Dr Kristina Gulordava

Université de Genève

Word order is one of the most intriguing aspects of the human language. Why in some languages like French we can say 'une très belle soirée' or 'une soirée très belle' but in other languages like English or Russian we can use only the first ordering? The availability of two options in French is the case of word order variation. By studying which of the orders speakers prefer in particular contexts, we can reveal general factors that shape word order in a language and across languages. The focus of this thesis is on one factor known as dependency length minimisation (DLM). DLM is the tendency for related parts of an expression (dependents) to appear linearly adjacent: while an order such as 'tasty eat a cake' might be possible in some languages, it is more common to observe an order 'eat a tasty cake' or 'eat a cake tasty' where the adjective ('tasty') describing the property of the noun ('cake') appears next to it.

This thesis analyses DLM phenomena in word order variation using a computational approach. The work capitalises on linguistic resources (texts annotated with relations of syntactic dependencies), available in many languages, and statistical methods which are essential for drawing generalisations from word order variation data. The new computational methodology allows us to statistically compare the extent to which languages follow the DLM principle and to reveal small but significant DLM effects in adjective variation in Romance languages. A computational model of incremental production is proposed to explain these observations in a unified way.