The origins of the scarcity of crossing dependencies in languages

C. Gómez-Rodríguez[1], R. Ferrer-i-Cancho[2] and J. L. Esteban[3]

[1] LyS Research Group
Departamento de Computación
Universidade da Coruña

[3] Complexity & Quantitatitve Linguistics Lab
LARCA Research Group
Departament de Ciències de la Computació
Universitat Politècnica de Catalunya

[3] LOGPROG Research Group
Departament de Ciències de la Computació
Universitat Politècnica de Catalunya

COMPLEXITAT

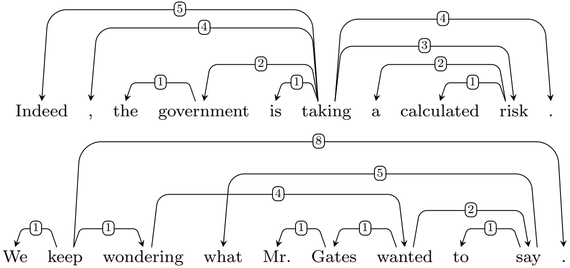
UPC

Qualico 2018 (Wroclaw), 6 July 2018

A big challenge for theoretical linguistics: How to explain language patterns as well as linguistic variation

- Generative enterprise: universal faculty of language + linguistic variation.
- Principles and Parameters framework [Chomsky, 1981].
- Invariant innate principles + open parameters (specified through the language specific input that children receive).
- Famous switchboard metaphor (Higginbotham).
- Zipf's principle of least effort [Zipf, 1949]. A humble professor of German.

A principle: Dependencies tend to not cross when drawn above the sent.



Discovered by Hays and Lecerf in the 1960s and confirmed on about 30 languages:

- $p(C = 0)$ is very high [Ferrer-i-Cancho et al., 2018].
- The average C does not exceed 1 for most of the languages [Gómez-Rodríguez and Ferrer-i-Cancho, 2017].
- C is actually small based on different baselines [Ferrer-i-Cancho et al., 2018].

Fundamental issue for the actual complexity of human languages: **projective (=easy to parse) grammar implies no crossings.**

A parameter: Headedness.

- Whether a head should follow or precede its dependents.
- Principles and parameters theory [Baker, 2001]
- Branching direction:
  - Left-branching (head final): Japanese, Turkish.
  - Right-branching (head initial): English (more right-branching than left-branching)

The power of simplicity and the risk of overfitting

Barnsley fern

Real fern

The truth is that ...

- Headedness does not qualify as a parameter [Ferrer-i-Cancho, 2015b, Ferrer-i-Cancho and Gómez-Rodríguez, 2016].
- The scarcity of crossings neither as a principle [Ferrer-i-Cancho, 2006, Gómez-Rodríguez and Ferrer-i-Cancho, 2017].

Key: **principle of dependency distance minimization [Liu et al., 2017]**

Stay here for the latter. Read refs above for the former.

Why? Two major hypotheses

The scarcity of crossings dependencies arises,

- **Directly** from an underlying rule or principle of human languages that is responsible for this fact (including the possibility of some cognitive cost associated directly to crossings). Held by the **overwhelming majority of researchers**. Serious problems [Ferrer-i-Cancho and Gómez-Rodríguez, 2016]:
  - It requires heavy assumptions that compromise the parsimony of linguistic theory as a whole.
  - It involves explanations based on internal constraints of obscure nature.
- **Indirectly**, from the actual length of dependencies, which are constrained by a well-known psychological principle: dependency length minimization [Gómez-Rodríguez and Ferrer-i-Cancho, 2017].

Crossing theory

- $C \leq \binom{n-1}{2}$ but actually $C \leq \binom{n-2}{2}$ [Ferrer-i-Cancho et al., 2018]
- $C \leq |Q|$, where Q is the **set of pairs of edges of a graph that can potentially cross** when their vertices are arranged linearly in some arbitrary order. **Edges sharing a vertex cannot cross!**
- $|Q|$, the cardinality of Q, is the **potential number of crossings**.
- In a tree, one has

$$|Q| = n(n - 1 - \langle k^2 \rangle)/2, \tag{1}$$

  $|Q| = 0$ if and only if the tree is a star tree [Ferrer-i-Cancho, 2017].

1st predictor of C

- The number of edge crossings

$$C = \sum_{(e_1,e_2)\in Q} C(e_1, e_2), \tag{2}$$

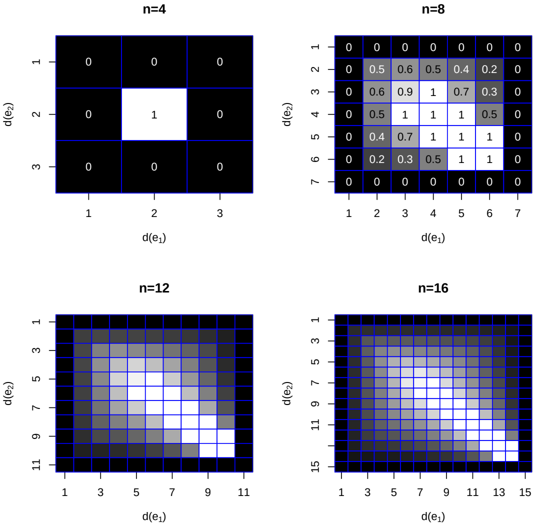
- $C(e_1, e_2)$ is an indicator variable ($C(e_1, e_2) = 1$ if the edges e_1 and e_2 cross and $C(e_1, e_2) = 0$ otherwise).

Null hypothesis that the vertices are arranged linearly at random (all possible orderings are equally likely). $p(C(e_1, e_2) = 1) = 1/3$ yields $E_0[C] = |Q|/3$ [Ferrer-i-Cancho, 2017].

2nd predictor of C

- Introducing knowledge about the length of the dependencies (edges of length 1 or $n-1$ are not crossable). $p(C(e_1,e_2)=1)$ is replaced by $p(C(e_1,e_2)=1|d(e_1),d(e_2))$, obtaining [Ferrer-i-Cancho, 2014]
- $$E_2[C] = \sum_{(e_1,e_2)\in Q} p(C(e_1,e_2)=1|d(e_1),d(e_2)), \qquad (3)$$
- $p(C(e_1,e_2)=1|d(e_1),d(e_2))$ depends only on n, $d(e_1)$ and $d(e_2)$ [Ferrer-i-Cancho, 2014].
- $E_0[C]$ is a true expectation while $E_2[C]$ is not!

$p(C(e_1, e_2) = 1 | d(e_1), d(e_2))$



Evaluation of the predictors: materials

- Corpora in version 2.0 of the HamleDT collection of treebanks [Zeman et al., 2014, Rosa et al., 2014]: a harmonization of existing treebanks for **30 different languages** into two well-known annotation styles: Prague dependencies [Hajič et al., 2006] and Universal Stanford dependencies [de Marneffe et al., 2014].
- Preprocessing: nodes corresponding to punctuation tokens or null elements were removed (non-punctuation nodes that had a punctuation node as their head were attached as dependents of their nearest non-punctuation ancestor). Same treatment for null elements.
- A syntactic dependency structure was included in our analyses if (1) it defined a tree and (2) the tree was not a star tree.

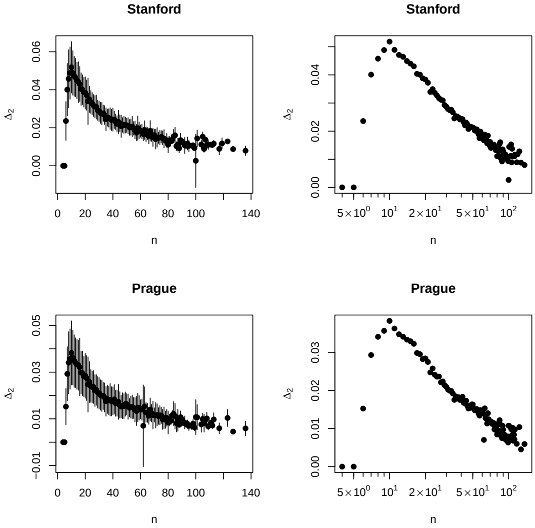
Evaluation of the predictors: methods

- C, number of crossings of the linear arrangement of a graph in general.
- C_{true}, the number of crossings of the syntactic dependencies of a real sentence.
- **relative number of crossings**, i.e. $\bar{C} = C/|Q|$ or $\bar{C}_{true} = C_{true}/|Q|$ [Ferrer-i-Cancho, 2014].
- **relative error** of a predictor [Ferrer-i-Cancho, 2014]

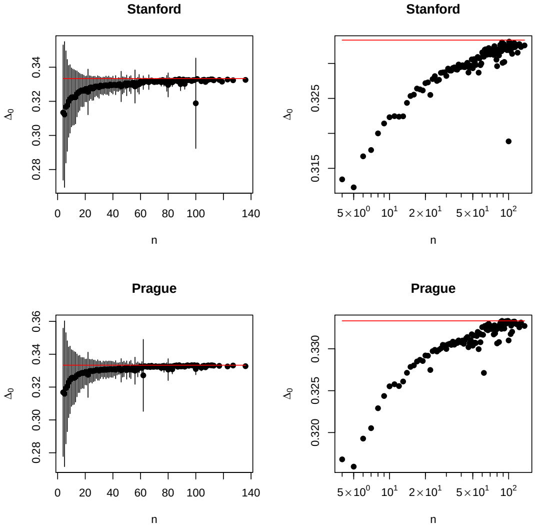
$$\Delta_x = E_x\left[\bar{C}\right] - \bar{C}_{true} = (E_x[C] - C_{true})/|Q|. \tag{4}$$

Δ_0 will be used as a baseline for Δ_2.
Δ_0 converges to $1/3$ for sufficiently long sentences when C_{true} is small [Ferrer-i-Cancho, 2014].

Results I


Stanford
Stanford
Prague
Prague
Δ_2
n
0 20 40 60 80 100 140
0.00 0.02 0.04 0.06
0.00 0.02 0.04
5×10^0 10^1 2×10^1 5×10^1 10^2
-0.01 0.01 0.03 0.05
0.00 0.01 0.02 0.03

Results II


Stanford
Stanford
Prague
Prague
Δ_0
n
0.28
0.30
0.32
0.34
0.36
0.315
0.320
0.325
0.330
0
20
40
60
80
100
140
5×10^0
10^1
2×10^1
5×10^1
10^2

Results III

Mixing sentences of different lengths:

- The average Δ_2, the relative error of the predictor $E_2[C]$, is small: it does not exceed 5%.
- The average Δ_2 is at least 6 times smaller than the baseline $\Delta_0 \approx 30\%$.

Controlling for sentence length (grouping sentences by length):

- Average over group averages of Δ_2: it does not exceed 4.3%.
- The average Δ_2 is at least 7 times smaller than the baseline error, again $\Delta_0 \approx 30\%$.
- The minimum size of a group is one sentence; the qualitative results are very similar if the minimum size is set to 2.

Concluding remarks I

principle of dependency length minimization [Ferrer-i-Cancho, 2015a]

$\downarrow$

dependency lengths

$\downarrow$

the actual **number of crossings** in sentences

To explain the low frequency of crossings in world languages, it many not be necessary to recur to recur to

- A ban of crossings by grammar (e.g., [Hudson, 2007, Tanaka, 1997]).
- A principle of minimization of crossings [Liu, 2008]
- A competence-plus [Hurford, 2012] limiting the number of crossings.

Concluding remmarks II

- The computational complexity of languages: **mild context sensitivity** [Joshi, 1985].
- Realistic constraints on dependency lengths could help to define, maybe probabilistically, the **mild non-projectivity** of real sentences.
- A constraint on crossings does not qualify as a principle (headedness neither qualifies as a parameter).
- We are **liberating theoretical linguistics from the naive (atheoretic) induction behind dogmas of the 20th century** [Ferrer-i-Cancho and Gómez-Rodríguez, 2016].
- Challeging the reality of constituent structure [Chen et al., 2018]. Continuity constraint for free!

Baker, M. C. (2001).
The atoms of language.
Basic books, New York.

Chen, X., C. Gómez-Rodríguez, C., and i Cancho, R. F. (2018).
A dependency look at the reality of constituency.
Glottometrics, 40:104–106.

Chomsky, N. (1981).
Lectures on Government and Binding.
Foris, Dordretch.

de Marneffe, M.-C., Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J., and Manning, C. D. (2014).
Universal Stanford dependencies: a cross-linguistic typology.
In Chair), N. C. C., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. European Language Resources Association (ELRA).

Ferrer-i-Cancho, R. (2006).
Why do syntactic links not cross?
Europhysics Letters, 76(6):1228–1235.

Ferrer-i-Cancho, R. (2014).
A stronger null hypothesis for crossing dependencies.
Europhysics Letters, 108:58003.

Ferrer-i-Cancho, R. (2015a).
The placement of the head that minimizes online memory. A complex systems approach.
Language Dynamics and Change, 5:114–137.

Ferrer-i-Cancho, R. (2015b).
Reply to the commentary "Be careful when assuming the obvious", by P. Alday.
Language Dynamics and Change, 5:147–155.

Ferrer-i-Cancho, R. (2017).
Random crossings in dependency trees.
Glottometrics, 37:1–12.

Ferrer-i-Cancho, R. and Gómez-Rodríguez, C. (2016).
Liberating language research from dogmas of the 20th century.
Glottometrics, 33:33–34.

Ferrer-i-Cancho, R., Gómez-Rodríguez, C., and Esteban, J. L. (2018).
Are crossing dependencies really scarce?
Physica A: Statistical Mechanics and its Applications, 493:311–329.

Gómez-Rodríguez, C. and Ferrer-i-Cancho, R. (2017).
Scarcity of crossing dependencies: a direct outcome of a specific constraint?
Physical Review E, 96:062304.

Hajič, J., Panevová, J., Hajičová, E., Sgall, P., Pajas, P., Štěpánek, J., Havelka, J., Mikulová, M., Žabokrtský, Z., Ševčíková-Razímová, M., and Urešová, Z. (2006).
Prague dependency treebank 2.0.
CDROM CAT: LDC2006T01, ISBN 1-58563-370-4. Linguistic Data Consortium.

Hudson, R. A. (2007).
Language Networks: The New Word Grammar.

Oxford University Press.

Hurford, J. R. (2012).
Syntax in the light of evolution, chapter 3, pages 175–258.
Oxford University Press, Oxford.

Joshi, A. (1985).
How much context-sensisitivity is necessary for characterizing structural descriptions – tree adjoining grammars.
In Dowty, D., Karttunen, L., and Zwicky, A., editors, *Natural Language Processing: Psycholinguistic, Computational and Theoretical Perspectives*, pages 206–250. Cambridge University Press.

Liu, H. (2008).
Dependency distance as a metric of language comprehension difficulty.
Journal of Cognitive Science, 9:159–191.

Liu, H., Xu, C., and Liang, J. (2017).
Dependency distance: a new perspective on syntactic patterns in natural languages.
Physics of Life Reviews, 21:171–193.

Rosa, R., Mašek, J., Mareček, D., Popel, M., Zeman, D., and Žabokrtský, Z. (2014).
HamleDT 2.0: Thirty dependency treebanks stanfordized.
In Chair), N. C. C., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. European Language Resources Association (ELRA).

Tanaka, H. (1997).
Invisible movement in sika-nai and the linear crossing constraint.
Journal of East Asian Linguistics, 6:143–188.

Zeman, D., Dušek, O., Mareček, D., Popel, M., Ramasamy, L., Štěpánek, J., Žabokrtský, Z., and Hajič, J. (2014).
HamleDT: Harmonized multi-language dependency treebank.
Language Resources and Evaluation, 48(4):601–637.

Zipf, G. K. (1949).
Human behaviour and the principle of least effort.
Addison-Wesley, Cambridge (MA), USA.