Fiche de projet tutoré / Project form

Testing (some of) Greenberg's Linguistic Universals on the Universal Dependencies corpora using a graph rewriting tool

---

Encadrement / Supervisors

Team : Sémagramme team, LORIA

Supervisors :

- Bruno Guillaume (Bruno.Guillaume@loria.fr)
- Karën Fort (Karen.Fort@loria.fr)

Description / Description

In 1963, J. H. Greenberg proposed 45 linguistic universals. For instance:

> [Universal 2] *In languages with prepositions, the genitive almost always follows the governing noun, while in languages with postpositions it always precedes*.

Since then, some of these universals (especially word order universals) have been tested on large corpora. With the Universal Dependencies (UD) project in 2014, many corpora are available in a common annotation framework. On these corpora, it is now possible to make cross-linguistic experiments on a large number of languages: UD includes treebanks in 85 languages. For instance, Gerdes *et al*. recently tested Universals 19 and 25 on the UD corpora.

We propose to extend this work using Grew[1], a graph rewriting tool developed at Loria.
With Grew, it is possible to make complex requests on annotated corpora and hence to have statistics about the frequencies of syntactic constructions.

During the first period, the students will have to read and understand the state of the art on the subject: Greenberg's paper and previous study of its universals on corpora. In parallel, then they will learn how to use Grew.

For the second part, they will study which universals are testable with Grew on UD, select some of them and perform a cross-linguistic study in order to (in)validate the 1963 universals.

---

[1] http://match.grew.fr

**Informations diverses : matériel nécessaire, contexte de réalisation /
Various information: material, context of realization**

An online tool is available for querying graphs with Grew, so no special material will be needed to start.

Livrables et échéancier / Deliverable and schedule

December 20: Report on the bibliography given at the beginning of the project and a list of selected universal to be tested on UD corpora.

Mai 26: Final report with cross-linguistic studies on universal selected

Bibliographie /References

Kim Gerdes, Sylvain Kahane and Xinying Chen. (2019) *Rediscovering Greenberg's Word Order Universals in UD*, Proceedings of the Universal Dependencies Workshop (UDW), SyntaxFest, Paris.

Joseph H. Greenberg. (1963) *Some universals of grammar with particular reference to the order of meaningful elements*, in Universals of Language, pp. 73-113.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty and Daniel Zeman (2016), *Universal Dependencies v1: A Multilingual Treebank Collection* Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC)

Kartik Sharma, Kaivalya Swami, Aditya Shete and Samar Husain (2019). *Can Greenbergian universals be induced from language networks?*, Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories (TLT), SyntaxFest, Paris.