

Invited Talk

Monday 26th August 2019

Dependency distance minimization: facts, theory and predictions

Ramon Ferrer-i-Cancho

Universitat Politècnica de Catalunya

Abstract

Quantitative linguistics is a branch of linguistics concerned about the study of statistical facts about languages and their explanation aiming at constructing a general theory of language. The quantitative study of syntax has become central to this branch of linguistics. The fact that the distance between syntactically related words is smaller than expected by chance in many languages led to the formulation of a dependency distance minimization (DDm) principle.

From a theoretical standpoint, DDm is in conflict with another word order principle: surprisal minimization (Sm). In single head structures, DDm predicts that the head should be put at the center of the linear arrangement, while Sm predicts that it should be put at one of the ends. In spite of the massive evidence of the action of DDm and the trendy claim that languages are optimized, attempts to quantify the degree of optimization of languages according to DDm have been rather scarce. Here we present a new optimality measure indicating that languages are optimized to a 70%. We confirm two old theoretical predictions: that the action of DDm is stronger in longer sentences and that DDm is more likely to be beaten by Sm in short sequences (resulting in an anti-DDm effect), while shedding new light on the kind of tree structures where DDm is more likely to be shadowed. Finally, we review various theoretical predictions of DDm focusing on the scarcity of crossing dependencies. We challenge the belief that formal constraints on dependency trees (e.g., projectivity or relaxed versions) are real rather than epiphenomenal.

The talk is a summary of joint work with Carlos Gomez-Rodriguez, Juan Luis Esteban, Morten Christiansen, Lluís Alemany-Puig and Xinying Chen.

Short bio

Ramon Ferrer-i-Cancho is associate professor at Universitat Politècnica de Catalunya and the head of the Complexity and Quantitative Linguistics Lab. He is a language researcher in a broad sense. His research covers different levels of the organization of life: from human language to animal behavior and down farther to the molecular level. One of his main research objectives is the development of a parsimonious but general theory of language and communication integrating insights from probability theory, information theory and the theory of spatial networks. In the context of syntax, he pioneered the study of dependency lengths from a statistical standpoint putting forward the first baselines and the principle of dependency distance minimization. He also introduced the hypothesis that projectivity, the scarcity of crossings dependencies and consistent branching are epiphenomena of that principle.