Dependency Distance VS Frequency

Xinying Chen, University of Ostrava

Kim Gerdes, Sorbonne Nouvelle, LPP(CNRS), Almanach (Inria)

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education
CZ.02.2.69/0.0/0.0/16_027/0008472

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

Outline

- Motivation

Outline

- Motivation
- Hypothesis

Outline

- Motivation
- Hypothesis
- Materials and methods

Outline

- Motivation
- Hypothesis
- Materials and methods
- Results and discussion

Motivation

- The neglected individual dependency distance

Motivation

- The neglected individual dependency distance
  - The popularity of Mean Dependency Distance

Motivation

- The neglected individual dependency distance
  - The popularity of Mean Dependency Distance

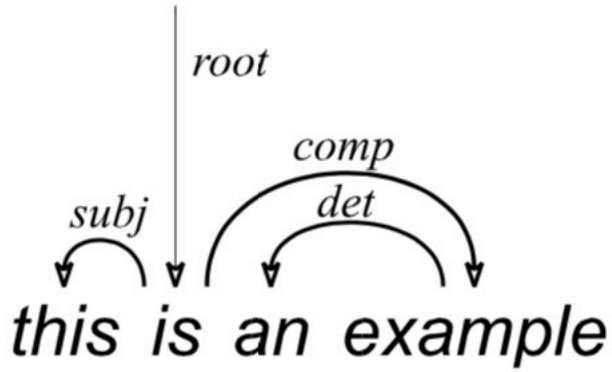


Figure 1: Example dependency tree in SUD analysis.

Motivation

- The neglected individual dependency distance
  - The popularity of Mean Dependency Distance

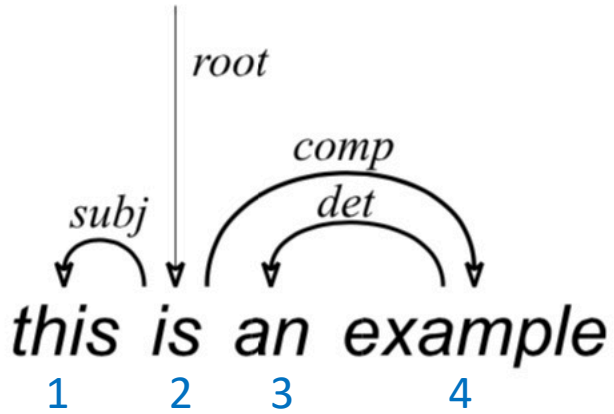


Figure 1: Example dependency tree in SUD analysis.

Motivation

- The neglected individual dependency distance
  - The popularity of Mean Dependency Distance

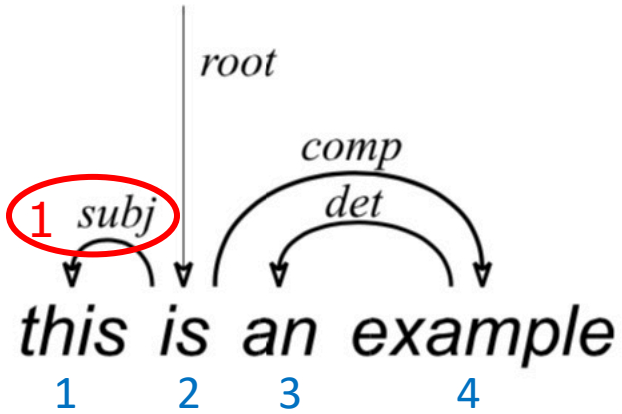


Figure 1: Example dependency tree in SUD analysis.

Motivation

- The neglected individual dependency distance
  - The popularity of Mean Dependency Distance

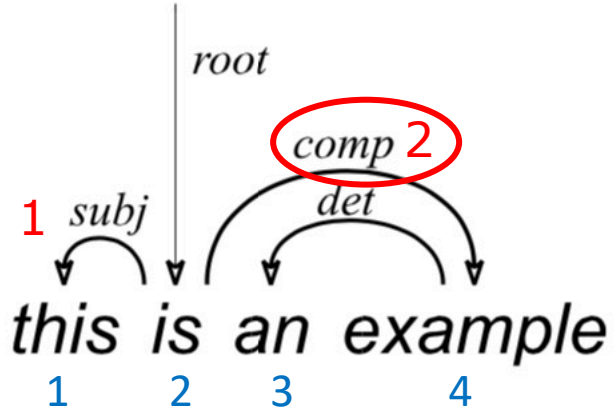


Figure 1: Example dependency tree in SUD analysis.

Motivation

- The neglected individual dependency distance
  - The popularity of Mean Dependency Distance

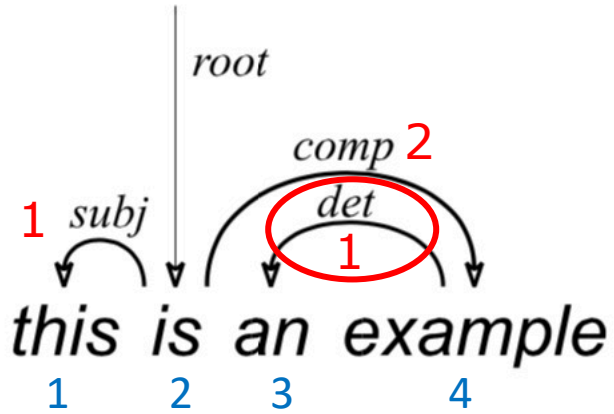


Figure 1: Example dependency tree in SUD analysis.

Motivation

- The neglected individual dependency distance
  - The popularity of Mean Dependency Distance

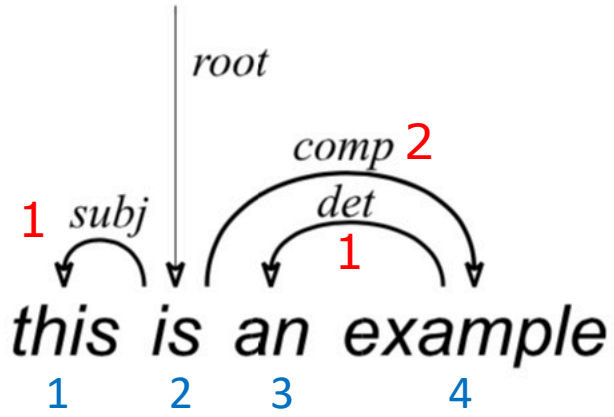


Figure 1: Example dependency tree in SUD analysis.

Mean dependency distance
(1+2+1) / 3 = 1.33

Motivation

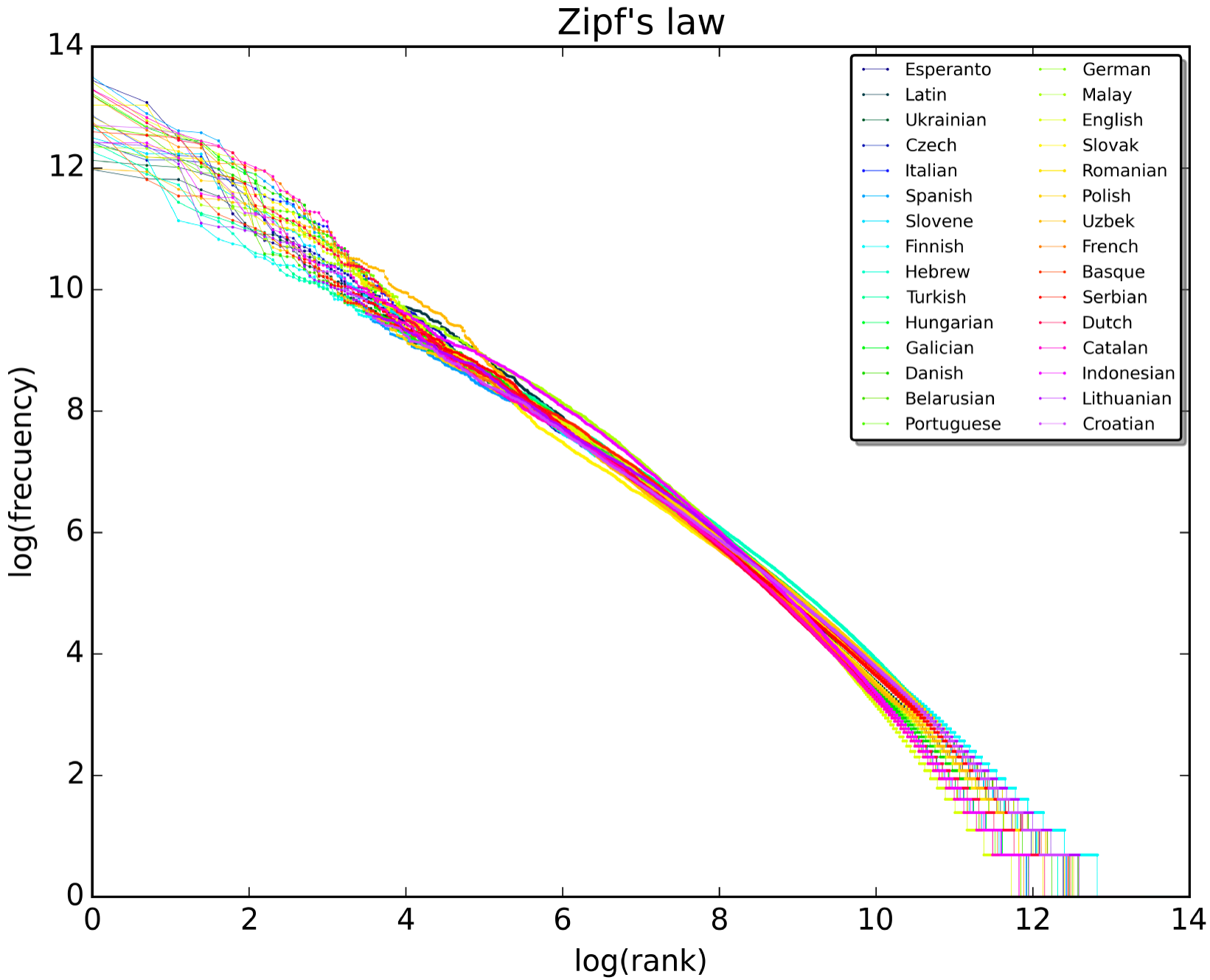
- The neglected individual dependency distance
  - The popularity of Mean Dependency Distance
  - Individual dependency distances provide more details of the fluctuation than the average
  - Maybe not sentences, but language use

Motivation

- The neglected individual dependency distance
- Power Law everywhere in language

Motivation

- The neglected individual dependency distance
- Power Law everywhere in language
  - Zipf’s Law (distribution law)



A plot of the rank versus frequency for the first 10 million words in 30 Wikipedias (dumps from October 2015) in a log-log scale.
https://en.wikipedia.org/wiki/Zipf%27s_law

Motivation

- The neglected individual dependency distance
- Power Law everywhere in language
  - Zipf’s Law (distribution law)
  - Synergetic framework (relations between two features of one linguistic unit)

Motivation

- The neglected individual dependency distance
- Power Law everywhere in language
  - Zipf’s Law (distribution law)
  - Synergetic framework (relations between two features of one linguistic unit)

    Eg. word length VS word frequency VS numbers of meaning

Motivation

- The neglected individual dependency distance
- Power Law everywhere in language (lexical level)

Motivation

- The neglected individual dependency distance
- Power Law everywhere in language (lexical level)
- Why not syntax level?

Motivation

- The neglected individual dependency distance
- Power Law everywhere in language (lexical level)
- Why not syntax level?
  - Empirical data or treebanks ? (language, size, accuracy, annotation schema)

Motivation

- The neglected individual dependency distance
- Power Law everywhere in language (lexical level)
- Why not syntax level?
  - Empirical data or treebanks ? (language, size, accuracy, annotation schema)

Motivation

- The neglected individual dependency distance
- Power Law everywhere in language (lexical level)
- Why not syntax level?
  - Empirical data or treebanks ? (language, size, accuracy, annotation schema)
  - New reality: Data is there !!!

Hypothesis

- *The relation between dependency distance and frequency can be formulated as a non-linear function (probably a power law function).*

Materials and methods

- PUD English Treebank from Surface-syntactic Universal Dependencies (SUD)

Materials and methods

- PUD English Treebank from Surface-syntactic Universal Dependencies (SUD)
  - Why PUD ?

Materials and methods

- PUD English Treebank from Surface-syntactic Universal Dependencies (SUD)
  - Why PUD ?
    - a **parallel treebank** with a wide range of languages, namely Arabic, Chinese, Czech, English, Finnish, French, German, Hindi, Indonesian, Italian, Japanese, Korean, Portuguese, Russian, Spanish, Swedish, Thai, and Turkish.

Materials and methods

- PUD English Treebank from Surface-syntactic Universal Dependencies (SUD)
  - Why PUD ?
  - Why SUD ?

Materials and methods

- PUD English Treebank from Surface-syntactic Universal Dependencies (SUD)
  - Why PUD ?
  - Why SUD ?
    - equivalent to UD
    - closer to traditional dependency grammar
    - Simpler

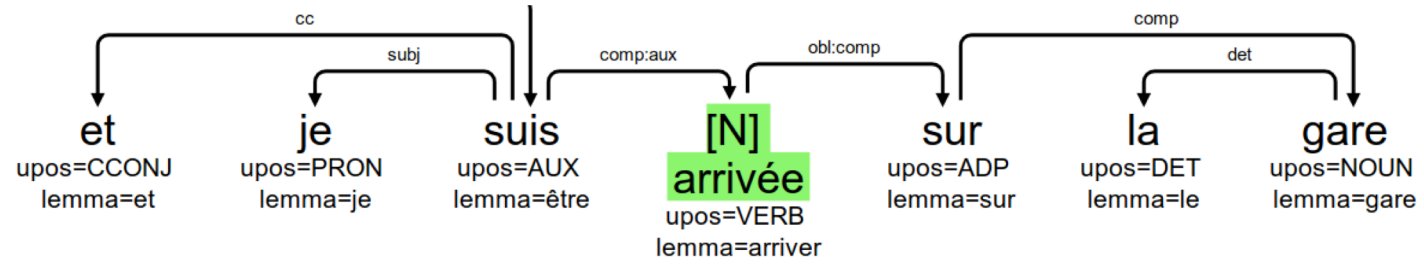
Gerdes, Kim, Bruno Guillaume, Sylvain Kahane, and Guy Perrier. "SUD or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD." In: Proceedings of the Universal Dependencies Workshop 2018.

Materials and methods

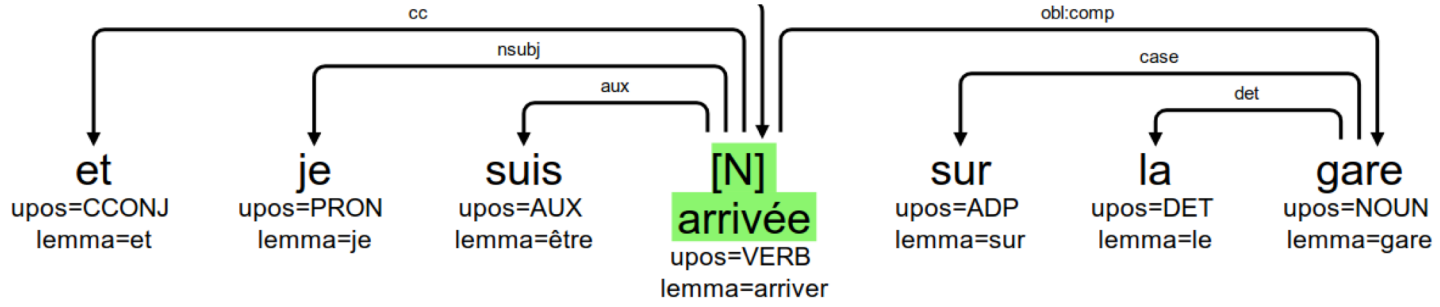
Dependency links and labels are based exclusively on **syntactic distributional criteria** → **functional heads** (auxiliairies, markers, adpositions)

Meaning-Text Theory, Mel'čuk 1988; Word Grammar, Hudson 1984, 2007; Prague Dependency Treebank, Hajič et al. 2017

SUD :



UD :



Materials and methods

- PUD English Treebank from Surface-syntactic Universal Dependencies (SUD)
- Computations

Materials and methods

- PUD English Treebank from Surface-syntactic Universal Dependencies (SUD)
- Computations
  - Individual dependency distance
  - frequencies of each dependency distance
  - non-linear formulations

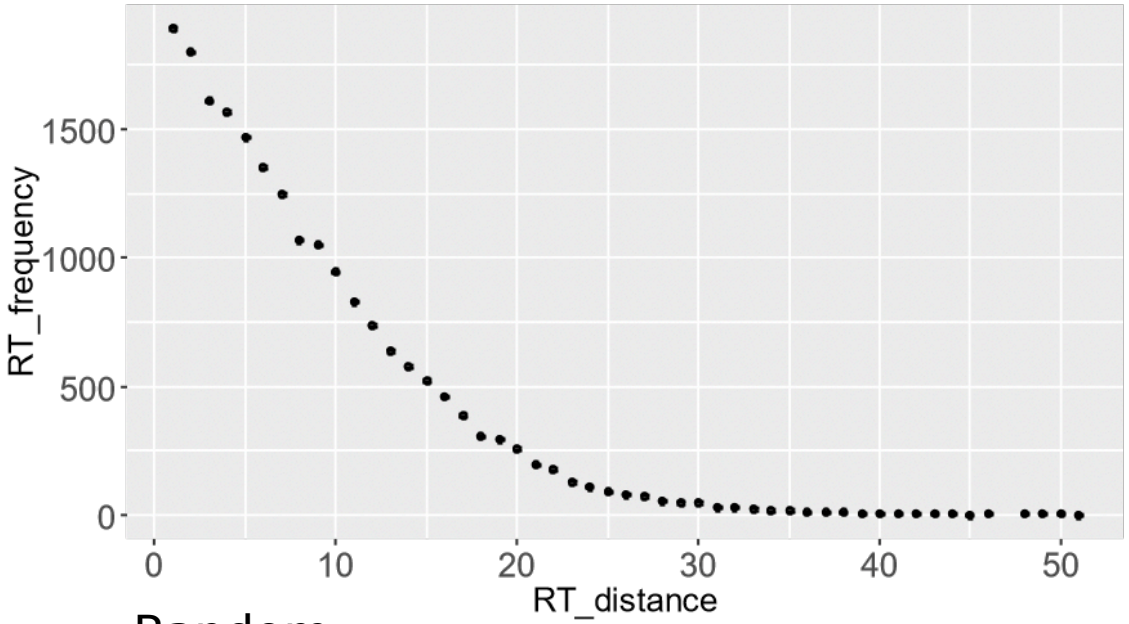
Materials and methods

- PUD English Treebank from Surface-syntactic Universal Dependencies (SUD)
- Computations
- Experiments

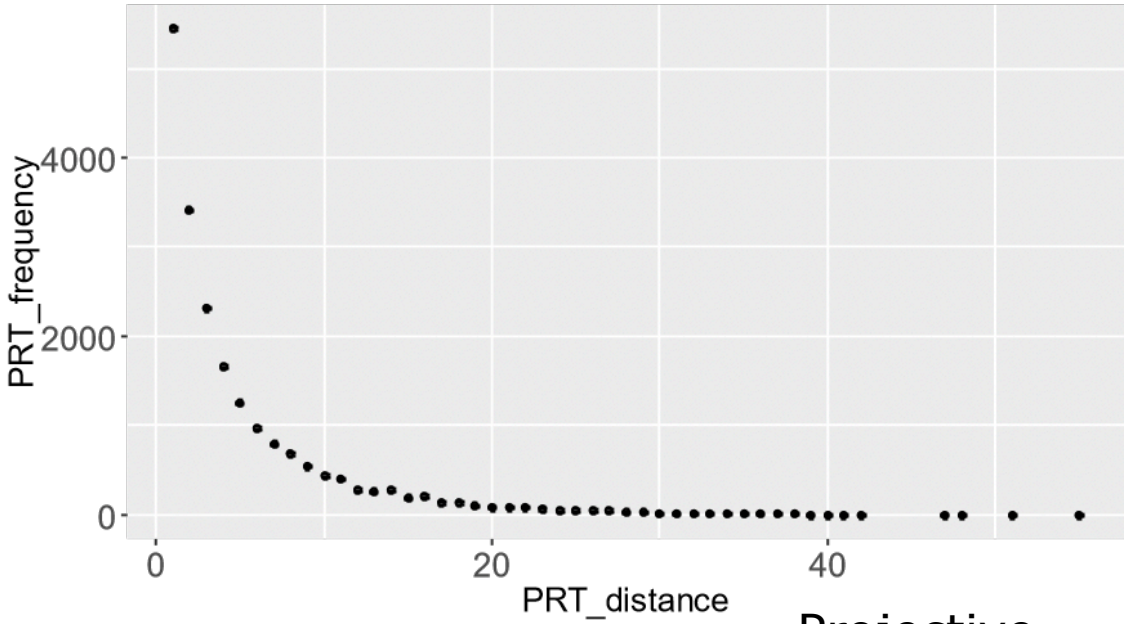
Materials and methods

- PUD English Treebank from Surface-syntactic Universal Dependencies (SUD)
- Computations
- Experiments
  - random baseline (random trees)
  - control group (projective random trees)
  - repeat the test for syntactic dependencies (the role of syntax)
    - subj, aux, cop, case, mark, cc, dislocated, vocative, expl, discourse, det, clf

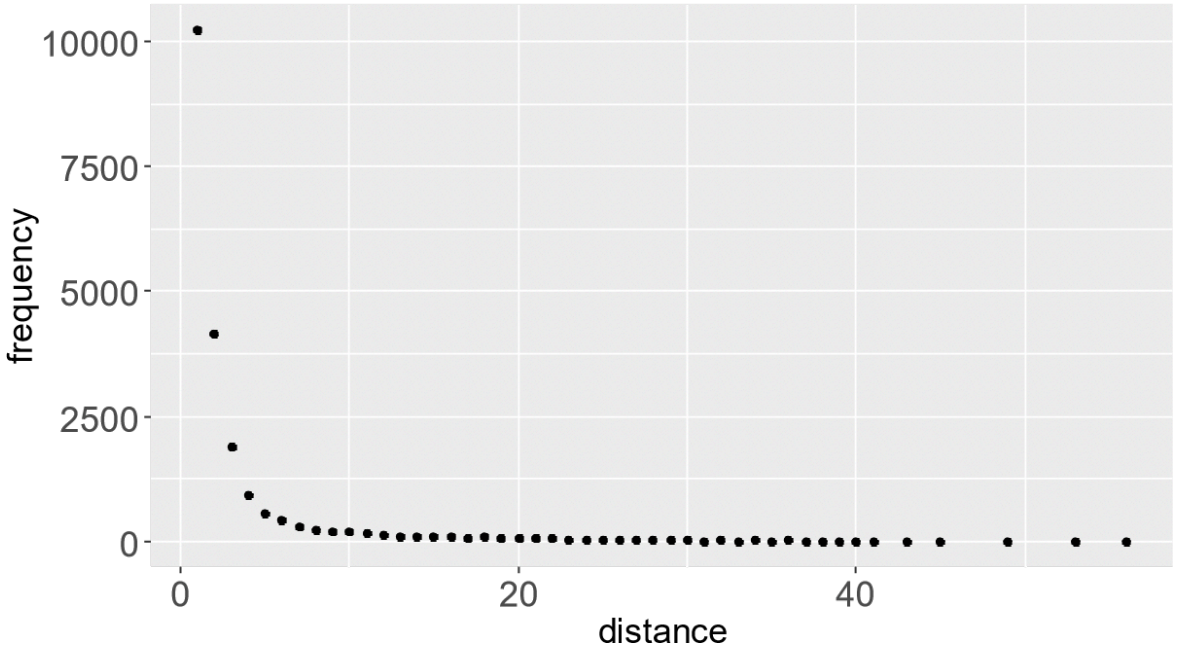
Results and discussion



Random



Projective



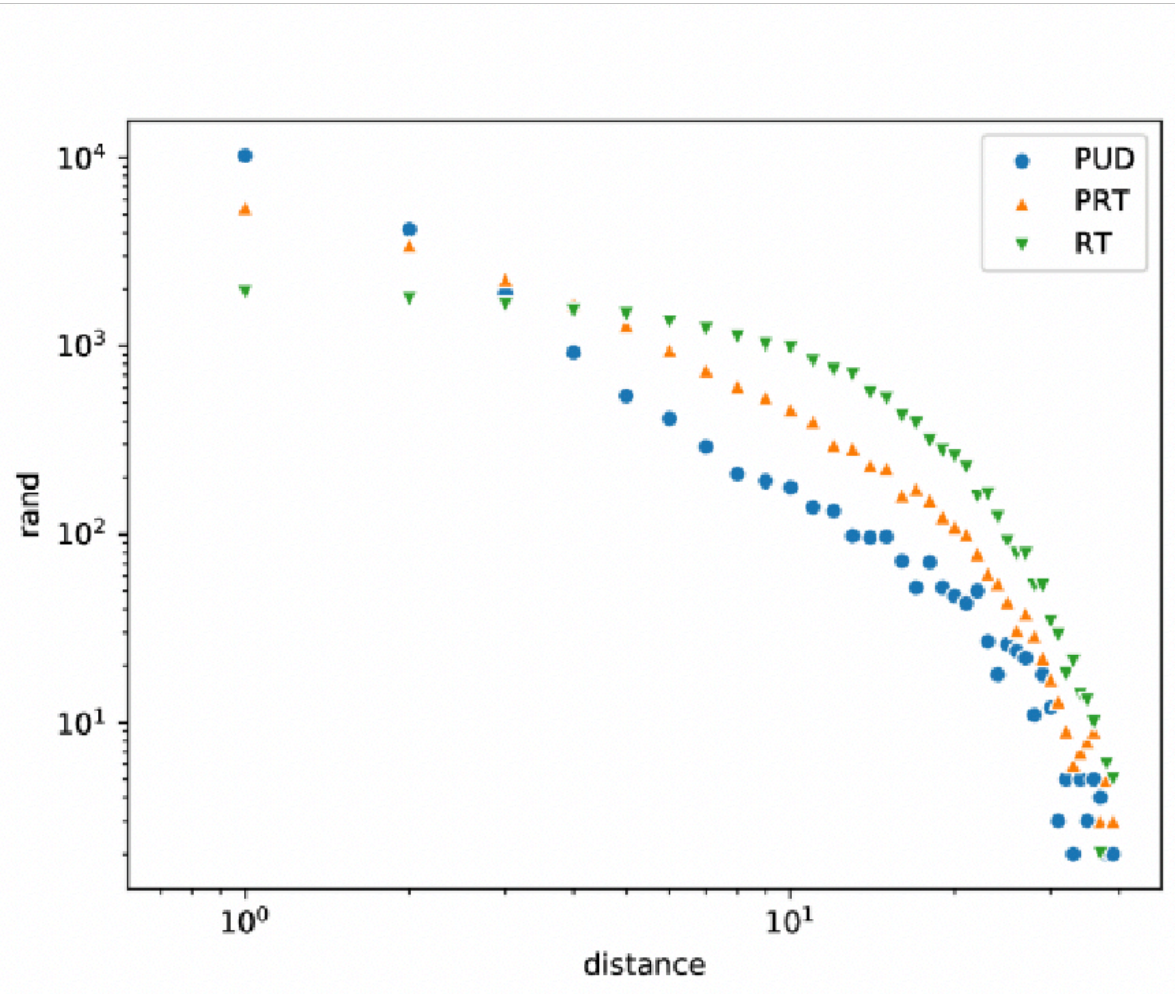
Nature

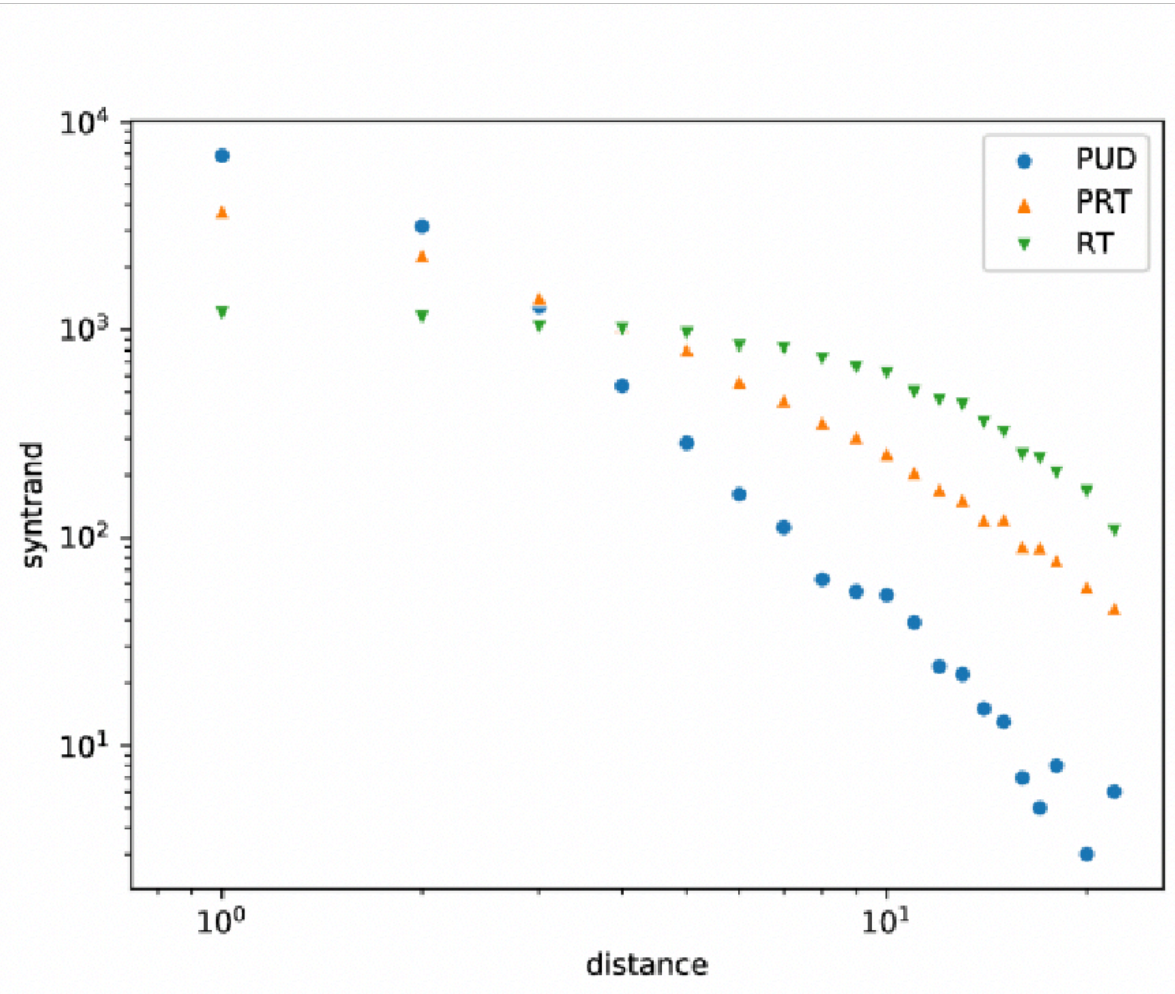
Treebank	Non-linear Model	Function	R^2
PUD English	Quadratic	$y=2963.44-206x+3.1x^2$	0.34
	Exponent	$log(y)=7.11-0.16x$	0.92
	Logarithm	$y=4100.8-1262log(x)$	0.49
	Power Law	$log(y)=10.71-2.56log(x)$	0.91
Random Trees	Quadratic	$y=1883.88-106.28x+1.43x^2$	0.98
	Exponent	$log(y)=8.42-0.17x$	0.98
	Logarithm	$y=2220.88-611.66log(x)$	0.96
	Power Law	$log(y)=11.23-2.37log(x)$	0.74
Projective Random Trees	Quadratic	$y=2551.07-168.63x+2.49x^2$	0.62
	Exponent	$log(y)=7.99-0.17x$	0.97
	Logarithm	$y=3258.25-972.05log(x)$	0.75
	Power Law	$log(y)=11.28-2.55log(x)$	0.84

Materials and methods

- PUD English Treebank from Surface-syntactic Universal Dependencies (SUD)
- Computations
- Experiments
  - random baseline (random trees)
  - control group (projective random trees)
  - **repeat the test for syntactic dependencies (the role of syntax)**
    - **subj, aux, cop, case, mark, cc, dislocated, vocative, expl, discourse, det, clf**

Syntactic Data Set	Non-linear Model	Function	R^2
PUD English	Quadratic	$y=2963.44-206x+3.1x^2$	0.44
	Exponent	$log(y)=7.11-0.16x$	0.81
	Logarithm	$y=4100.8-1262log(x)$	0.56
	Power Law	$log(y)=10.71-2.56log(x)$	0.97
Random Trees	Quadratic	$y=1883.88-106.28x+1.43x^2$	0.98
	Exponent	$log(y)=8.42-0.17x$	0.97
	Logarithm	$y=2220.88-611.66log(x)$	0.95
	Power Law	$log(y)=11.23-2.37log(x)$	0.74
Projective Random Trees	Quadratic	$y=2551.07-168.63x+2.49x^2$	0.6
	Exponent	$log(y)=7.99-0.17x$	0.95
	Logarithm	$y=3258.25-972.05log(x)$	0.73
	Power Law	$log(y)=11.28-2.55log(x)$	0.89


PUD
PRT
RT
rand
distance
10^4
10^3
10^2
10^1
10^0
10^1


PUD
PRT
RT
syntrand
distance
10^4
10^3
10^2
10^1
10^0
10^1

Results and discussion

- Hypothesis ✓

Results and discussion

- Hypothesis ✓
- Power law model is probably a better choice for representing the relation between dependency distance and frequency.

Results and discussion

- Hypothesis ✓
- Power law model is probably a better choice for representing the relation between dependency distance and frequency.
- Projectivity has a major role as the responsible factor for the power-law function of dependency distance.

Thank You!