Detecting constraints on clitic climbing – with the help of corpora and psycholinguistic tests

The talk aims to show how corpora can be used to study fairly complex phenomena. We will base the discussion on the example of constraints on clitic climbing (CC) in Bosnian, Croatian and Serbian (BCS). According to Junghanns (2002: 58) CC is associated with matrix embedding constructions: “In komplexen syntaktischen Ausdrücken bewegt sich ein klitisches Pronomen aus der Einbettung in die Matrix.” An example of CC out of an infinitival complement is given in (1) where the clitical pronoun ga ‘him’ is realised in the second position of the matrix clause (Wackernagel position); in other cases, however, CC does not take place as in (2) where the clitic ih stays in the complement clause.

(1)  
Milan  ga₂  mora₁  vidjeti₂.  

Milan  him.ACC  must.3PRS  see.INF  

‘Milan must see him.’  

Stjepanović (2004: 179f)

(2)  
Bojim₁  se₁  testirati₂  ih₂.  

afraid.1PRS  REFL  test.INF  them.ACC  

‘I am afraid to test them.’  

hrWaC v2.2

Although clitics in Bosnian, Croatian and Serbian (BCS) have attracted considerable attention in the syntactic literature (cf. Franks & King 2000, Browne 2014, or Bošković 2004), the syntactic conditions and constraints for CC are seriously understudied in comparison to e.g. Czech (e.g. Junghanns 2002). Jurkiewicz-Rohrbacher et al. (2017a, 2017b), Hansen et al. (2018) are the first descriptions of CC in BCS based on empirical investigations. Basing on the data obtained from massive web corpora {bs, hr, sr}WaC (Ljubešić & Klubička 2014), the raising-control dichotomy of matrix predicates is shown to be a relevant factor of CC. Apart from that, it is found out that reflexivity plays a major role. Kolaković et al. (accepted), on the other hand, tackle the question of register as a relevant factor by comparing results from Forum subcorpus of hrWaC v2.2, Croatian Language Repository (Ćavar & Brozović Rončević 2012) Croatian National Corpus (Tadić 2009) while examining the same types of matrix predicates.

First, the talk presents the results of the corpus based and corpus driven studies mentioned above, discusses in detail the particular steps of a corpus approach, ranging from the formulation of queries, coping with tagging errors, to the statistical analysis of the data. Second, it will show how these results feed into a major psycholinguistic experiment recently carried
out in Croatia (7 experiments x 40 participants = 280 participants). The logistic regression mixed models based on data from the speeded yes-no grammaticality judgment tasks with OpenSesame free software provide the additional evidence for constraints on CC. Finally, we shall present the main findings of our study on constraints on CC in BCS. We will argue that the constraints encompass i) what we call tied islands, ii) predicate types and iii) mixed cluster effects.

Bibliography


