

*Max Wahlström* (Helsinki)

## A SPEECH CORPUS STUDY OF TWO IMPERSONAL STRATEGIES IN BULGARIAN AND CROATIAN<sup>1</sup>

In this paper, I address the use of two human impersonal strategies in Croatian and Bulgarian: the so-called *man*-impersonal and the second person singular. Here, impersonality is understood in a semantic and pragmatic sense: impersonality is the reduction of referentiality or lack thereof. Consider the following phrase: *when you reach zero gravity, you start to float*. Despite the use of 2sg verb form, it is not in reference to the addressee, but to any human being. Such impersonal strategies as the use of 2sg above are called human impersonals or reference impersonals (Siewierska 2011). Sometimes the term impersonality is used in a syntactic sense meaning the lack of subject argument or lack of subject agreement on the verb (cf. Bg. *bezičnost*). However, the lack of a subject argument or the lack of subject agreement is independent of semantic and pragmatic impersonality (explicit subject, but impersonal or generic reading: ...*you start to float*; no nominative subject, but referential reading: Bg. *studeno mi e*).

Human impersonals in South Slavic languages constitute a largely uncharted territory, especially regarding spoken language. While the term impersonality (Ru. *impersonalnost'*) occurs in grammatical descriptions and research, it is used almost exclusively to denote syntactic phenomena. However, the term general-personal or generic-personal (Cr. *uopćeno-lično*, Bg. *obobšteno-lično*) is used in reference to human impersonal strategies. Occasionally, also the term indefiniteness (Cr. *neodređenost*) is used. In grammatical descriptions of Bulgarian, *čovek* 'person' has been identified in passing as a human impersonal device (see, e.g., Vlahova-Rujkova 2009; Feuillet 1996: 253), and some, mostly contrastive studies address it (Venkova 1997; Dimova 1981). Sometimes *čovek* in this function is called, rather infelicitously,

---

<sup>1</sup> This paper reports a part of an ongoing study, contrasting human impersonal strategies in Western Romance and South Slavic (Posio & Wahlström [under preparation]).

an expletive subject (*ekspletiven podlog*; see, e.g., Atanasov 2015: 44), although its use is syntactically rather unrestricted: it can act, for instance, as a shared subject of a converbial (*deepričastie*) clause. In Bulgarian, where NPs are obligatorily marked for definiteness, the impersonal uses of *čovek* are formally distinct from nouns with specific reference by lacking the definite (or indefinite) article. The article is omitted also with nouns that are used in reference to a whole species, but only *čovek* and its suppletive plural *hora* may be used also impersonally. Grammatical descriptions of Croatian typically do not mention the impersonal or generic use of the noun *čovjek*, but there are a few studies addressing its status (Kordić 2002; Marojević 1977).

The two human impersonal strategies contrasted in this study, the use of the noun denoting ‘person’ and the 2sg, are both reported in the respective linguistic traditions of Bulgarian and Croatian. However, there are no quantitative studies addressing their scope and frequency, especially in spoken language. In addition, speech corpora allow the study of 2sg as an impersonal strategy, since it is typically used only in informal contexts. In typological literature, it has been suggested that the impersonalization strategy based on the noun ‘person, man’ is a feature of the so-called Standard Average European, centered around languages such as French and German (Ramat & Sansò 2007). In addition, it is claimed that the grammaticalization *man*-impersonals would be especially favored in languages that require an explicit subject (non-*pro* drop languages; Siewierska 2011). It has been shown that there is a statistically significant difference in the frequency of overt subjects between Bulgarian and BCMS (Bosnian, Croatian, Montenegrin, Serbian), Bulgarian displaying more overt subjects (Seo 2001). It may be hypothesized that Bulgarian could, therefore, show preference for an impersonal strategy with an overt subject.

The choice to contrast Bulgarian with Croatian is due to two main reasons. The availability of suitable speech corpora being an important factor, speech data gathered exclusively from the territory of the Republic of Croatia has yet some added benefits. While some South Slavic dialects spoken in Serbia are transitional between Bulgarian, Macedonian, and Serbian, the Croatian speech area is geographically discontinuous from Bulgarian. Spoken Croatian offers therefore a clearly distinct point of comparison with spoken Bulgarian varieties, while still being relatively closely related.

The Bulgarian data comes from the *Corpus of Spoken Bulgarian* (CSB; Aleksova 1994) and consists of 35 spontaneous family conversations, collected between mid-80’s–mid-90’s. The corpus contains 79,431 words. There

are 65 working-age and 6 underage informants. All informants were residents of the capital Sofia at the time of recording, but may have been born in other parts of Bulgaria. The Croatian data is based on the *Croatian Adult Spoken Language Corpus* (HrAL; Kuvač Kraljević, Hržica 2016). The corpus consists of 165 predominantly spontaneous conversations among friends, relatives, or acquaintances. It was collected between 2010 and 2016, contains 285,811 words and includes 617 adult informants. The corpus was collected with the aim of representing conversations in as many locations within the Republic of Croatia as possible, but with no explicit goal of targeting speakers of local dialects. The topics of the conversations are not restricted to particular themes, but daily life topics and discussions of recent personal experiences dominate. The conversational character of the recordings keeps turns short and most informants are in familiar terms with each other — there is, for instance, very little use of the polite 2pl in addressing the other conversants. In general, both corpora being analyzed are characterized by being fairly informal, and having the second person singular as the default address form.

The data shows that the *man*-impersonal strategy exists in Bulgarian, although it is not an especially frequent impersonal device in comparison to other impersonal strategies. However, despite the larger size of the Croatian corpus, the impersonal use of *čovjek* seems very marginal, with only a handful of reliable examples. Yet Croatian displays much more frequent impersonal use of the 2sg than Bulgarian. The inverse correlation between these two strategies seems to be in line with the hypothesis: Bulgarian, with its preference for overt subject marking, may also prefer a noun-based impersonal strategy. However, since there is a drastic difference in the overall frequency between these two strategies in both languages, further studies must extend the comparison to other human impersonal strategies, such as the third person plural and the reflexive pronoun *se*.

#### REFERENCES

- Aleksova 1994 — *Aleksova K.* Corpus of spoken Bulgarian. Sofia: 1994. (Accessed at <http://folk.uio.no/kjetilrh/bulg/Aleksova/>, on 21.01.2019).
- Atanasov 2015 — *Atanasov A.* Klasifikacija na različnite predikati v sãvremennija bãlgarski ezik. *Bãlgarski ezik* 62(3), 2015. S. 38–53.
- Dimova 1981 — *Dimova A.* Die Polysemie des deutschen Pronomens „man“ unter Berücksichtigung seiner Äquivalente im Bulgarischen // *Deutsch als Fremdsprache* 18(1), 1981. S. 38–44.

- Feuillet 1996 — *Feuillet J.* Grammaire synchronique du bulgare. Paris: Institut d'études slaves, 1996.
- Giacalone Ramat, Sansò 2007 — *Giacalone Ramat A., Sansò A.* The spread and decline of indefinite man-constructions in European languages // Europe and the Mediterranean as linguistic areas: convergencies from a historical and typological perspective / Ramat P., Roma E. (eds.). P. 95–132. Amsterdam; Philadelphia: John Benjamins Publishing Company, 2007.
- Kordić 2002 — *Kordić S.* Riječi na granici punoznačnosti. Zagreb: Hrvatska sveučilišna naklada, 2002.
- Kuvač Kraljević, Hržica 2016 — *Kuvač Kraljević J., Hržica G.* Croatian adult spoken language corpus (HrAL). Zagreb, 2016. (Accessed at <https://ca.talkbank.org/access/Croatian.html>, on 21.01.2019).
- Marojević 1977 — *Marojević R.* Rečenice sa uopšteno-ličnim značenjem u ruskom jeziku u poređenju sa srpskohrvatskim // Južnoslovenski filolog 33, 1977. S. 99–116.
- Posio, Wahlström (under preparation) — *Posio P., Wahlström M.* Human impersonals in Western Romance and South Slavic: A speech corpus study of man-impersonals.
- Seo 2001 — *Seo S.* The frequency of null subject in Russian, Polish, Czech, Bulgarian and Serbo-Croatian: An analysis according to morphosyntactic environments. Bloomington: Indiana University PhD thesis, 2001.
- Siewierska 2011 — *Siewierska A.* Overlap and complementarity in reference impersonals. Man-constructions vs. third person plural impersonals in the languages of Europe // Impersonal Constructions. A cross-linguistic perspective / Malchukov A., Siewierska A. (eds.). Amsterdam; Philadelphia: John Benjamins, 2011. P. 57–89.
- Venkova 1997 — *Venkova V.* Šaštstvitelnite čovek i hora v portugalskija i v bälgarskija ezik // Šapostavitelno ezikoznanie 22(4), 1997. S. 5–12.
- Vlahova-Rujkova 2009 — *Vlahova-Rujkova R.* Praktičeska gramatika: bälgarski ezik. Sofija: Klet Bälgarija, 2009.