

## DARIO LEČIĆ

The University of Sheffield (PhD) & Sveučilište u Zadru

### *WHAT CAN CORPORA TELL US ABOUT MENTAL GRAMMARS?*

Tuesday, May 16, 1745–1915, room 420, nám. J. Palacha 2

ABSTRACT: The author presents the results of his PhD research entitled *Morphological Doublets in Croatian: a multi-methodological analysis*. Morphological doubletism (also called morphological synonymy, overabundance) is the situation in language when there are two (or more) grammatical morphemes with the same function within an inflectional paradigm, which are not in complementary distribution, i.e. are completely interchangeable. Some examples from Croatian: *car-om/car-em* as the instrumental singular of a-declension, *bitk̄a/bitk̄i /bitākā* as the genitive plural of e-declension, *čist iji/čišć-i* as two comparative forms etc.

Having determined the corpus distributions of the individual forms in Croatian corpora, the author conducted several questionnaire surveys among native speakers of Croatian, testing for acceptability and production of individual forms. The results have shown that the differences in the corpus distributions of doublets are reflected in the speakers' answers, i.e. the acceptability/production of one form decreases as its level of "domination" over its doublet pair decreases. Such results, confirmed in three independent surveys, lead us to the conclusion that speakers' mental grammars/lexicons contain a piece of information about the frequency of particular forms (in the words of Joan Bybee, "Frequency distributions matter and are a part of grammar").

### *WHAT TO DO WHEN RULES FAIL US: COMPUTATIONAL MODELLING OF (CROATIAN) LANGUAGE*

Wednesday, May 17, 1410–1540, room 424, Celetná 20

ABSTRACT: If a speaker (or learner) of Croatian consults grammars and other reference manuals of the Croatian language in search of a solution to some language question (more specifically morphological doublets), more often than not they will be left confused as they will find ambiguous answers in a single manual or opposite answers in different manuals. If they take a look at how this item is used in real life (derived from a corpus), they might find completely different distributions. Where does that leave the speaker/learner? Whose "rules" should they follow? Is a rule-based system applicable at all or should they rely on a different principle?

In recent years, numerous computational models of language have been developed that try to dig into the processes taking place in the speakers' minds when processing and producing language. Some of these models rely solely on memory and eliminate rules from this process altogether. We test two such models (Analogical Model and Tilburg Memory-Based Learner) on Croatian, using morphological doublets as test cases. More specifically, we try to see whether the answers these models produce would be more in line with the answers given by grammar books or by the patterns found in the corpora. On top of this, the performance of these models (which we interpret as "artificial" speakers of Croatian) is compared to the results of real speakers collected by means of a survey.

### *MOŽEMO LI VJEROVATI HRVATSKIM GRAMATIKAMA (I ZAŠTO NE)?*

Thursday, May 18, 1230–1410, room 308A, nám. J. Palacha 2

ABSTRACT: Na hrvatskom tržištu knjiga trenutačno se može pronaći desetak normativnih priručnika (gramatika, jezičnih savjetnika i pravopisa). Kako bi se opravdalo njihovo izdavanje, svaki od ovih priručnika mora donijeti nešto novo, odnosno mora sadržavati izmijenjena pravila i preporuke u odnosu na prethodne priručnike. No, koliko svi ovi priručnici realno oslikavaju hrvatski jezik? Koliko se govornici u jezičnoj neodumici mogu osloniti na njih?

U ovom izlaganju uspoređujemo preporuke koje možemo naći u hrvatskim normativnim priručnicima za neka morfološka i sintaktička pitanja (morfološke dublete, slaganje itd.) s primjerima upotrebe iz korpusa. Iz ovakvog pregleda postaje jasno da hrvatska normativna literatura nije u toku s uporabnom normom. Većina normativnih priručnika svoje preporuke temelji na jeziku "dobrih" pisaca, što nije jezik kojim govori većina Hrvata. Osim toga, većina je normativnih priručnika previše ekskluzivna, odnosno preporučuju isključivo jedan oblik iako se oba upotrebljavaju podjednako.