

# Dependency Length Minimization vs. Word Order Constraints: An Empirical Study on 55 Treebanks

Xiang Yu, Agnieszka Falenska, Jonas Kuhn

Institute for Natural Language Processing (IMS),  
University of Stuttgart

August 26, 2019

# Background

- Dependency Length Minimization hypothesis (Gibson, 1998)

# Background

- Dependency Length Minimization hypothesis (Gibson, 1998)
  - natural language tends to have shorter dependency length

# Background

- Dependency Length Minimization hypothesis (Gibson, 1998)
  - natural language tends to have shorter dependency length
  - in order to reduce cognitive burden

# Background

- Dependency Length Minimization hypothesis (Gibson, 1998)
  - natural language tends to have shorter dependency length
  - in order to reduce cognitive burden
- Empirical comparison to random order baseline supports the hypothesis (Futrell et al., 2015)

# Background

- Dependency Length Minimization hypothesis (Gibson, 1998)
  - natural language tends to have shorter dependency length
  - in order to reduce cognitive burden
- Empirical comparison to random order baseline supports the hypothesis (Futrell et al., 2015)
- In this work:

# Background

- Dependency Length Minimization hypothesis (Gibson, 1998)
  - natural language tends to have shorter dependency length
  - in order to reduce cognitive burden
- Empirical comparison to random order baseline supports the hypothesis (Futrell et al., 2015)
- In this work:
  - more incremental baselines to show the effect of DLM in detail

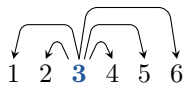
# Background

- Dependency Length Minimization hypothesis (Gibson, 1998)
  - natural language tends to have shorter dependency length
  - in order to reduce cognitive burden
- Empirical comparison to random order baseline supports the hypothesis (Futrell et al., 2015)
- In this work:
  - more incremental baselines to show the effect of DLM in detail
  - relation between DLM and word order constraints



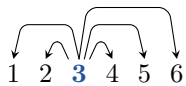
- 55 languages from Universal Dependencies 2.4
- Take the largest treebank to represent each language
- Keep sentences of maximum length of 50
- Use only projective trees
- Punctuation is ignored in calculation

# Baselines

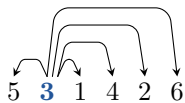


(a) observed

# Baselines

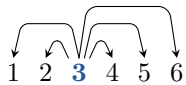


(a) observed

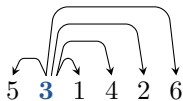


(b) free order

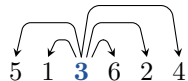
# Baselines



(a) observed

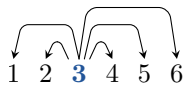


(b) free order

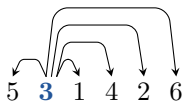


(c) same valency

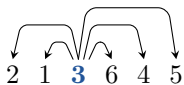
# Baselines



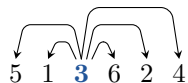
(a) observed



(b) free order

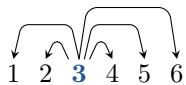


(d) same side

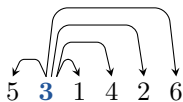


(c) same valency

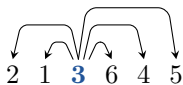
# Baselines



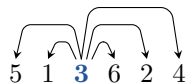
(a) observed



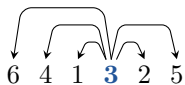
(b) free order



(d) same side



(c) same valency



(e) optimal

# Baseline Comparison

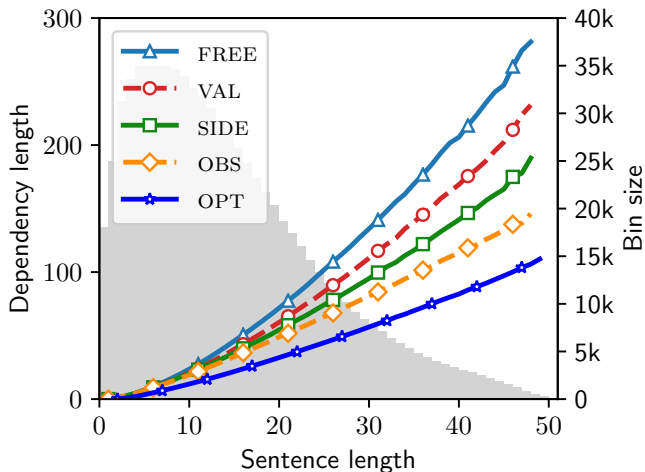
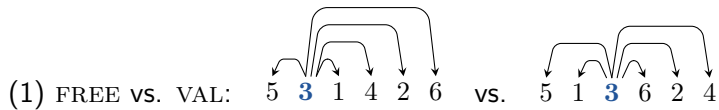


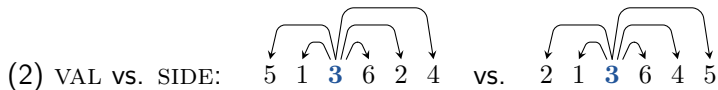
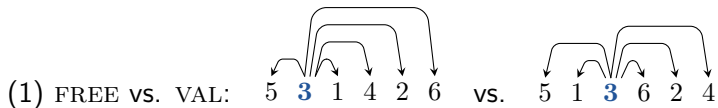
Figure: Average DL vs. sentence length across 55 treebanks.

# Analysis





# Analysis



# Analysis

(1) FREE vs. VAL: 5 **3** 1 4 2 6 vs. 5 1 **3** 6 2 4

(2) VAL vs. SIDE: 5 1 **3** 6 2 4 vs. 2 1 **3** 6 4 5

(3) SIDE vs. OBS: 2 1 **3** 6 4 5 vs. 1 2 **3** 4 5 6

## (1) Dependent Distribution: FREE vs. VAL

- Difference is the dependent distribution

## (1) Dependent Distribution: FREE vs. VAL

- Difference is the dependent distribution
- More balanced dependent distribution leads to shorter DL (Ferrer-i-Cancho, 2015)

## (1) Dependent Distribution: FREE vs. VAL

- Difference is the dependent distribution
- More balanced dependent distribution leads to shorter DL (Ferrer-i-Cancho, 2015)
- Define **imbalance** of a head as:

$$imbalance = \frac{|l - r|}{l + r}$$

where  $l$  and  $r$  are the numbers of left and right dependents

## (1) Dependent Distribution: FREE vs. VAL

- Difference is the dependent distribution
- More balanced dependent distribution leads to shorter DL (Ferrer-i-Cancho, 2015)
- Define **imbalance** of a head as:

$$imbalance = \frac{|l - r|}{l + r}$$

where  $l$  and  $r$  are the numbers of left and right dependents

- FREE has average imbalance of 0.67, VAL (same as OBS) has 0.47

# Word Order Constraints

- Two types of word order constraints:

# Word Order Constraints

- Two types of word order constraints:

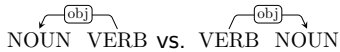
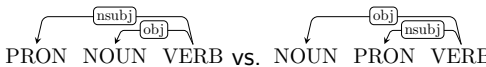
- dependent direction: NOUN VERB vs. VERB NOUN





# Word Order Constraints

- Two types of word order constraints:

- dependent direction: NOUN VERB vs. VERB NOUN  

- sibling ordering: PRON NOUN VERB vs. NOUN PRON VERB  


# Word Order Constraints

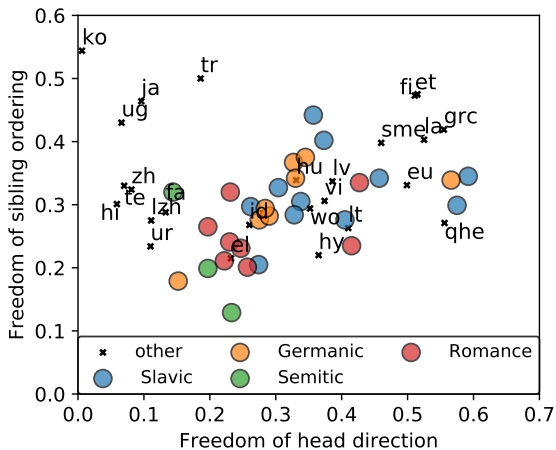
- Two types of word order constraints:

- dependent direction: NOUN VERB vs. VERB NOUN

- sibling ordering: PRON NOUN VERB vs. NOUN PRON VERB

- Freedom is defined as the entropy of the constraints in the corpus

# Word Order Freedom



## (2) Head Direction: VAL vs. SIDE

- Difference is the exact dependents on each side

## (2) Head Direction: VAL vs. SIDE

- Difference is the exact dependents on each side
- Flip one dependent in the real tree to the other side of the head, and observe the change of DL

## (2) Head Direction: VAL vs. SIDE

- Difference is the exact dependents on each side
- Flip one dependent in the real tree to the other side of the head, and observe the change of DL
- Consider two scenarios:

## (2) Head Direction: VAL vs. SIDE

- Difference is the exact dependents on each side
- Flip one dependent in the real tree to the other side of the head, and observe the change of DL
- Consider two scenarios:
  - majority case: flip the dependent that follows the dominant order

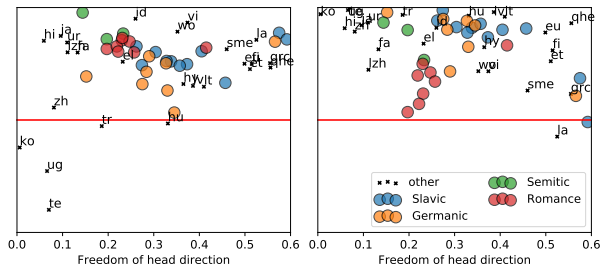
## (2) Head Direction: VAL vs. SIDE

- Difference is the exact dependents on each side
- Flip one dependent in the real tree to the other side of the head, and observe the change of DL
- Consider two scenarios:
  - majority case: flip the dependent that follows the dominant order
  - minority case: flip the dependent that is an exception of the order



## (2) Head Direction: VAL vs. SIDE

- Difference is the exact dependents on each side
- Flip one dependent in the real tree to the other side of the head, and observe the change of DL
- Consider two scenarios:
  - majority case: flip the dependent that follows the dominant order
  - minority case: flip the dependent that is an exception of the order



(a) Majority case

(b) Minority case

### (3) Sibling Ordering: SIDE vs. OBS

- Difference is the ordering of same-side siblings

### (3) Sibling Ordering: SIDE vs. OBS

- Difference is the ordering of same-side siblings
- Swap two dependents on the same side, and observe the change of DL

### (3) Sibling Ordering: SIDE vs. OBS

- Difference is the ordering of same-side siblings
- Swap two dependents on the same side, and observe the change of DL
- Consider two scenarios:

### (3) Sibling Ordering: SIDE vs. OBS

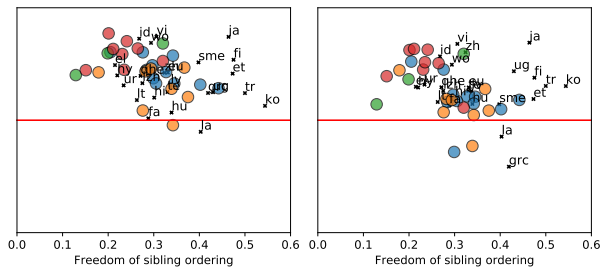
- Difference is the ordering of same-side siblings
- Swap two dependents on the same side, and observe the change of DL
- Consider two scenarios:
  - majority case: swap the dependents that follows the dominant order

### (3) Sibling Ordering: SIDE vs. OBS

- Difference is the ordering of same-side siblings
- Swap two dependents on the same side, and observe the change of DL
- Consider two scenarios:
  - majority case: swap the dependents that follows the dominant order
  - minority case: swap the dependents that is an exception of the order

### (3) Sibling Ordering: SIDE vs. OBS

- Difference is the ordering of same-side siblings
- Swap two dependents on the same side, and observe the change of DL
- Consider two scenarios:
  - majority case: swap the dependents that follows the dominant order
  - minority case: swap the dependents that is an exception of the order



(a) Majority case

(b) Minority case

# Summary

- Break down the effect of DLM in three aspects:



# Summary

- Break down the effect of DLM in three aspects:
  - dependent distribution
  - head direction
  - sibling ordering

# Summary

- Break down the effect of DLM in three aspects:
  - dependent distribution
  - head direction
  - sibling ordering
- The relation between DLM and word order:

# Summary

- Break down the effect of DLM in three aspects:
  - dependent distribution
  - head direction
  - sibling ordering
- The relation between DLM and word order:
  - dominant word order complies with the DLM preference

# Summary

- Break down the effect of DLM in three aspects:
  - dependent distribution
  - head direction
  - sibling ordering
- The relation between DLM and word order:
  - dominant word order complies with the DLM preference
  - exceptions in word order also has shorter DL

# Summary

- Break down the effect of DLM in three aspects:
  - dependent distribution
  - head direction
  - sibling ordering
- The relation between DLM and word order:
  - dominant word order complies with the DLM preference
  - exceptions in word order also has shorter DL
- Evidence of DLM shaping the word order constraints and motivating the exceptions when necessary

## Future Work

- Explore the idiosyncrasy of annotation guidelines

# Future Work

- Explore the idiosyncrasy of annotation guidelines
- Analyze OBS vs. OPT:  
how word order constraints conflict with DLM

# Future Work

- Explore the idiosyncrasy of annotation guidelines
- Analyze OBS vs. OPT:  
how word order constraints conflict with DLM
- Use a statistical linearizer:  
more nuanced than extracted patterns



# References I

- Ferrer-i-Cancho. 2015. The placement of the head that minimizes online memory: a complex systems approach. *Language Dynamics and Change*, 5(1):114–137.
- 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.
- Edward Gibson. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1):1–76.