

A mini-workshop on linguistic issues in computational modelling

**Room 432 Battelle Centre Universitaire d'Informatique
University of Geneva**

13th June 14:00-18:30

Program at a glance

Opening Welcome

14:00-14:40 The Morphosyntactic Encoding of Core Arguments – A Cross-Linguistic Perspective
Joakim Nivre — Uppsala University

14:40-15:20 Colorless green recurrent networks dream hierarchically
Kristina Gulordava — University of Geneva

15:20-16:00 Self-Organized Sentence Processing and Dependency Length Minimization
Whitney Tabor — University of Connecticut

16:00-16:20 Coffee break

16:20-17:00 Linguistic Yardsticks: Evaluating Language Technology Using Insights from Linguistic Theory
Laura Rimell — DeepMind

17:00-17:40 Lack of evidence of meaning effects in locality
Paola Merlo. — University of Geneva

17:40-18:20 Communicative Efficiency, Uniform Information Density, and the Rational Speech Act theory
Roger Levy — Massachusetts Institute of Technology

Closing

**14:00-14:40 The Morphosyntactic Encoding of Core Arguments –
A Cross-Linguistic Perspective
Joakim Nivre
Uppsala University**

Languages use essentially three mechanisms to encode grammatical relations like subject and object: word order, case marking and agreement marking (or indexing). The relative importance of different mechanisms vary across languages and they also interact in complex ways. For example, it appears that predominantly verb-final languages favor case marking, while verb-initial languages favor agreement marking and verb-medial languages disfavor both marking strategies (Siewierska and Bakker, 2012). Most of these generalizations, however, are stated at the level of complete languages, and much less is known about how the different encoding strategies are distributed and interact in specific sentences in a given language. In this talk, I will present very preliminary results from an exploration of word order and case marking for core argument relations based on treebanks annotated in the Universal Dependencies project. On the one hand, I will look at word order distributions for verb, subject and object in transitive main clauses and discuss different ways of measuring word order freedom in terms of entropy, including variants of relation order entropy and arc direction entropy (Futrell et al., 2015). On the other hand, I will look at the presence of different types of case marking in the same transitive main clauses and see how these patterns correlate with word order distributions.

**14:40-15:20 Colorless green recurrent networks dream hierarchically
Kristina Gulordava
University of Geneva**

Recurrent neural networks (RNNs) have achieved impressive results in a variety of linguistic processing tasks, suggesting that they can induce non-trivial properties of language. We investigate here to what extent RNNs learn to track abstract hierarchical syntactic structure. We test whether RNNs trained with a generic language modeling objective in four languages (Italian, English, Hebrew, Russian) can predict long-distance number agreement in various constructions. We include in our evaluation nonsensical sentences where RNNs cannot rely on semantic or lexical cues ("The colorless green ideas I ate with the chair sleep furiously"), and, for Italian, we compare model performance to human intuitions. Our language-model-trained RNNs make reliable predictions about long-distance agreement, and do not lag much behind human performance. We thus bring support to the hypothesis that RNNs are not just shallow-pattern extractors, but they also acquire deeper grammatical competence.

<https://arxiv.org/abs/1803.11138>

**15:20-16:00 Self-Organized Sentence Processing and Dependency Length
Minimization**

**Whitney Tabor
University of Connecticut**

Self-Organized Sentence Processing (SOSP) is a computational sentence processing framework in which small elements (words, morphemes) interact via continuous bonding-processes to form larger constituents. The framework has the advantageous property that it generates both grammatical and ungrammatical structures, thus making it suitable for modeling various known phenomena of aberrant language behavior—e.g., agreement attraction, local coherence, center-embedding difficulty. Here, we sketch an approach offered by this framework to dependency length minimization. Whereas some accounts of such phenomena have argued that languages minimize dependencies in order to minimize demands on memory, such arguments have a teleological flavor; they leave one wondering how languages managed to set themselves up to behave optimally in this regard. In SOSP, dependency length minimization follows as an emergent feature of the word-interactions. Basically, when a language affords multiple orders for a given meaning, there is a competition between the different structures that can express the meaning. Short-dependency structures can form more easily, so they tend to beat out their competition. In a survey of many languages, Futrell, Mahowald, & Gibson (2015) found that dependency length minimization is present in all of their test languages, but it is weaker in head-final than head-initial languages. We offer a possible insight into this asymmetry.

(Work with Julie Franck)

**16:20-17:00 Linguistic Yardsticks: Evaluating Language Technology Using
Insights from Linguistic Theory**

**Laura Rimell
DeepMind**

Language technology has achieved remarkable success on practical tasks, such as machine translation and sentiment analysis, while incorporating very little theoretical linguistic knowledge. However, the appearance of success may be deceiving, because standard evaluation metrics for language technology under-represent relatively rare but linguistically interesting phenomena. Poor performance in these areas will be increasingly noticeable as technology advances and users expect more human-like behaviour. I will describe work that evaluates language technology using linguistic yardsticks: datasets designed to focus on specific phenomena, such as the semantic understanding of relative clauses, and I will

consider how they may point the way toward improvements in natural language processing.

17:00-17:40 Lack of evidence of meaning effects in locality

Paola Merlo
University of Geneva

Some of the oldest and most established generalisations in syntax are 'constraints over variables', the observation that extractions from certain positions are ungrammatical and that this ungrammaticality derives from formal constraints, without semantic or processing influences. Recently, this point of view has been weakened by results arguing for semantic effects in syntactic islands (Gibson) or showing that ungrammaticality is graded and sensitive to notions such as animacy (Villata and Franck). For modelling, many of the notions invoked in these explanations could be made more precise. We study the locality theory of Relativised Minimality, whose core is the notion of 'intervener'. For a semantic version of this theory to be at play in explaining extraction violations or infelicities, the notion of intervener must be defined in semantic terms. We formalise the notion of 'semantic' as the popular notion of 'lexical word embeddings' and the notion of 'similarity' used to defined interveners as a distance between word embedding vectors. We present preliminary results where, under these formal, precise definitions, we fail to find semantic effects in extraction from weak islands and agreement errors in object relative clauses. While negative results are always hard to interpret, there is at least one theory and one encoding of this theory under precise conditions that shows that extraction constraints are not subject to semantic modulation.

(Work with Francesco Ackermann)

17:40-18:20 Communicative Efficiency, Uniform Information Density, and the Rational Speech Act theory

Roger Levy
Massachusetts Institute of Technology

One major class of approaches to explaining the distribution of linguistic forms is rooted in communicative efficiency. For theories in which an utterance's communicative efficiency is itself dependent on the distribution of linguistic forms in the language, however, it is less clear how to make distributional predictions that escape circularity. I propose an approach for these cases that involves iterating between speaker and listener in the Rational Speech Act theory. Characteristics of the fixed points of this iterative process constitute the distributional predictions of the theory. Through computer simulation I apply this approach to the well-studied

case of predictability-sensitive optional function word omission for the theory of Uniform Information Density, and show that the approach strongly predicts the empirically observed negative correlation between phrase onset probability and rate of function word use.

<https://psyarxiv.com/4cgxh>

Practical Information

Centre Universitaire d'informatique (CUI)

Battelle - bâtiment A

7, route de Drize

CH-1227 Carouge

