Edyta Jurkiewicz-Rohrbacher & Zrinka Kolaković &Björn Hansen & Dušica Filipović-Đurđević

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*   ***ga*$_2$**   *mora*$_1$ *vidjeti*$_2$.

       Milan   him.ACC   must.3PRS   see.INF

       'Milan must see him.'   Stjepanović (2004: 179f)

(2)   *Bojim*$_1$ ***se*$_1$**   *testirati*$_2$ ***ih*$_2$**.

       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

Aljović, N. (2004) "Cliticization Domains: Clitic Climbing in Romance and in Serbo-Croatian." In: Crouzet, O. et alii (eds.) *Proceedings of JEL'2004 Domain(e)s,* Université de Nantes, 169-175.

Aljović, N. (2005) "On clitic climbing in Bosnian/Croatian/Serbian". In: Leko, N. (ed.) *Lingvistički vidici* 34:(05). Sarajevo: Forum, 58-84.

Bošković, Ž. (2001) *On the nature of the syntax-phonology interface: cliticization and related phenomena*. Amsterdam: Elsevier.

Browne, W. (2014) "Groups of Clitics in West and South Slavic Languages". In: Kaczmarska, E.; Nomachi, M. (eds.) *Slavic and German in Contact: Studies from Areal and Contrastive Linguistics*. Slavic Eurasian Studies 26, 81-96

Ćavar, D., Brozović-Rončević, D. (2012) "Riznica: The Croatian Language Corpus". In: Prace filologiczne 63, 51-65.

Franks, S. & King, T. H. (2000) *A Handbook of Slavic clitics*. Oxford: OUP.

Hansen, B.; Kolaković, Z.; Jurkiewicz-Rohrbacher, E.; (2018) "Clitic climbing and infinitive clusters in Bosnian, Croatian and Serbian − a corpus-driven study." In: Fuß, Eric et al., *Grammar and Corpora 2016*. Heidelberg: Heidelberg University Publishing (heiUP).

Junghanns, U. (2002) „Clitic climbing im Tschechischen". In: *Linguistische Arbeitsberichte* 80, 57-90.

Jurkiewicz-Rohrbacher, E.; Kolaković, Z.; Hansen, B. (2017) "Web Corpora – the best possible solution for tracking rare phenomena in underresourced languages: clitics in Bosnian, Croatian and Serbian". In: Bański, P. et al. (eds.): *Proceedings of the Workshop on Challenges in the Management of Large Corpora and Big Data and Natural Language Processing (CMLC-5+BigNLP) 2017 including the papers from the Web-as-Corpus (WAC-XI) guest section*. Mannheim: Institut für Deutsche Sprache, 49-55.

Jurkiewicz-Rohrbacher, E.; Hansen, B.; Kolaković, Z. (2017) "Clitic climbing, finiteness and the raising-control distinction. A corpus-based study." In: *Jazykovedný časopis* 68:(2), 179-190.

Kolaković, Z.; Hansen, B.; Jurkiewicz-Rohrbacher, E.; (accepted) "Uspon zanaglasnice, dihotomija dizanje – kontrola i stilska varijacija". *6. Hrvatski sintaktički dani - Sintaksa zavisnosložene rečenice*, 17.-19.05.2018 Osijek.

Ljubešić, N., Klubička, F. (2014) "{bs,hr,sr}WaC — Web corpora of Bosnian, Croatian and Serbian". In: Bildhauer, F., Schäfer, R. (eds.) *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*. Gothenburg, 29-35.

Stjepanović, S. (2004) "Clitic Climbing and Restructuring with "Finite Clause" and Infinitive Complements". In: *Journal of Slavic Linguistics* 12: 1, 173-212.

Tadić, M. (2009) "New version of the Croatian National Corpus". In: Hlaváčková, D. et al. (eds.) *After Half a Century of Slavonic Natural Language Processing*. Masaryk University, Brno, 199-205.