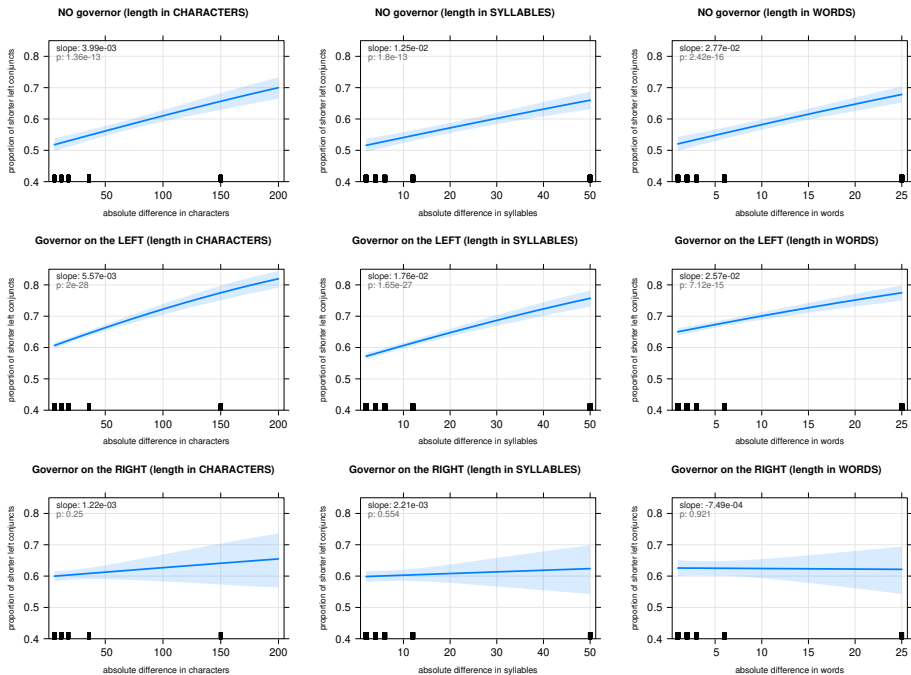
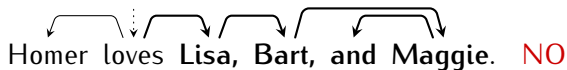


Figure 1: Proportions of shorter left conjuncts depending on the absolute difference of conjunct lengths (with confidence bands)

- **Bouquet/Stanford** (Universal Dependencies):



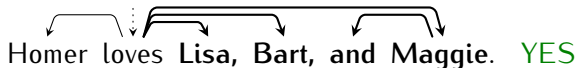
- **Chain/Moscow**:



- **Conjunction-headed/Prague**:



- **Multi-headed/London**:



Replication of Przepiórkowski and Woźniak 2023, but:

- instead of PTB we used large parts of the **Corpus of Contemporary American English** (COCA; Davies 2023),
- over 380M words (compared to 1.25M in PTB),
- 11.5M coordinations (compared to 24.5K in PTB),
- automatically parsed with Stanza (Qi *et al.* 2020) (instead of manual annotation in PTB).

Results:

- left conjuncts tend to be shorter (confirmed),
- this tendency grows with length difference,
- **only when the governor is on the left or absent** (confirmed)
(*I saw Bart and Lisa; Homer came and sneezed*),
- **not when it is on the right** (confirmed)
(*Ted and Ned laughed*).

Replication of Przepiórkowski and Woźniak 2023, but:

- instead of PTB we used large parts of the **Corpus of Contemporary American English (COCA; Davies 2023)**
- over 380M words (compared to 1.25M in PTB),
- 11.5M coordinations (compared to 24.5K in PTB),
- automatically parsed with Stanza (Qi *et al.* 2020) (instead of manual annotation in PTB).

Results:

- left conjuncts tend to be shorter (confirmed),
- this tendency grows with length difference,
- **only when the governor is on the left or absent** (confirmed)
(*I saw Bart and Lisa; Homer came and sneezed*),
- **not when it is on the right** (confirmed)
(*Ted and Ned laughed*).

Replication of Przepiórkowski and Woźniak 2023, but:

- instead of PTB we used large parts of the **Corpus of Contemporary American English** (COCA; Davies 2023),
- **over 380M words** (compared to 1.25M in PTB),
- 11.5M coordinations (compared to 24.5K in PTB),
- automatically parsed with Stanza (Qi *et al.* 2020) (instead of manual annotation in PTB).

Results:

- left conjuncts tend to be shorter (confirmed),
- this tendency grows with length difference,
- **only when the governor is on the left or absent** (confirmed)
(*I saw Bart and Lisa; Homer came and sneezed*),
- **not when it is on the right** (confirmed)
(*Ted and Ned laughed*).

Replication of Przepiórkowski and Woźniak 2023, but:

- instead of PTB we used large parts of the **Corpus of Contemporary American English** (COCA; Davies 2023),
- **over 380M words** (compared to 1.25M in PTB),
- **11.5M coordinations** (compared to 24.5K in PTB),
- **automatically parsed** with Stanza (Qi *et al.* 2020) (instead of manual annotation in PTB).

Results:

- **left conjuncts tend to be shorter** (confirmed),
- **this tendency grows with length difference**,
- **only when the governor is on the left or absent** (confirmed)
(*I saw Bart and Lisa; Homer came and sneezed*),
- **not when it is on the right** (confirmed)
(*Ted and Ned laughed*).

Replication of Przepiórkowski and Woźniak 2023, but:

- instead of PTB we used large parts of the **Corpus of Contemporary American English** (COCA; Davies 2023),
- **over 380M words** (compared to 1.25M in PTB),
- **11.5M coordinations** (compared to 24.5K in PTB),
- **automatically parsed** with Stanza (Qi *et al.* 2020) (instead of manual annotation in PTB).

Results:

- **left conjuncts tend to be shorter** (confirmed),
- **this tendency grows with length difference**,
- **only when the governor is on the left or absent** (confirmed)
(*I saw Bart and Lisa; Homer came and sneezed*),
- **not when it is on the right** (confirmed)
(*Ted and Ned laughed*).

Replication of Przepiórkowski and Woźniak 2023, but:

- instead of PTB we used large parts of the **Corpus of Contemporary American English** (COCA; Davies 2023),
- **over 380M words** (compared to 1.25M in PTB),
- **11.5M coordinations** (compared to 24.5K in PTB),
- **automatically parsed** with Stanza (Qi *et al.* 2020) (instead of manual annotation in PTB).

Results:

- **left conjuncts tend to be shorter** (confirmed),
- this **tendency grows with length difference**,
- **only when the governor is on the left or absent** (confirmed)
(*I saw Bart and Lisa; Homer came and sneezed*),
- **not when it is on the right** (confirmed)
(*Ted and Ned laughed*).

Replication of Przepiórkowski and Woźniak 2023, but:

- instead of PTB we used large parts of the **Corpus of Contemporary American English** (COCA; Davies 2023),
- **over 380M words** (compared to 1.25M in PTB),
- **11.5M coordinations** (compared to 24.5K in PTB),
- **automatically parsed** with Stanza (Qi *et al.* 2020) (instead of manual annotation in PTB).

Results:

- **left conjuncts tend to be shorter** (confirmed),
- this **tendency grows with length difference**,
- **only when the governor is on the left or absent** (confirmed)
(*I saw Bart and Lisa; Homer came and sneezed*),
- **not when it is on the right** (confirmed)
(*Ted and Ned laughed*).

Replication of Przepiórkowski and Woźniak 2023, but:

- instead of PTB we used large parts of the **Corpus of Contemporary American English** (COCA; Davies 2023),
- **over 380M words** (compared to 1.25M in PTB),
- **11.5M coordinations** (compared to 24.5K in PTB),
- **automatically parsed** with Stanza (Qi *et al.* 2020) (instead of manual annotation in PTB).

Results:

- **left conjuncts tend to be shorter** (confirmed),
- this **tendency grows with length difference**,
- **only when the governor is on the left or absent** (confirmed)
(*I saw Bart and Lisa; Homer came and sneezed*),
- **not when it is on the right** (confirmed)
(*Ted and Ned laughed*).

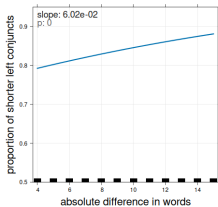
Replication of Przepiórkowski and Woźniak 2023, but:

- instead of PTB we used large parts of the **Corpus of Contemporary American English** (COCA; Davies 2023),
- **over 380M words** (compared to 1.25M in PTB),
- **11.5M coordinations** (compared to 24.5K in PTB),
- **automatically parsed** with Stanza (Qi *et al.* 2020) (instead of manual annotation in PTB).

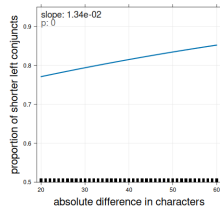
Results:

- **left conjuncts tend to be shorter** (confirmed),
- this **tendency grows with length difference**,
- **only when the governor is on the left or absent** (confirmed)
(*I saw Bart and Lisa; Homer came and sneezed*),
- **not when it is on the right** (confirmed)
(*Ted and Ned laughed*).

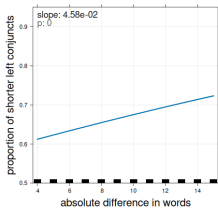
Governor on the LEFT (length in WORDS)



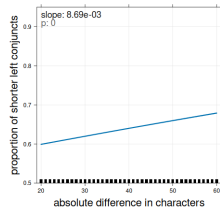
Governor on the LEFT (length in CHARACTERS)



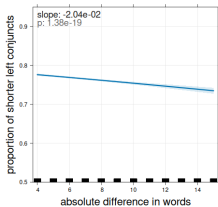
NO governor (length in WORDS)



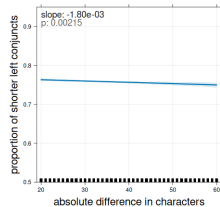
NO governor (length in CHARACTERS)



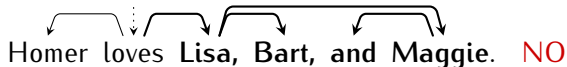
Governor on the RIGHT (length in WORDS)



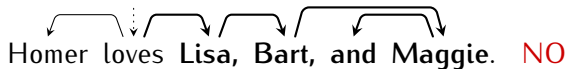
Governor on the RIGHT (length in CHARACTERS)



- **Bouquet/Stanford** (Universal Dependencies):



- **Chain/Moscow**:



- **Conjunction-headed/Prague**:



- **Multi-headed/London**:



See the paper for various subtleties!

Talk to us at the poster session!

See the paper for various subtleties!

Talk to us at the poster session!

- Davies, M. (2008–2023). The Corpus of Contemporary American English (COCA). Available online at <https://www.english-corpora.org/coca/>.
- Ficler, J. and Goldberg, Y. (2016). Coordination annotation extension in the Penn Tree Bank. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 834–842, Berlin, Germany.
- Gibson, E., Schütze, C. T., and Salomon, A. (1996). The relationship between the frequency and the processing complexity of linguistic structure. *Journal of Psycholinguistic Research*, 25(1), 59–92.
- Marcus, M. P., Santorini, B., and Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19, 313–330.
- Przepiórkowski, A. and Woźniak, M. (2023). Conjunct lengths in English, Dependency Length Minimization, and dependency structure of coordination. In A. Rogers, J. Boyd-Graber, and N. Okazaki, eds., *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 15494–15512, Toronto, Canada. Association for Computational Linguistics.
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J., and Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 101–108.